

עיבוד תמונות

מיכאל אלעד



הקדמה

ספר זה נכתב כאסופת הרצאות לקורס מתקדם בעיבוד תמונות בפקולטות להנדסת חשמל ולאחר מכן מדעי המחשב בטכניון. בגירסה חדשה זו (אוקטובר 2003) של הספר נוספו מספר פרקים המכסים נושאים בטומוגרפיה ממוחשבת, ייצוגים מולטי-רזולוציוניים, עיבוד תמונות צבע, ונספח בו מובאות הרחבות בנושאים שונים כגון קוונטיזציה, שחזור והתמרות.

אין בכוונתי לחדש בספר זה אלא לרכז באופן קריא ונוח (ובעיקר בשפה העברית) דיון בנושאים מגוונים בעיבוד תמונות. דגש מרכזי מושם על עדכניות החומר המובא כאן, ועל ריבוי בדוגמאות לשם המחשתן. הדוגמאות במהלך הפרקים נוצרו ע"י תוכנת MATLAB.

בניית ספר זה מתבססת על מספר מקורות המתוארים בהמשך. זוהי עדיין גירסה לא סופית ולצערי עדיין ישנן שגיאות במהלך הפרקים. אני מקווה לתקן ולעדכן ספר זה לכלל ספר נוח ואמין ללימוד עיבוד תמונות ולהוסיף לתוכנו, כך שיוכל להוות בסיס לשני קורסים – מתחיל ומתקדם בתחום.

אני רוצה להודות לד"ר דורון שקד על סיועו בקביעת תוכן הספר, ושיחותיו עמי על נושאים שונים הנסקרים בו.

מקורות

- [1] Anil. K. Jain, *Fundamentals of Digital Image Processing*, Prentice Hall Englewood Cliffs, New-Jersey, 1989.
- [2] Kenneth R. Castleman, *Digital Image Processing*, Prentice Hall Englewood Cliffs, New-Jersey, 1996.
- [3] Brend Jahne, *Digital Image Processing – Concepts, Algorithms, and Scientific Applications*, Springer Verlag, 1995.
- [4] Boaz Porat, *A Course In Digital Signal Processing*, Wiley, New-York, 1997.
- [5] Dan E. Dudgeon and Russell M. Mersereau, *Multi-Dimensional Digital Signal Processing*, Prentice-Hall, Englewood-Cliffs, New Jersey, 1984.
- [6] George Wolberg, *Digital Image Warping*, IEEE Computer Society Press, Los-Alamitos California, 1990.
- [7] Nahum Kiriati, Doron Shaked, and Nir Sochen, *Lecture Notes in Digital Image Processing*, The Computer-Science Department, The Technion, Israel Institute of Technology (Obtained by personal communication).
- [8] Alfred M. Bruckstein, *Image Processing – Lecture Notes*, the Computer Science Department, The Technion, Israel Institute of Technology.
- [9] Michael Lindenbaum, *Lecture Notes in Digital Image Processing*, The Computer-Science Department, The Technion, Israel Institute of Technology. Web-site: <http://www.cs.technion.ac.il/~mic/isp.html>
- [10] Hagit Zabrodzki Hel-Or, *Lecture Notes in Image Processing*, The Computer-Science Department, Haifa University. Web-site: http://cs.haifa.ac.il/courses/image_p.

ומאמרים שונים.

עיבוד תמונות

תוכן עניינים

1	<u>פרק 1 – כללי</u>
2	1.1 תמונה דיגיטלית
5	1.2 קלט/פלט לתמונות דיגיטליות
8	1.3 סוגיות בטיפול בתמונות
10	1.4 עיבוד תמונה ותחומים סמוכים
11	1.5 יישומים לעיבוד ואנליזה של תמונות
15	<u>פרק 2 – עיבוד אותות בדו-מימד</u>
15	2.1 מערכות ליניאריות בדו-מימד – אותות מעל הרצף
19	2.2 התמרת פוריה דו-מימדית ותכונותיה
23	2.3 מערכות ליניאריות בדו-מימד – אותות דיסקרטיים
25	2.4 התמרת פוריה לאותות דיסקרטיים ותכונותיה
29	2.5 אותות ומערכות ליניאריות בתמך סופי
39	2.6 התמרת פוריה דיסקרטית (DFT) לאותות בתמך סופי
41	<u>פרק 3 – דגימה ושחזור של תמונות</u>
41	3.1 דגימת אחידה של אותות דו-מימדיים
44	3.2 שחזור מדגימה אחידה
46	3.3 תופעת הקיפול בתמונות
51	3.4 משפטי דגימה מוכללים לדו-מימד
55	<u>פרק 4 – קוונטיזציה לתמונות</u>
55	4.1 קוונטיזציה סקלרית
59	4.2 קוונטיזר מקס-לויד
67	4.3 קוונטיזציה עם שיקולים חזותיים
71	4.4 קוונטיזציה וקטורית וקוונטיזציית צבע

75	<u>פרק 5 – עיבודים בסיסיים לתמונות</u>
76	5.1 פעולות נקודה והיסטוגרמת תמונה
83	5.2 פעולות עיצוב היסטוגרמה
96	5.3 פעולות מרחביות בסיסיות
107	5.4 פעולות מרחביות מסתגלות
113	<u>פרק 6 – שחזור תמונה</u>
113	6.1 בעיית השחזור הקלאסית
116	6.2 משערכים סטטיסטיים – ML ו-MAP
122	6.3 שחזור מבוסס משערך ML
125	6.4 שחזור מבוסס משערך MAP
132	6.5 שיפור פונקציות ההסתברות לתמונות
140	6.6 בעיות היפוך אחרות והקשר לבעיית השחזור הקלאסית
143	<u>פרק 7 – התמרות דיסקרטיות בדו-מימד</u>
143	7.1 התמרות ליניאריות - בסיס
148	7.2 התמרת ה-DFT הדו-מימדית ותכונותיה
151	7.3 התמרת ה-DCT הדו-מימדית ותכונותיה
156	7.4 התמרת Hadamard דו-מימדית ותכונותיה
160	7.5 התמרת Haar ומבוא ל-Wavelet
168	7.6 התמרות מבוססות תוכן תמונה
171	7.7 התמרות Over-Complete
175	<u>פרק 8 – דחיסת תמונה</u>
175	8.1 יסודות בתורת האינפורמציה
181	8.2 יתרונות בתמונה וניצולן
184	8.3 קידוד אנטרופיה - אלגוריתם הופמן
188	8.4 קידוד עם שגיאות - גישת החיזוי
191	8.5 קידוד עם שגיאות - גישת ההתמרה
195	<u>פרק 9 – עיבוד סדרות של תמונות</u>
196	9.1 תנועה בין תמונות
201	9.2 שערך תנועה בין תמונות - גישת בלוקים
205	9.3 שערך תנועה בין תמונות גישה מלאה

210 9.4 ניקוי רעש בסדרות של תמונות
215 9.5 דחיסת סדרות של תמונות ומבוא ל-MPEG
219 <u>פרק 10 - שחזור תמונה מהיטלים</u>
220 10.1 בעיות שחזור מהיטלים וניסוחן המתמטי
221 10.2 התמרת ראדון
223 10.3 משפט החתך המרכזי והיפוך התמרת ראדון
226 10.4 שחזור מהיטלים ע"י גישות תדר
228 10.5 שחזור מהיטלים ע"י קונבולוציה
229 10.6 שחזור מהיטלים ע"י גישות אלגבריות
235 <u>פרק 11 - מבנים פירמידליים ושימושיהם בתמונות</u>
236 11.1 הפירמידה הגאוסית
240 11.2 הפירמידה הלפלסיאנית
249 11.3 התמרת Wavelet ע"י Lifting Scheme
257 <u>פרק 12 - מבוא לתורת הצבע</u>
258 12.1 צבע – מהפיזיקה לעין האנושית
261 12.2 התאמות צבע
265 12.3 שיטות ייצוג צבע בתמונות
271 12.4 ייצוג תמונות צבע
276 12.5 שיפור ושחזור תמונות צבע
283 12.6 דחיסת תמונות צבע
285 <u>נספחים:</u>
286 נספח א' – הרחבות לקוונטיזר אופטימלי
294 נספח ב' – התמרת KLT כהתמרה אופטימלית והרחבותיה
299 נספח ג' – הקצאת סיביות אופטימלית בדחיסה
302 נספח ד' – נושאים מתקדמים בשחזור תמונות

פרק 1 – כללי

1	<u>פרק 1 – כללי</u>
2	1.1 תמונה דיגיטלית
5	1.2 קלטופלט לתמונות דיגיטליות
8	1.3 סוגיות בטיפול בתמונות
10	1.4 עיבוד תמונה ותחומים סמוכים
11	1.5 יישומים לעיבוד ואנליזה של תמונות

במסגרת פרק זה ננסה להגדיר מהי תמונה דיגיטלית, ומה ניתן ורצוי לבצע עליה. שאלה זו והתשובות האפשריות לה מגדירים תחום מחקר ופיתוח מרתק, המוכר בשם הכולל "עיבוד תמונות". חשוב לציין כי תחום זה קיים מלכתחילה בעיקר בשל ההתקדמות המטאורית בנושא המיחשוב. כל העיבודים אותם נכיר במסגרת ספר זה הינם "דיגיטליים", דהיינו – אלגוריתמים מבוססי מחשב, וזאת בניגוד לעיבוד אנלוגי או אופטי. לפיכך, שם ראוי יותר לספר זה יהיה "עיבוד וניתוח ספרתי של תמונות".

איפוא אנו פוגשים תמונות? מדוע כלל חשוב לדון בעיבודן? אלבום תמונות משפחתי, הטלוויזיה, האינטרנט, קריאה בעיתון, פרסומות רחוב ועוד –אלה הם מקורות אשר מתרכזים בהעברת מידע ויזואלי אלינו. מכנה משותף מרכזי למקורות אלו ואחרים הוא רצונם לנצל את שתי המצלמות המוצבות בפנינו – העיניים. חוש הראיה מספק לנו בני האדם ולמגוון יצורים אחרים את עיקר המידע על-פיו אנו מתמצאים במרחב, פועלים, ואפילו חושבים. מחקרים מלמדים שחלק ניכר במוח האדם משמש לעיבוד התמונה הנקלטת בעיניים, וניתוחה. בהקשר זה, האמירה "תמונה אחת שווה אלף מילים" ממחישה את כמות האינפורמציה שניתן לדלות מתמונה אופיינית.

כיצד המוח "מעבד" את התמונה? איזה פעולות התמונה עוברת בדרכה מהחיישנים בעיניים ועד לגמר עיבוד זה? שאלות מסוג זה שאלו את עצמם חוקרים במהלך מאה זו. על אף ריבוי המחקרים בתחום זה, האנושות רחוקה מלהבין את התהליכים המוחיים הכרוכים בטיפול במידע הויזואלי המוזן למוח. מוטיבציה ראשונה, אם כן, למחקר ב"עיבוד תמונות" הוא הניסיון להבין, ולאחר מכן לחקות את דרך פעולת המוח ולפתור בעיות הקשורות בעיבוד מידע ויזואלי, אשר ידוע לנו שהמוח פותר. כאמור, הבנתנו את דרך פעולת המוח בהקשר זה הינה בסיסית ביותר.

מכיוון הפוך, "עיבוד תמונות" עוסקת גם בהגדרת בעיות הקשורות בתמונות ופתרון ע"י אלגוריתמים ממוחשבים, תוך ניתוק מהשאלה על דרך פעולתו של המוח. דווקא כיוון זה מניב תוצאות רבות ואיכותיות, וזהו הכיוון בו "עיבוד תמונה" זכתה בדרך כלל להוות פריצת דרך.

בספר זה אנו סוקרים פרקי יסוד ב-"עיבוד תמונות", וגישתנו תהיה כזו המנותקת ממערכת הראיה. "עיבוד תמונה" אינו תיאוריה מגובשת! הפרקים אותם נסקור נשענים על תורות שונות ומקורות שונים. ניתן להתייחס ל-"עיבוד תמונה" כארגז כלים מתוחכמים לפתרון בעיות המתעוררות לפרקים בטיפול בתמונות דיסקרטיות. זו גם הסיבה שלימוד אמיתי של "עיבוד תמונות" חייב להיות מלווה בתרגילי מחשב בהם מתרגלים את האלגוריתמים (כלי העבודה מארגז הכלים) הנלמדים.

בפרק זה אנו מתחילים את הדיון בהגדרת מהותה של תמונה דיסקרטית, אפיון מקורותיה ויעדיה. לאחר מכן אנו מזכירים בקצרה את מגוון הפעולות הנחוצות על תמונות, כשאנו מסווגים משימות אלה לתחומי מחקר שונים שהתגבשו במהלך שני העשורים האחרונים. לסיום פרק זה אנו מזכירים מספר מצומצם של יישומים המערבים ברמה זו או אחרת עיבוד תמונות.

1.1 תמונה דיגיטלית

על מנת שנוכל להפעיל אלגוריתם על תמונה, נדרש קודם להביאה אל המחשב. אלגוריתם אופייני יניח כי התמונה הינה קובץ מחשב לכל דבר! איך נוצר קובץ זה, ואיך הוא קשור לתמונה אמיתית? ובכן, נניח כי נתונה לנו תמונה על נייר – נניח בשלב זה כי התמונה היא תמונת שחור-לבן, כפי שמתואר בציור 1.1. הרעיון הבסיסי הינו להתייחס לתמונה זו כפונקציה, שלכל מקום במישור (x,y) תייחס "רמת אפור" $f(x,y)$. אנו נניח כי גוון שחור מוחלט הוא גובה 0 ולבן מוחלט הוא גובה 1, וגווי האפור שביניהם יקבלו ערכים ממשיים בתוך האינטרוול $[0,1]$. ציור 1.1 מראה כיצד נראית התמונה הן כתמונה והן כפונקציה.

על-פי הציור הנ"ל, הפינה השמאלית העליונה הינה שחור מוחלט ולכן גובהה הוא אפס, והפינה הימנית התחתונה בעלת גובה 1 בשל היותה לבן מוחלט. בתמונה המתוארת ישנו מעבר חלק משחור ללבן, והדבר ניכר גם בתמונה וגם ביריעה בתיאור התלת-מימדי.

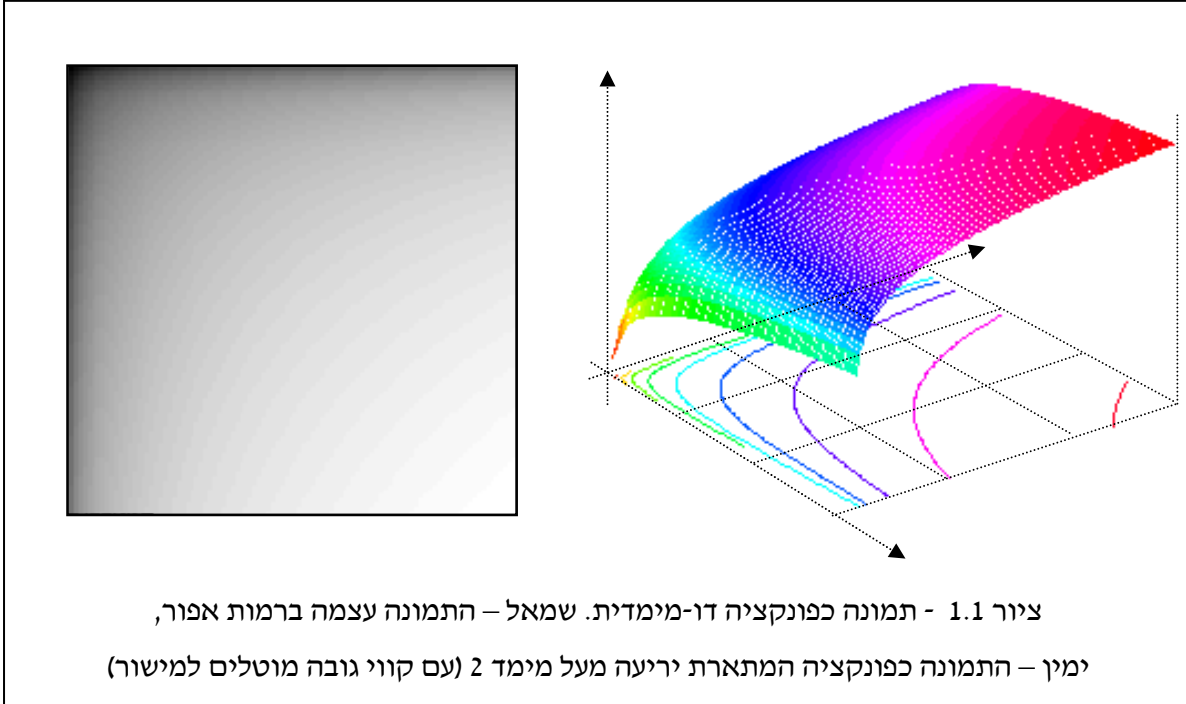
הערה: יתכן שבהדפסת התמונה (צד שמאל של ציור 1.1) ניכרות מדרגות בה. מדרגות אלה אינן חלק מהתמונה ומקורן במגבלות תהליך ההדפסה, עליו עוד נדבר בהמשך.

ברור כי הפונקציה $f(x,y)$ מוגדרת מעל תחום הגדרה מוגבל Ω (תחום התמך של התמונה), וחסרת משמעות מחוצה לו. באופן פורמלי יותר נאמר כי התמונה היא התאמה מהצורה הבאה:

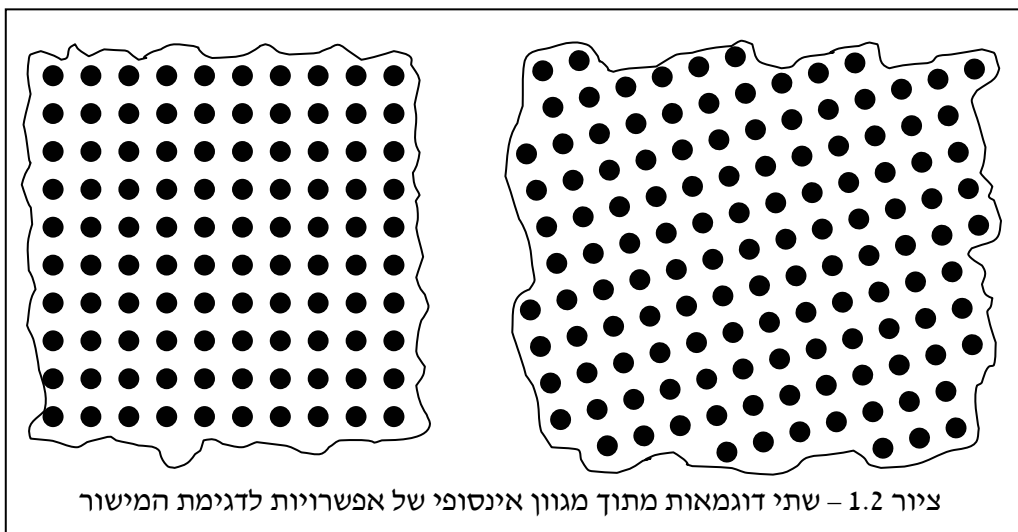
$$f : (x,y) \in \mathfrak{R}^{\Omega} \rightarrow \mathfrak{R}^{[0,1]}$$

כל זה טוב ויפה, אך בודאי לא מאפשר טיפול ספרתי, בשל שתי בעיות:

1. התמונה מוגדרת מעל כל נקודה (x,y) ממשיים בתחום ההגדרה, וזה אומר אינסוף בלתי ניתן למניה של נקודות.
2. בכל נקודה מקבלת הפונקציה ערך ממשי באינטרוול $[0,1]$ - ערך זה ידרוש ייצוג מדויק ע"י אינסוף סיביות.



שתי בעיות אלו מיתרגמות לצורך בכמות אינסופית של סיביות לשם ייצוג מלא של הפונקציה. הפתרון לבעיה הראשונה הינו דגימה מרחבית. מושג הדגימה מוכר לנו בהקשר לאותות חד-מימדיים, ואף ראינו שבמצבים מסוימים (כשהאות חסום תדר, ומרווח הדגימה נקבע בהתאם לכך) ניתן לקבוע כי דגימה אינה מאבדת מידע. גם על אותות דו-מימדיים ניתן ליישם דגימה, כשהפעם ישנו חופש גדול הרבה יותר לקביעת אופן הדגימה. ציור 1.2 ממחיש שתי צורות דגימה אפשריות, כשכל שמבדיל ביניהם הוא ייחוס הצירים. בשני המקרים מדובר בדגימה אחידה בשני הצירים, אך בסיבוב שונה ביחס לצירים הראשיים.



קיימות כמובן מגוון אופציות אחרות לדגימה שלא מוזכרות כאן. נושא הדגימה חשוב ואנו נקדיש לו דיון נפרד. בשלב זה נניח כי מטעמי נוחיות בחרנו לדגום את הפונקציה במרווח D בשני הצירים וכך קיבלנו את האות הדו-מימדי הדיסקרטי הבא:

$$f(x, y)|_{x=mD, y=nD} = f(mD, nD) = f[m, n]$$

אנו נסמן בסוגריים מלבניים אותות דיסקרטיים, ובסוגריים עגולים אותות מעל הרצף (כשאנו מאמצים, באופן זה, את סימוניו של פרופ' בועז פורת בסיפרו על עיבוד אותות ספרתיים). אם תחום ההגדרה של התמונה מלבני - $\Omega = [0, (M-1)D] \times [0, (N-1)D]$, אזי הפונקציה הדיסקרטית מוגדרת מעל הערכים

$$f : [m, n] \in [0, M-1] \times [0, N-1] \rightarrow \mathbb{R}^{[0,1]}$$

כך שהתמונה בנויה כמערך דו מימדי של N על M אלמנטים ממשיים בתחום 0 עד 1. כאשר מרווח הדגימה D נבחר להיות קטן, גודל התמונה גדול, ופרטים קטנים יותר ניתנים להכרה. במקרה זה אנו נאמר כי הרזולוציה של התמונה טובה יותר. רזולוציה מתייחסת לאו דווקא לגודל התמונה אלא ליכולת האבחנה בפרטים קטנים.

אם נחזור לציור 1.1, נתח תמונה מסוים על ערכיו מתוארים בציור 1.3. אנו רואים ערכים ממשיים כך שבפינה השמאלית העליונה הערך הנמוך ביותר, ועם ההתקדמות ימינה ולמטה, הערך גדל קלות.



נפתור כעת את הבעיה השנייה – במקום לייצג את רמות האפור כמספר ממשי בתחום $[0,1]$, נמיר אותן לייצוג בעל מספר סיביות קבוע וידוע מראש. לדוגמה, נוכל להחליט כי כל רמת אפור תיוצג ב- L

סיביות ע"י ההמרה - $\text{round}\left\{\left(2^L - 1\right)f[m,n]\right\}$. פעולה זו מעבירה כל ערך בתחום $[0,1]$ לערך שלם בתחום $[0, 2^L - 1]$. פעולה זו קרויה קוונטיזציה (בעברית - כימוי). ברור כי פעולת הקוונטיזציה מאבדת מידע! אבל, בזכות אובדן זה, כעת התמונה הינה מקבץ סופי של סיביות.

התמונה הדיגיטלית אותה בנינו ע"י דגימה וקוונטיזציה מיוצגת ע"י NM נקודות תמונה הקרויות פיקסלים (Pixel - קיצור של Picture Element), וכל אחד מהם מיוצג ע"י L סיביות. לכן, סך של MNL סיביות נחוץ לאחסונה של התמונה. גדלים אופייניים לגודל תמונה יהיו $N, M = 50$ לתמונה קטנה במיוחד, ועד $N, M = 10^5$ ואף יותר להדמאות לוויין ותצלומי אויר. באשר לעומק הקוונטיזציה, ערך אופייני מאוד הוא $L=8$, אך ישנם מקורות עם ערכי L גבוהים יותר. כך יוצא שנפח תמונה מינימליסטית הוא $2,500\text{Byte}$, תמונה בגודל סביר תצרוך $100\text{Kbytes} - 1\text{Mbyte}$, ותמונות חריגות עשויות להגיע עד 10Gbytes .

תיאור נפחי המידע הנ"ל מחזיר אותנו לאמירה מתחילת פרק זה על הקשר ההדוק בין ההתקדמות בעולם המיחשוב ו-"עיבוד תמונות". גם היכולת לאגור כמויות אלו, וכמובן היכולת לעבדן מציבים רף גבוה למחשבים. לדוגמה, גם פעולה פשוטה הכרוכה ב- 20 מכפלות לפיקסל בתמונה, הופכת להיות עצומה כאשר מדובר במליוני פיקסלים.

כל שנאמר עד כה התייחס לתמונות בשחור-לבן. אחת הדרכים המקובלות ביותר לייצוג של תמונות צבע היא מבנה RGB. כל נקודה (x,y) בתחום ההגדרה של התמונה תקבל לא ערך אחד, כמקודם, אלא שלושה ערכים המתייחסים לעוצמות האדום (Red), הירוק (Green) וכחול (Blue). ע"י צירופים בעוצמות שונות של שלושת צבעי יסוד אלה אנו מסוגלים לייצר מגוון רחב מאוד של גוונים. לכן, לאחר דגימה וקוונטיזציה, במקום מספר אחד בן L סיביות, תמונה צבעונית תדרוש שלושה ערכים כאלו, ולכן נפחה יהיה גדול פי שלוש. ישנם גם ייצוגים אחרים לתמונות צבע, וקיים דיון מעמיק בכלל בנושא צבע. במסגרת ספר זה נתמקד בתמונות שחור-לבן עם הצצות חטופות לצבע מפעם לפעם.

כנקודה אחרונה במסגרת סעיף זה נזכיר כי כל האמור לעיל מתייחס לתמונה בודדה - Still. מה באשר לסרטים? ובכן, סרט הינו פונקציה של המקום (x,y) אך גם של הזמן t . לכן, סרט היא פונקציה במימד 3 מהצורה $f(x,y,t)$. ממש כשם שדגמנו ב- (x,y) נוכל גם לדגום בציר הזמן, וכך יהפוך סרט רצף לסידרה ארוכה של תמונות Still. קצב הדגימה בציר הזמן נקבע לפי תכונות מערכת הראיה שלנו - מקובל כי הצגת 30 תמונות בשניה תיצור לעין אנושית (ולמוח שעובד מאחוריה) תחושה של רצף. לכן, למשל אם תמונה בודדה בסרט היא בגודל של 10Kbyte (תמונה קטנה של 100 על 100 פיקסלים עם 8 סיביות לפיקסל), הרי שסרט כזה באורך של שעה ידרוש נפח אחסון של כ- 1Gbyte . נפחים אלו הם הסיבה לעיכוב שחל ב-"עיבוד תמונה" בנושא הטיפול בוידאו. רק בעשור האחרון מעיזים להיכנס לנושא זה ולהציע כלים ואלגוריתמים לפעולות מורכבות בוידאו.

1.2 קלטופלט לתמונות דיגיטליות

תיארנו כיצד מיוצגת תמונה דיגיטלית במחשב, וראינו כי נחוצות פעולות דגימה וקוונטיזציה על מנת להביאה ממצב של פונקציה רצף לקובץ מחשב תקין. איזה התקנים מבצעים פעולות אלו? כיצד? באופן דומה, בהינתן תמונה כקובץ מחשב, נרצה לעיתים להתבונן בה, אם על מנת לקבל את תוצאת העיבוד, או בשל רצוננו לראות את התמונה במהלך שלבי עיבודה. גם כאן עולה השאלה כיצד עושים זאת? במסגרת סעיף זה נציג באופן שטחי מענה לשאלות הנ"ל. המטרה המרכזית היא להבין כיצד נוצרים מסלולים שלמים המתחילים בהרכשת תמונה, העברתה למחשב לשם עיבוד, ומסתיימים בהפקתה להתקן פלט.

מקור אפשרי ראשון של תמונה הוא מצלמה דיגיטלית, כגון מצלמות הווידאו הביתיות האופייניות, ומצלמות ה- Still החדישות בהן במקום סרט צילום ישנה מדיה דיגיטלית. מצלמות כאלו מורכבות סביב מטריצת חיישנים הרגישים לאור (כגון גלאי CCD), אשר קובעים מתחוזרם חשמלי באופן יחסי ישר לכמות הפוטונים הפוגעת בהם. כל חיישן מייצר ערך עבור פיקסל בודד, ולכן מטריצות חיישנים אלו צריכות להיות גדולות (מאות על מאות של חיישנים על chip בודד). המתחוזרם הנוצרים מומרים ישירות ל-L הסיביות המייצגות את רמת האפור בנקודה זו וכך מתקבל ערכו של הפיקסל. פרמטר מרכזי בקביעת איכותה של מצלמה כזו הוא מספר הגלאים במטריצת החיישנים. מצלמות Still דיגיטליות עם מערך של 480 על 640 נחשבות לרמה הבסיסית ביותר (זוהי רזולוציית VGA), ואיכותן רחוקה מלהתחרות במצלמת סרט קונונציונלית. לשם קבלת איכות מתחרה בסרט צילום, על המערך להכיל כ- 2000 על 2000 חיישנים ויותר.

כאלטרנטיבה למערך דו-מימדי של חיישנים, ניתן להציע מערך חד-מימדי וסריקה מכנית (של המערך, מראה מטילה או פריזמה) לקבלת אוסף שורות היוצרות את התמונה. גישה זו מקובלת כאשר נדרשת רזולוציה גבוהה מאוד ולפיכך מערך חיישנים ארוך (בסדר גודל של אלפים). הבעיה העיקרית עם גישה זו הוא הצורך במרכיבי מכניקה מדויקת במצלמה.

על מנת לקלוט תמונת צבע נדרש בעיקרון שימוש בשלושה מערכי חיישנים, כשכל אחד מהם מחובר למסנן אופטי לבצע הרלוונטי. כך ייצא למשל כי שכבת ה-R תיוצר ע"י מערך גלאים עם פילטר אופטי המונע מכל צבע למעט אדום לחדור. לכן, מצלמת צבע כזו תהיה בפועל שלוש מצלמות שונות על אותו ציר אופטי לשם קבלת התאמה מוחלטת ביניהן במובן הגיאומטרי.

כל האמור לעיל מתייחס לדרך בה פועלת מצלמה דיגיטלית אופיינית הן לתמונות Still והן לוידאו (סדרות של תמונות ברצף). חשוב לציין כי מצלמה אלקטרונית אינה מוגבלת לחישה של אור נראה – בהינתן גלאים החשים כל תופעה פיזיקלית אחרת וממירים אותה לאותות חשמליים, ניתן לסדרם במערך דו-מימדי ולהגדיר את תוצאת החישה כתמונה דיגיטלית לכל דבר. כך הדבר בצילום אינפרה-אדום, המספק תמונות של אורך גל אחר הקשור לחום של עצמים, כך גם בצילום רנטגן, צילומים סיסמיים של רעידות על-פני כדור הארץ, ועוד.

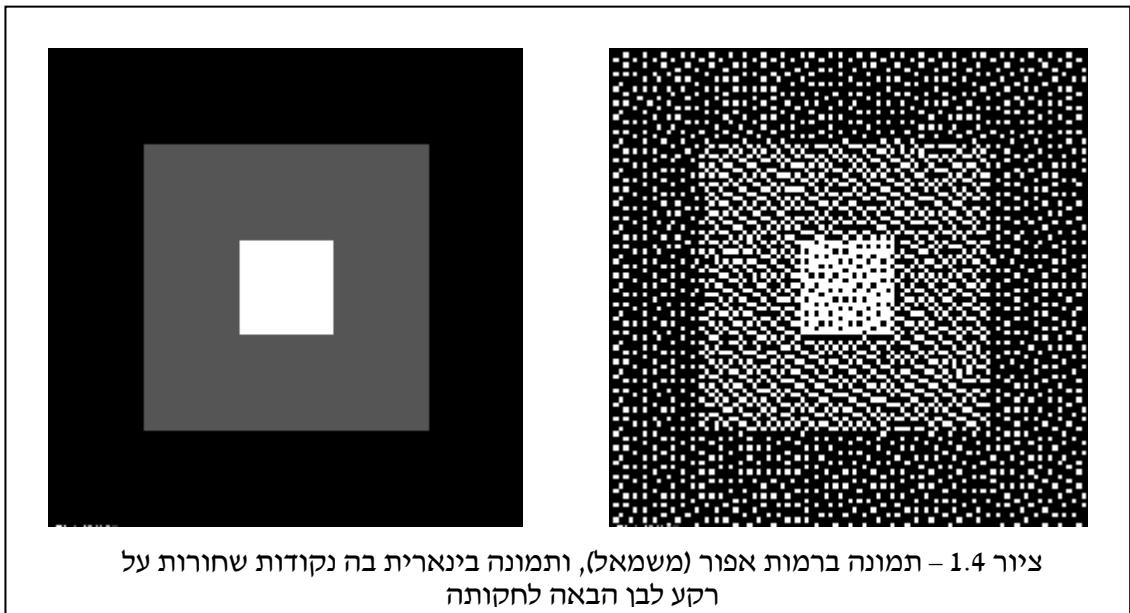
מקור מקובל אחר לתמונות הוא עותק קשה, דהיינו תצלום על נייר או על סרט צילום. הכלי המקובל לשם המרתה של תמונה כזו לקובץ הוא סורק (Scanner). סורק אינו אלא מצלמה דיגיטלית הפועלת בתנאים נוחים, ולכן מצלמה זו יכולה להיות פשוטה. התנאים הנוחים מתייחסים לכך שהעצם המצולם אינו זו, ולכן אין סיבה להשתמש במערך דו-מימדי יקר – די במערך חד-מימדי הנע מצד אחד של התמונה לצידה האחר. כמו כן, אין צורך ל-"לכוד" את התמונה בחטף, וניתן לעשות זאת בקצב קריאת המידע ע"י המחשב. הצילום נעשה בתנאי תאורה מבוקרים ולכן אין צורך בעיבודים אלקטרוניים מקדימים של תיקון תחום דינאמי. כמובן שאם הסריקה היא של סרט צילום בתשליל (Negative), יש להפוך את היוצרות ולהמיר את השחור לערך גבוה ואת הלבן לערך נמוך.

פרמטר מרכזי באפיון סורק היא רזולוציית הסריקה שלו, דהיינו המרווח המינימלי אשר הוא יכול לקבוע בין דגימות. סורקים אופייניים לשימושים ביתיים מספקים רזולוציה של עד 600 dpi, דהיינו 600 נקודות באינטש' אחד. לא פעם בפרסומים על סורקים מוזכרות רזולוציות בסדר גודל של 9600 dpi – חשוב להבהיר כי לרוב ערכים אלו מתייחסים לרזולוציה מלאכותית. התמונה נסרקת בפועל ברזולוציה של 600 dpi, ואז מוגדלת הלאה ע"י אינטרפולציה לגודל הנקוב. אינטרפולציה אינה תחליף לרזולוציית אמת. אנו נעמוד על ההבדלים בדיון בנושא דגימה, ולאחר מכן שכנדון בסוגיית הגדלת תמונה.

מקור שלישי ואחרון שיוזכר כאן לתמונות הוא תמונות מחושבות, או תמונות סינטטיות. צילום מכ"ם (SAR), צילום אולטראסאונד, צילום MRI וצילום CT – כל אלה יוצרים תמונות. תמונות אלו נוצרות ע"י מדידת תופעה פיזיקלית כלשהי לאו דווקא במערך דו-מימדי. עיבוד ספציפי של האות המתקבל יוצר תמונה מלאכותית אשר חשובה ליישומים רבים. לדוגמה, בצילום MRI התמונה הנוצרת מתארת חתך (שלא ניתן היה לקבלו בדרכים נעימות) בגופו של המצולם, כשרמות האפור קשורות לצפיפות המים בכל רקמה. גם לו חתכנו את הגוף המצולם באותו מקום בדיוק, לא היינו מקבלים את אותה תמונה. כדוגמה שניה, בצילום SAR, החזרים מכ"ם מייים מתועדים כתמונה. מוט ברזל בשטח יתבטא בנקודה לבנה חזקה, ולעומתו, גם אם נפרוס בד ענק בצבע לבן בשטח, לא נראה אותו כלל.

עד כה סקרנו אפשרויות להזנת תמונה אל המחשב. כעת נדון בקצרה בהתקני הפלט – המרת קובץ תמונה לתצוגה. האפשרויות העיקריות הינן הצגה למסך או הדפסה לנייר. באשר לשימוש במסך, הגישה האופיינית היא שימוש ב-CRT (קיימות גם אפשרויות אחרות כגון מסכי LCD) בהם קרן אלקטרוניים פוגעת במשטח הממיר פגיעת אלקטרוניים בפליטת פוטונים, תכונה קרויה פלואוריצנטיות. ע"י שליטה אלקטרונית בעוצמת קרן האלקטרוניים ומיקום פגיעתה במסך, מתקבלת נקודה בודדה בעוצמה נשלטת – זוהי רמת האפור הנדרשת. המסך עצמו מכוסה במערך דו-מימדי של נקודות חומר פלואורצנטי בצפיפות גבוהה הנותנת את התמונה. חשיבות רבה יש לשני פרמטרים במסך מעין זה – מהירות סריקת הקרן, אשר מיתרגמת למהירות בנייתה של התמונה על המסך, ורזולוציית המסך – המרווח בין נקודות סמוכות (פיקסלים) על פני המסך. הצגת תמונת צבע נעשית ע"י שימוש בשלושה צבעי חומר פלואורצנטי (R, G, ו-B) בשלושת של נקודות סמוכות, ופגיעה בכל אחד מאלו בעוצמה הנכונה לשם קבלת שילוב הצבע הנחוץ.

ובאשר להדפסה, הדפסה איכותית של תמונה פירושה המרת מטריצת הפיקסלים למגוון כל הגוונים הנחוץ על פני נייר ההדפסה. קיימות מדפסות Continuous-Tone אשר עושות פעולה זו בדיוק, אך השימוש בהן נדיר ויקר. מדפסות אלו פועלות כמו פיתוח על נייר צבע בצילום ביתי אופייני – חשיפת נייר רגיש לאור במידה מבוקרת בכל נקודה. כתחליף לגישה יקרה זו המחייבת לא פעם נייר פיתוח, מדפסות אופייניות (InkJet ו-Laser) מציעות גישה שונה לפיה רק מקבץ מצומצם מאוד של צבעים וגוונים ניתן להפקה. על מנת ליצור את התמונה הנתונה, נעשה שימוש בגישה ה-Half-Tone, אשר יוצרת אשליה של מגוון כל הצבעים וגוונים ע"י פיזור של נקודות בגודל, פיזור וצבע (מתוך המקבץ המצומצם האפשרי) נשלטים. ציור 1.4 ממחיש גישה זו ע"י דוגמה. עבור מדפסת היכולה לצבוע נקודות בשחור בלבד בגודל נקודה ובמיקום נשלט, כך נראית ההמרה מתמונת רמת אפור לתמונה בינארית – צפיפות גדולה לרמות נמוכות ונמוכה לרמות גבוהות. גישה זו נהוגה במדפסות ביתיות, בתמונות עיתון ועוד.



ביצוע Halftoning לצבע מורכב יותר, אך בנוי על אותו עיקרון. אנו נקדיש דיון לשיטות Half-Tone בהמשך.

1.3 סוגיות בטיפול בתמונות

כבר מהדיון על בנייתה של תמונה כקובץ ע"י דגימה וקוונטיזציה, והפקתה לעותק קשה ע"י Halftoning, פגשנו מספר אפליקציות המחייבות מענה ונדונות במסגרת עיבוד תמונות. בנוסף לנושאים אלו, בהינתן תמונה דיגיטלית, ניתן לדון במגוון רחב של בעיות מעניינות אחרות ודרכים לפתרונן.

בשלב ראשון, ניתן להתייחס לתמונה כאות דו-מימדי, ולהרחיב את שלמדנו במסגרת עיבוד אותות לדו-מימד. באופן זה נוכל לשאול כיצד להגדיר ולבצע התמרת פוריה דו-מימדית (רציפה ודיסקרטית, כמו באותות חד-מימדיים), כיצד לבצע סינון לתמונה, וכיצד לתכנן מסננים להגשמת יעדים

מוגדרים מראש על התמונה. גם נושא הדגימה הדו-מימדית נופל למסגרת זו של עיבוד אותות, וכפי שראינו, עולות אפשרויות מעניינות בהרחבת משפט הדגימה לדו-מימד. חשוב עם זאת להבהיר כי עיבוד אותות דו-מימדיים אינו "עיבוד תמונה". "עיבוד תמונות" כולל בתוכו הרבה יותר מאשר עיבוד אותות גרידה - אנו נפגוש המון בעיות להן אין מענה מסגרת עיבוד אותות, ויידרש טיפול ספציפי הנשען על פרקי מחקר אחרים.

בעיבוד תמונות מקובל להגדיר מספר משימות יסוד בהם משפחות שונות של בעיות ואלגוריתמים.

נושא ראשון בדיון על עיבוד תמונות הוא הפרק הדין בשיפור תמונה. לא אחת מתקבלת תמונה ירודה באיכותה בשל תנאי צילום לא מוצלחים, בשל פעולות לא נכונות בעת המרתה לקובץ, או בשל תכונות בעייתיות של העצם המצולם. במקרים אלו נרצה להציע אלגוריתם המשפר את איכות התמונה ויוצר תמונת פלט טובה יותר באיזה שהוא מובן. הנקודה הבעייתית בהקשר זה היא קביעת מדד לאיכות תמונה. מגוון כלים רחב מאוד פותח במהלך השנים כדי לתת מענה לבעיות שונות בשיפור תמונות. בדיון על שיפור תמונה נעשית אבחנה בין צורך לשפר תמונה שגורם קלקולה אינו מוגדר, ובין בעיות שחזור בהן ידוע מה עברה התמונה בתהליך הקלקול, ונדרשת מערכת ההופכת את התהליך במידת האפשר. אנו נקדיש מספר פרקים לדיון בנושאים אלו, ונסקור את הגישות הרווחות.

נושא שני בעיבוד תמונה הוא הפרק הדין בדחיסת תמונות. מהדיון עד כה ברור כי תמונות הן צרכן זיכרון תובעני. שאלה טבעית שעולה לפיכך היא האם ניתן לייצג תמונה ללא פגיעה באיכותה בפחות סיביות. בדחיסת מידע בכלל ובדחיסת תמונות בפרט מבחינים בין דחיסה ללא אובדן, בה לאחר פריסה חוזרים בדיוק למקור, ובין דחיסה אובדנית, אשר בפריסה חוזרת קרוב אך לא בדיוק למקור. אנו נראה כי ניתן בקלות לדחוס תמונה לכדי עשירית ופחות מנפחה המקורי ללא שינויים ניכרים לעין. כל הדיון בדחיסה נסמך על תורת האינפורמציה – זהו פרק חדש יחסית בתקשורת ספרתית אשר, מטבע הדברים, ייסקר בשטחיות רבה במסגרת ספר זה.

נושא שלישי בעיבוד תמונה דן בייצוגה של תמונה. פגשנו כבר נושא אחד מדיון זה – נושא ההמרה ל-Half-Tone. נושא הקוונטיזציה חיוני הן לבנייתה של תמונה דיגיטלית והן לצורך דחיסת תמונה – נושאים אלו ואחרים ייסקרו במסגרת ספר זה.

בכל הפעולות עד כה קיבלנו קובץ תמונה, ביצענו עיבוד, וקיבלנו קובץ תמונה כפלט (בדחיסה, אנו מתייחסים למערכת שדוחסת ולאחר מכן פורסת). לא אחת נרצה להכניס תמונה לעיבוד שמוצאו אינו תמונה, אלא החלטה. למשל – האם יש אנשים בתמונה הנתונה, ואם כן כמה והיכן? גלה את מיקומם של הקווים הישרים בתמונה וכיו. סוגיות אלו בעיבוד תמונה משויכות לתחום הקרוי ראייה ממוחשבת, ואנו נרחיב על האבחנה בין התחומים הללו בסעיף הבא.

נושא אחרון שיידון במסגרת ספר זה הוא הטיפול בסדרות של תמונות (סרטי וידאו). גם בסדרות של תמונות מדברים על שיפור, שחזור, דחיסה, ייצוג, ופעולות ראייה ממוחשבת. עניין ייחודי ומרתק שעולה בעיבוד סדרות הוא נושא התנועה בין תמונות, ושימוש בו לשם פתרון הבעיות הנ"ל בטיב גבוה

יותר, בהשוואה לטיפול פרטני בתמונות בודדות. אנו נקדיש פרק אחד לנושא זה, וניגע באופן זה בקצהו של קרחון של תוצאות מחקריות ואלגוריתמים.

חשוב להבהיר כי ספר זה רחוק מלכסות את מגוון הנושאים בעיבוד תמונות. המטרה היא סקירה ראשונית, ומתן הפנייה לעבודות המשך בתחום. עם זאת מושם דגש על עדכניות החומר, והבאת תוצאות חדשניות ככל האפשר.

1.4 עיבוד תמונה ותחומים סמוכים

הטיפול הממוחשב בתמונות אינו נחלתו הבלעדית של התחום הקרוי "עיבוד תמונות". קיימים פרקים סמוכים אשר אף להם זיקה ישירה לעיבוד ספרתי של תמונות. חלוקה מקובלת כבר נרמזה בסעיף קודם, בו עמדנו על האבחנה בין עיבוד תמונה וראייה ממוחשבת. בעוד שבעיבוד תמונה, הקלט לאלגוריתם והפלט הם תמונות, הרי שראייה ממוחשבת דנה באלגוריתמים בהם הקלט הוא תמונה והפלט הינו מידע אחר.

בתוך תחום הראייה הממוחשבת מקובל להגדיר משימות ברמה נמוכה (Low-Level Vision), בינונית (Mid-Level Vision) וגבוהה (High-Level Vision). ראייה ממוחשבת ברמה נמוכה דנה בהפקת מסקנות ראשוניות מתמונה בצורה של מידע לא מעובד וחסר משמעות פיזיקלית ישירה. כך הדבר בהפקת וקטורי תנועה בין צמד תמונות, מיקום שפות או צורות פרימיטיביות אחרות כגון ישרים, פינות וכו'. בראייה ממוחשבת ברמה בינונית אלגוריתמים מספקים מידע בעל ערך מתקדם יותר בעל ערך פיזיקלי, כגון מפת עומק תלת-מימדית מצמד תמונות, או תיאור של תנועת המצלמה במרחב בהתבסס על התמונות הנתונות. בראייה ממוחשבת ברמה גבוהה האלגוריתמים מסיקים מסקנות דמויות אנוש כגון בתמונה זו רואים שולחן, בתמונה זו ישנם שלושה אנשים וכדומה.

בתוך הפרק הרחב הקרוי ראייה ממוחשבת מצוי פרק העוסק בזיהוי צורות – Pattern Recognition. פרק זה דן בשאלה כיצד מוצאים עצם מסויים בתמונה באופן יעיל ואמין. יש לזכור כי תמונה מהווה היטל מעולם תלת-מימדי למישור דו-מימדי, ובכך ישנו אובדן רב. בשל כך, גם משימות אלמנטריות לכאורה כגון זיהוי פניו של אדם מתמונה, תחת האפשרות שהוא פונה קדימה או לצדדים, הינה בעיה קשה מאוד. פרק זה של זיהוי צורות שואב מתיאוריות רחבות יותר כגון "למידה חישובית" (Computational Learning), רשתות נוירונים, בינה מלאכותית (Artificial Intelligence) ועוד.

קיימים יישומים רבים לראייה ממוחשבת ולזיהוי צורות, וניכרת חדירה חזקה של תחום זה לתעשייה ולמוצרים בעשור האחרון. נזכיר כמה דוגמאות כשהן מחולקות לקטגוריות. בתחום הקרוי זיהוי תווים ניתן להזכיר קריאה אוטומטית של טפסים, איתור וקריאת בר-קוד, קריאה ממוכנת של צ'קים, המרת דף כתוב (מודפס או כתב יד) לקובץ ASCII, ומיון אוטומטי של מכתבים עפ"י קריאת המיקוד. בטיפול בתמונות רפואיות נזכיר ספירת תאי דם, ספירת תאי זרע ומציאת פגמים בהם, ואיתור גידולים בתמונות רפואיות. באוטומציה תעשייתית ורובוטיקה נזכיר בקרת איכות למעגלים מודפסים, זיהוי חלקים בפס ייצור ומיונם, בקרת תנועה ע"י משוב חזותי, איתור מכשולים בתנועה ועוד. במיפוי

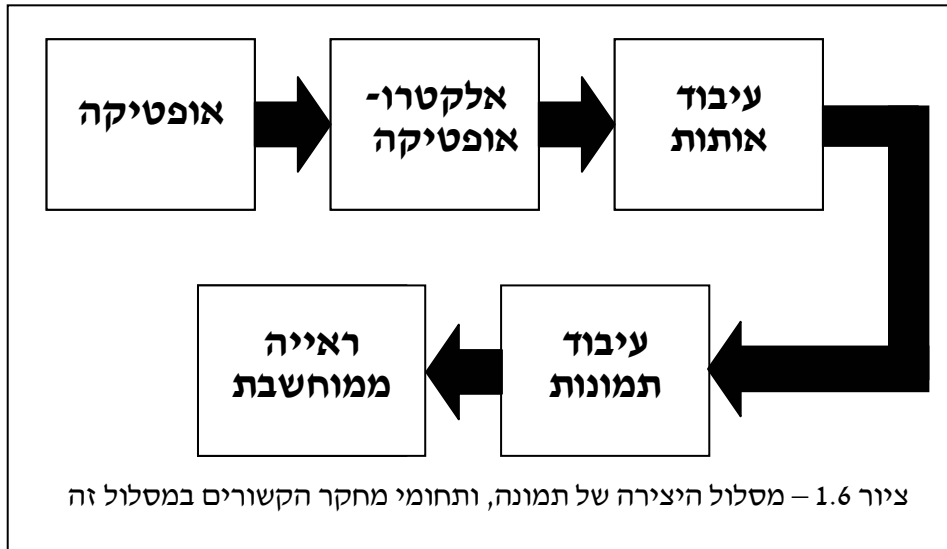
וחישה מרחוק (צילום ע"י מטוס/לוויין) ישנם עריכת מפות מתוך תצלום אוויר, קישור תמונות שונות וואו מפות זו לזו מבחינה גיאומטרית, גילוי תנועה ושינויים בשטח, איתור מטרות ועוד. בזיהוי פלילי דנים בזיהוי טביעת אצבע תוך יחוס למאגר קיים, זיהוי חתימה, זיהוי פנים ועוד.

בתחילת סעיף זה אמרנו ש- "עיבוד תמונה" מתמקדת במשימות בהן הן הכניסה והן היציאה הם תמונות. "ראיה ממוחשבת" לעומת זאת דנה במשימות בהן הכניסה היא תמונה והמוצא אינו תמונה. בהמשך טבעי להגדרות הנ"ל, התחום בו הקלט אינו תמונה אלא מידע כלשהו, והפלט הוא תמונה, קרוי גרפיקה ממוחשבת. במידה רבה קיימת דואליות בין ראייה ממוחשבת וגרפיקה ממוחשבת, ולהרבה בעיות שניתן להציג בפרק האחד, ניתן להציג בעיות דואליות בפרק השני. למשל, אם בראיה ממוחשבת מתוך תמונות נרצה חילוץ של מפת התלת-מימד של העצמים המעורבים, הרי שבגרפיקה נשאל כיצד מגופים תלת-מימדיים בייצוג כלשהו בונים תמונה ריאליסטית (Rendering). ציור 1.5 ממחיש את המפגש בין עיבוד תמונות, ראייה ממוחשבת, וגרפיקה ממוחשבת.



ניתן גם להציע חלוקה אחרת של פרקים הקשורים בעיבוד תמונות, המבוססת על תהליך עיבוד של תמונה מרגע יצירתה ועד סיום הטיפול בה במערכת דמוית ראיית אנוש. בשלב ראשון מצלמה (העין) אמורה לרכוש את התמונה, וזאת ע"י אופטיקה מתאימה (העדשה בעין), לאחר מכן התקנים אלקטרו-אופטיים (אותם גלאי CCD למשל במצלמה, או החיישנים בעין הקרויים Rods ו-Cones), ובסיום עיבוד אותות אלמנטרי של תיקון תחום דינאמי, דגימה וקוונטיזציה. ואמנם, לאופטיקה, אלקטרו-אופטיקה ועיבוד אותות קשר ישיר וחשוב לעיבוד תמונות (שלבם אלו מתבצעים ברטינה בעין).

לאחר בנייתה של התמונה, יבואו אלגוריתמי דחיסה ושיפור איכות השייכים ל- Main Stream בעיבוד תמונות (פעולות אלו מבוצעים בדרך מהעין למוח, ובמוח הראיה עצמו). אחריהם תתחיל פעולת ניתוח (ראייה ממוחשבת) לשם הפקת מידע מהתמונות, כגון זיהוי האנשים, אבחנה בתנועה וכו'. כך אנו רואים מעבר מאופטיקה ועד ראייה ממוחשבת ולמעשה בינה מלאכותית. ציור 1.6 סוקר את המסלול הנ"ל על הפרקים אותם הוא רותם.



1.5 יישומים לעיבוד ואנליזה של תמונות

לסיום פרק זה רצוננו להמחיש את חשיבותו ומעשיותו של התחום הקרוי "עיבוד תמונות". זאת נעשה ע"י מספר דוגמאות למוצרים ומערכות בהן משולבים אלגוריתמים לעיבוד תמונה, וראיה ממוחשבת.

דוגמה 1 – וידאו-פון ושיחה ע"י Net-Meeting: הרעיון ששיחת טלפון תלווה בצילום הדדי, כך שהדוברים ישוחחו זה עם זה, מלהיבה חברות ומהנדסים קרוב לעשרים שנה, אך לאחרונה חל מפנה משמעותי בתחום זה, וניתן לראות מערכות של ממש המיישמות רעיון זה. הסיבה לעיכוב עד כה היה רוחב הפס הנדרש להעברת אות וידאו. במילים אחרות, כפי שראינו כבר בפרק זה, כמות הסיביות הנדרשת לתיאור סדרת תמונות היא עצומה, ולא מעשית בערוצי תקשורת טלפוניים רגילים. לדוגמה, 30 תמונות צבע בגודל של 100 על 100 פיקסלים ידרשו שידור של קרוב ל-1Mbyte לשניה, וזה עבור תמונה קטנה מאוד. ערוץ טלפון נדיב לעומת זאת ייתן 64Kbits לשניה. ההתפתחויות האחרונות בנושא דחיסת וידאו אפשרו להמיר את הקצב המקורי בקצב הטלפון האופייני, ולהשיג דחיסה ביחס של מעל 100, תוך שהסרט הנפרס משמר הרבה מאיכותו המקורית. כיום ניתן לשוחח זה עם זה מעל פני האינטרנט תוך שימוש ברעיון בסיסי זה, כשהוא מלווה בפתרון בעיות מורכבות בנושא של תזמון חבילות השידור לשם השגת שידור אפקטיבי בזמן אמת.

דוגמה 2 – מצלמה דיגיטלית ואלבום תמונות משפחתי ממוחשב: חברות רבות מציעות כיום מצלמה דיגיטלית כתחליף למצלמת היד הקונבנציונלית. גם במקרה זה, מדובר במוצר שעוכב בשל מגבלות טכנולוגיות. עד לאחרונה מקובל היה לטעון שמערכי CCD בגדלים של מעל 1000 על 1000 יקרים מידי. גם נושא המדיה השומרת את התמונות זכה להתקדמות טכנולוגית בשל היכולת לדחוס את התמונות. כיום, מוצעות מצלמות ברזולוציה המתקרבת לביצועי סרט צילום רגיל. הרעיון הוא לצלם ישירות לקובץ הנשמר כשהוא דחוס, ולכן מאפשר שמירה של 10 עד 200 תמונות (תלוי ביצרן, ובעוד גורמים). כל מצלמה כזו כוללת בתוכה סידרת תהליכים בהם התמונה מעובדת לתיקון הצבעים, תיקון ההארה ועוד – אלגוריתמים קונבנציונליים המבוצעים בחומרה על ASIC.

אחת המגבלות בשימוש במצלמות יד דיגיטליות לתצלומי Still הוא הצורך בהדפסת התמונות. עד לפני כ- 5 שנים מקובל היה שהדפסה באיכות טובה כדוגמת פיתוח מסרט לנייר פיתוח היא בלתי אפשרית. כיום בשל ההתקדמות במדפסות הזרקת דיו, ניתן להפיק תמונות באיכות טובה על נייר רגיל. במקביל, בשל היות התמונות ספרתיות, מוצע לנהלן בתוכנות לאלבומים וירטואליים. באופן זה ניתן לבצע חיפושים חכמים באלבום כגון – מצא את דוד בכל התמונות באלבום, הראה לי את כל התמונות בהן דודה נחמה עומדת ליד דוד משה, וכדומה. כל הרעיון הוא שתוכנת האלבום תבצע חיפוש מבוסס תוכן בתמונות, ולא על-פי מאפייני טקסט המלווים כל תמונה. אפליקציות מתוחכמות לחיפושים מבוססי תוכן מוצעות בעת האחרונה גם בהקשר של חיפוש תמונות ב-WEB.

דוגמה 3 – מדפסות: כבר הזכרנו שכל מדפסת מכילה בתוכה אלגוריתם להמרת תמונה מערכי הרצף - Continuous Tone – ל-Half-Tone. במדפסות סטנדרטיות גם משולבים אלגוריתמים אחרים לתיקון צבעים בתמונה, למתיחת תחום דינאמי ועוד.

דוגמה 4 – גילוי פגמים בפס ייצור: נתחיל בדוגמה פשוטה – פס ייצור נוסע ועליו תפוזים. על התפוזים להיכנס לארגזים כשהם מסווגים לפי גודלם. מצלמת וידאו, המוצבת מעל פס הייצור ומחוברת למחשב PC רגיל ובו כרטיס דגימה, ותוכנת זמן אמת לעיבוד התמונה הנקלטת, יכולים לתת בקרה על פעולת הפס באופן שסיווג התפוזים ייעשה באופן אוטומטי לחלוטין. תפיסה זו חלה מסתבר לא רק על תפוזים. תעשיות הכרטיסים המודפסים (PCB) ותעשיית המיקרו-אלקטרוניקה (ייצור מעגלים על שבבי סיליקון) כבר מזמן עושות שימוש בבקרה מבוססת ראייה ממוחשבת לשם פיקוח על פגמים בייצור.

דוגמה 5 – טלוויזיה דיגיטלית: כל עוד אין מוצרים ביתיים המשלבים עיבוד תמונה באופן המוני, הרי שניתן לקבוע כי תחום זה אינו ממצה את הפוטנציאל שתולים בו. מגבלה זו עומדת לפוג ברגע שבכל בית תהיה טלוויזיה דיגיטלית. טלוויזיה כזו פירושה שהעברת השידורים נעשה באופן דיגיטלי, ולכן דחיסה ופריסה מעורבים בתהליך. יתרה מזו – ברגע שהתמונות מגיעות באופן דיגיטלי אין מניעה מלבצע עליהן עיבוד מתקדם. למשל, שיפור של איכות שידור במקרים בעייתיים, המרת התמונות לשידור סטריאוסקופי שיאפשר ראיית עומק, הקפאת תמונה, הגדלתה והדפסתה, אחסון סרטים במדיה דיגיטלית, ועוד.

פרק 1 - כללי

לרשימה הנ"ל ניתן להוסיף עוד יישומים כגון עריכת וידאו ואפקטים, שירותי אינטרנט לשיפור תצלומים מהאלבום המשפחתי, שחזור תמונות מקולקלות כתוצאה מקלקול המצלמה, ועוד. כמו כן הזכרנו יישומים רבים בהקשר לראיה ממוחשבת בסעיף קודם.

פרק 2 – עיבוד אותות בדו-מימד

15	פרק 2 – עיבוד אותות בדו-מימד
15	2.1 מערכות ליניאריות בדו-מימד - אותות מעל הרצף
19	2.2 התמרת פוריה דו-מימדית ותכונותיה
23	2.3 מערכות ליניאריות בדו-מימד – אותות דיסקרטיים
25	2.4 התמרת פוריה לאותות דיסקרטיים ותכונותיה
29	2.5 אותות ומערכות ליניאריות בתמך סופי
39	2.6 התמרת פוריה דיסקרטית (DFT) לאותות בתמך סופי

הזכרנו בפרק הקודם שעבוד תמונות אינו רק עיבוד אותות דו-מימדי. עם זאת, רבים ממושגי היסוד בעיבוד תמונות, כמו מסננים, מערכות ליניאריות ושאינן ליניאריות, התמרת פוריה ועוד, מקורם בעיבוד אותות. בפרק זה נתמקד בהרחבת המוכר לנו מעיבוד אותות חד-מימדיים לדו-מימד. בבסיס פרק זה ישנה ההנחה כי קיים לקורא רקע בעיבוד אותות חד-מימדיים.

2.1 מערכות ליניאריות בדו-מימד – אותות מעל הרצף

עבור אות דו-מימדי מעל הרצף, $f(x,y)$, המוזן למערכת H , אנו מגדירים את מוצא המערכת ע"י

$$g(x, y) = H\{f(x, y)\}$$

מערכת תיקרא ליניארית אם ורק אם מתקיים

$$H\{\alpha f_1(x, y) + \beta f_2(x, y)\} = \alpha H\{f_1(x, y)\} + \beta H\{f_2(x, y)\}$$

לכל צמד אותות $f_1(x, y), f_2(x, y)$, ופרמטרים α, β . תכונה זו מוכרת לנו היטב ממסגרת של מערכות חד-מימדיות.

מערכת H תיקרא חסרת זיכרון אם מוצאה במקום (x,y) תלוי אך ורק בערכו של אות הכניסה f במקום (x,y) . בעיבוד אותות חד-מימדיים, המשך טבעי למושג הזכרון הוא הסיבתיות. עם זאת, סיבתיות אינה טבעית בדיון על אותות דו-מימדיים, כיוון שלא קיים סדר בין נקודות במישור. אם נגדיר מראש סדר סריקה למישור לפי ציר זמן (כגון, סריקה שורה אחר שורה), ניתן לדבר על סיבתיות, אך זו תהיה רלוונטית רק לסריקה הנדונה. במקרה כזה בפועל המרנו את התמונה לאות חד-מימדי, ואז מושג הסיבתיות מתלכד עם סיבתיות חד-מימדית.

מערכת H תיקרא קבועה במקום אם תוצאת הזזת האות במקום בשיעור (X_0, Y_0) והזנתו למערכת מניבה תוצאה זהה להזנת האות המקורי והזזת מוצא המערכת בשיעור דומה, דהיינו

$$H\{f(x,y)\} = g(x,y) \Rightarrow H\{f(x-X_0, y-Y_0)\} = g(x-X_0, y-Y_0)$$

דוגמה 2.1: נציג מספר דוגמאות להמחשת תכונת הליניאריות, זיכרון, וקביעות במקום. בכל המקרים המתוארים ניתן להוכיח את הטענות השונות ישירות מתוך ההגדרות:

א. המערכת $H\{f(x,y)\} = f(x,y) + 3f(x-2, y+5)$ היא מערכת ליניארית. זוהי מערכת קבועה במקום, עם זיכרון.

ב. המערכת $H\{f(x,y)\} = f(x,y) + g(x,y)f(x-2, y+5)$ (עבור $g(x,y)$ קבוע וידוע מראש) היא ליניארית. זו מערכת שאינה קבועה במקום, וגם לה יש זיכרון.

ג. המערכת $H\{f(x,y)\} = f(x,y) \cdot f(x-2, y+5)$ אינה ליניארית. זו מערכת עם זיכרון והיא קבועה במקום.

ד. המערכת $H\{f(x,y)\} = f(x,y) + 1$ אינה ליניארית למרבה ההפתעה. זו מערכת חסרת זיכרון הקבועה במקום.

ה. המערכת $H\{f(x,y)\} = \text{Log}\{f(x,y)\}$ אינה ליניארית. זו מערכת חסרת זיכרון הקבועה במקום.

ו. המערכת $H\{f(x,y)\} = g(x,y) \cdot \text{Log}\{f(x,y)\}$ אינה ליניארית. בנוסף, זו מערכת תלוית מקום וחסרת זיכרון.

מערכת ליניארית H המוזנת בפונקציה הלם דו-מימדית $\delta(x-X_0, y-Y_0)$ מניבה במוצאה את פונקציה התגובה להלם $h(x,y, X_0, Y_0)$. במקרה הכללי של מערכת ליניארית, התגובה להלם היא פונקציה בארבעה משתנים כיוון שלכל כניסה דו-מימדית יש מוצא דו-מימדי (ולכן שני המשתנים הראשונים), ופונקציה זו עשויה להיות תלוית מקומו הדו-מימדי של ההלם המוזן למערכת (X_0, Y_0) .

מסתבר כי כל מערכת ליניארית מאופיינית לחלוטין ע"י תגובתה להלם, h . דהיינו, בהינתן הפונקציה $h(x,y, X_0, Y_0)$, נוכל לקבוע מהו מוצאה של המערכת H לכל כניסה שהיא. תכונה זו רואים מתוך כך שכל אות $f(x,y)$ ניתן לתיאור כצירוף ליניארי של פונקציות הלם ע"י

$$f(x, y) = \int_{\alpha} \int_{\beta} f(\alpha, \beta) \cdot \delta(x - \alpha, y - \beta) d\alpha d\beta$$

בשל ליניאריות המערכת H, מתקבל מוצאה לאות f(x,y) נתון ע"י האינטגרל הבא אשר זכה לשם

קונבולוציה

$$\begin{aligned} H\{f(x, y)\} &= H\left\{ \int_{\alpha} \int_{\beta} f(\alpha, \beta) \cdot \delta(x - \alpha, y - \beta) d\alpha d\beta \right\} = \\ &= \int_{\alpha} \int_{\beta} f(\alpha, \beta) \cdot H\{\delta(x - \alpha, y - \beta)\} d\alpha d\beta = \\ &= \int_{\alpha} \int_{\beta} f(\alpha, \beta) \cdot h(x, y, \alpha, \beta) d\alpha d\beta \end{aligned}$$

על פי האמור לעיל, כאשר המערכת H קבועה במקום צריך להתקיים כי

$$H\{f(x, y)\} = g(x, y) = \int_{\alpha} \int_{\beta} f(\alpha, \beta) \cdot h(x, y, \alpha, \beta) d\alpha d\beta \Rightarrow$$

$$1. H\{f(x - X_0, y - Y_0)\} = \int_{\alpha} \int_{\beta} f(\alpha - X_0, \beta - Y_0) \cdot h(x, y, \alpha, \beta) d\alpha d\beta = g(x - X_0, y - Y_0)$$

$$2. g(x - X_0, y - Y_0) = \int_{\alpha} \int_{\beta} f(\alpha, \beta) \cdot h(x - X_0, y - Y_0, \alpha, \beta) d\alpha d\beta$$

והשוואת שתי המשוואות התחתונות מניבה לאחר אלגברה אלמנטרית

$$\begin{aligned} \forall x, y, \alpha, \beta, X_0, \& Y_0 : h(x - X_0, y - Y_0, \alpha, \beta) = h(x, y, \alpha + X_0, \beta + Y_0) \\ \Rightarrow h(x, y, \alpha, \beta) &= h(x - \alpha, y - \beta) \end{aligned}$$

משמעות תוצאה זו היא שהתגובה להלם אינה משנה את צורתה במקום אלא זזה באותו שיעור בו זז ההלם שהוזן למערכת. לכן, במקרה זה, תגובת ההלם היא פונקציה של שני משתנים בלבד. במקרה זה מוצא המערכת מתקבל ע"י הקשר

$$\begin{aligned} H\{f(x, y)\} &= \int_{\alpha} \int_{\beta} f(\alpha, \beta) \cdot h(x, y, \alpha, \beta) d\alpha d\beta = \\ &= \int_{\alpha} \int_{\beta} f(\alpha, \beta) \cdot h(x - \alpha, y - \beta) d\alpha d\beta = f(x, y) \otimes h(x, y) \end{aligned}$$

וזהו פעולת הקונבולוציה למקרה של נערכת קבועה במקום.

דוגמה 2.2: בהמשך לדוגמה הקודמת, נתייחס למספר מקרים ונקבע את התגובה להלם עבורם :

א. ראינו כי המערכת $H\{f(x, y)\} = f(x, y) + 3f(x - 2, y + 5)$ ליניארית וקבועה במקום. תגובתה להלם נתונה ע"י $h(x, y) = \delta(x, y) + 3\delta(x - 2, y + 5)$.

ב. המערכת $H\{f(x, y)\} = f(x, y) + g(x, y)f(x - 2, y + 5)$ ליניארית, אך אינה קבועה במקום. תגובתה להלם נתונה ע"י $h(x, y, \alpha, \beta) = \delta(x - \alpha, y - \beta) + g(x, y)\delta(x - 2 - \alpha, y + 5 - \beta)$, וכפי שאנו רואים, זו אינה פונקציה של ההפרשים $(x - \alpha)$ ו- $(y - \beta)$ בלבד.

כיצד מתבטאת מערכת חסרת זיכרון בתגובתה להלם? בהתבסס על הנאמר קודם, נובע כי על תגובת ההלם להיות $h(x, y, \alpha, \beta) = h(x, y)\delta(x - \alpha, y - \beta)$, ואז

$$\begin{aligned} H\{f(x, y)\} &= \int_{\alpha} \int_{\beta} f(\alpha, \beta) \cdot h(x, y, \alpha, \beta) d\alpha d\beta = \\ &= \int_{\alpha} \int_{\beta} f(\alpha, \beta) \cdot h(x, y) \delta(x - \alpha, y - \beta) d\alpha d\beta = f(x, y) h(x, y) = g(x, y) \end{aligned}$$

כלומר, המוצא מתקבל ע"י הכפלת הכניסה בסקלר כלשהו שעשוי להיות תלוי מקום. במקרה של מערכת קבועה במקום וחסרת זיכרון מקבלים כי המוצא הוא הכפלה בסקלר קבוע h של הכניסה.

מערכת ליניארית וקבועה במקום תיקרא פרידה (ספרבילית) אם מתקיים כי ניתן לפרק את תגובת ההלם הדו-מימדית למכפלה של שני פונקציות חד-מימדיות $h(x, y) = h_1(x)h_2(y)$. במקרה זה, מוצא המערכת נתון ע"י

$$\begin{aligned} H\{f(x, y)\} &= \int_{\alpha} \int_{\beta} f(\alpha, \beta) \cdot h(x - \alpha, y - \beta) d\alpha d\beta = \\ &= \int_{\alpha} \int_{\beta} f(\alpha, \beta) \cdot h_1(x - \alpha) h_2(y - \beta) d\alpha d\beta = \int_{\alpha} h_1(x - \alpha) \int_{\beta} f(\alpha, \beta) \cdot h_2(y - \beta) d\alpha d\beta \end{aligned}$$

ומשמעות תוצאה זו היא שיש לבצע קונבולוציה חד-מימדית על כל שורה בפונקציה $f(x, y)$, ועל התוצאה שוב קונבולוציה חד-מימדית, הפעם על כל עמודה. אם גם האות f פריד נקבל

$$H\{f(x, y)\} = \int_{\alpha} f_1(\alpha) h_1(x - \alpha) d\alpha \int_{\beta} f_2(\beta) \cdot h_2(y - \beta) d\beta$$

ופירוש הדבר שבפועל בוצעו שתי קונבולוציות חד-מימדיות.

שרשור של מערכות ליניאריות מניב מערכת ליניארית – זאת ניתן להוכיח לפי הגדרה. כאשר מדובר במערכות קבועות במקום, ניתן להחליף את סדר פעולתן ללא גרימת שינוי במוצא המערכת הכוללת. אנו נפגוש תכונה זו בהמשך, כשנדון במערכות דיסקרטיות עם תמך סופי, ונקשור אותה למכפלה מתחלפת של מטריצות.

2.2 התמרת פוריה דו-מימדית ותכונותיה

התמרת פוריה דו-מימדית לפונקציות מעל הרצף מוגדרת באופן הבא

$$F(u, v) = F\{f(x, y)\} = \int_x \int_y f(x, y) \exp\{-j2\pi(ux + vy)\} dx dy$$

המשתנים u ו- v הם תדרים מרחביים אופקי ואנכי. ההתמרה ההפוכה נתונה ע"י

$$f(x, y) = F^{-1}\{f(x, y)\} = \int_u \int_v F(u, v) \cdot \exp\{+j2\pi(ux + vy)\} dudv$$

מלכתחילה אנו מגדירים את ההתמרה (קדימה ואחורה) עבור אותות חסומי אנרגיה המצויים ב- L_2 , ואיננו מרשים בפונקציות הנדונות אי-רציפויות מהסוג הראשון, בהן ערך הפונקציה קופץ בנקודה אחת לערך אחר ומיד חוזר. הסיבה לכך היא שהמרחק בין שתי פונקציות זהות למעט סידרה (אפילו אינסופית אך ברת מניה) של אי-רציפויות מסוג זה הוא 0, כך שבמרחב הפונקציות בו אנו פועלים הן נדמות זהות לכל דבר.

בטבלה 2.1 מתוארות מספר פונקציות בסיסיות והתמרתן. ניכר דמיון רב בין התמרת פוריה החד-מימדית, וזו הנדונה כאן. משמעותה של הפונקציה $F(u, v)$ היא תיאור האות הדו-מימדי כהרכב של תדרים מרחביים. התדר המרחבי (u, v) מופיע באות הנדון בעוצמה $F(u, v)$. לכן, תמונה חלקה יחסית תורכב ממרכיבי תדר נמוכים בדרך כלל, ותמונה מרובת מעברים מהירים תתרכז התדרים גבוהים.

$f(x, y)$	$F\{f(x, y)\} F(u, v) =$
$\delta(x, y)$	1
$\delta(x - X_0, y - Y_0)$	$\exp\{-j2\pi(X_0u + Y_0v)\}$
$\exp\{j2\pi(xU_0 + yV_0)\}$	$\delta(u - U_0, v - V_0)$
$\exp\{-\pi(x^2 + y^2)\}$	$\exp\{-\pi(u^2 + v^2)\}$
$-1 \leq x, y \leq 1 = 1$ for $\text{rect}(x, y)$	$4 \sin(u) \sin(v) / uv = \text{Sinc}(u, v)$
$\text{tri}(x, y) = \text{rect}(x, y) \otimes \text{rect}(x, y)$	$\text{Sinc}^2(u, v)$
$\sum_k \sum_j \delta(x - k, y - j)$	$\sum_k \sum_j \delta(u - k, u - j)$

טבלה 2.1 – צמדים של פונקציות בדו-מימד והתמרת הפוריה שלהן

להתמרת פוריה הדו-מימדית מספר תכונות חשובות אשר נסקור כאן ללא הוכחה. מרבית תכונות אלה מוכרות לנו מהתמרת פוריה בחד-מימד.

1. **יחידות:** לכל פונקציה $f(x,y)$ התמרה אחת ויחידה $F(u,v)$, ובביצוע ההתמרה ההפוכה, חוזרים לאותה פונקציה בדיוק. אין אובדן אינפורמציה בהתמרה קדימה או אחורה. כל זאת בכפוף לנאמר קודם על מרחב הפונקציות בו אנו פועלים.

2. **ליניאריות:** התמרת פוריה היא ליניארית, ולכן הפעלתה על צירוף ליניארי של תמונות זהה להפעלתה על כל תמונה בנפרד וצירוף ליניארי של התוצאות

$$\forall \alpha, \beta, f_1, f_2 : F\{\alpha f_1(x, y) + \beta f_2(x, y)\} = \alpha F\{f_1(x, y)\} + \beta F\{f_2(x, y)\}$$

3. **התמרת תמונה צמודה:** הפעלת התמרת פוריה על תמונה צמודה מניבה התמרה צמודה ומהופכת בצירים, דהיינו

$$F\{f(x, y)\} = F(u, v) \Rightarrow F\{f^*(x, y)\} = F^*(-u, -v)$$

מתכונה זו עולה שעבור תמונה ממשיית מתקבל כי התמרת הפוריה מקיימת $F(u, v) = F^*(-u, -v)$.

4. **היפוך צירים:** התמרת פוריה לתמונה עם היפוך צירים מניבה תוצאה הפוכת צירים

$$F\{f(x, y)\} = F(u, v) \Rightarrow F\{f(-x, -y)\} = F(-u, -v)$$

5. **ספרביליות ההתמרה:** ההתמרה ניתנת ליישום בשני שלבים בהם פועלים בחד-מימד. חשוב להבין כי תכונה זו לא מניחה דבר על התמונה $f(x,y)$ עצמה. ניתן לראות זאת ישירות מההגדרה, לפי

$$F(u, v) = \int_x \left[\int_y f(x, y) \exp\{-j2\pi v y\} dy \right] \exp\{-j2\pi u x\} dx$$

בשל התכונה הנ"ל, כאשר אות הכניסה פריד מתקבל כי ההתמרה אף היא פרידה, והינה מכפלת שתי התמרות פוריה חד-מימדיות

$$F(u, v) = \left[\int_x f_1(x) \exp\{-j2\pi u x\} dx \right] \left[\int_y f_2(y) \exp\{-j2\pi v y\} dy \right] = F_1(u) F_2(v)$$

6. **תגובת הלם:** למערכת ליניארית קבועה במקום המאופיינת ע"י התגובה להלם $h(x,y)$, תגובת התדר של המערכת הליניארית נתונה ע"י $H(u,v)=F\{h(x,y)\}$. אם למערכת כזו מוכנסת תמונה הרמונית $f(x,y) = \exp\{j2\pi(xU_0 + yV_0)\}$, מוצא המערכת יהיה

$$\begin{aligned} g(x,y) &= \int_{\alpha} \int_{\beta} \exp\{j2\pi(\alpha U_0 + \beta V_0)\} h(x-\alpha, y-\beta) d\alpha d\beta = \\ &= \int_{\alpha} \int_{\beta} \exp\{j2\pi((x-\alpha)U_0 + (y-\beta)V_0)\} h(\alpha, \beta) d\alpha d\beta = \\ &= \exp\{j2\pi(xU_0 + yV_0)\} \int_{\alpha} \int_{\beta} \exp\{-j2\pi(\alpha U_0 + \beta V_0)\} h(\alpha, \beta) d\alpha d\beta = \\ &= \exp\{j2\pi(xU_0 + yV_0)\} \cdot H(U_0, V_0) \end{aligned}$$

משמעות תוצאה זו היא שלאות הכניסה המוצע, המוצא הינו אותו אות כשהוא מוכפל בתגובת התדר. פיזיקלית פירוש הדבר שלכניסה הרמונית בתדר מרחבי מסוים, המוצא ממערכת ליניארית זהה בתדרו ומשנה רק את הפאזה והאמפליטודה בהתאם לאופי המערכת.

7. **הגדלה והקטנה:** תמונה מוגדלת ביחסים a ו- b בשני הצירים בהתאמה, תניב התמרה מוקטנת באותם יחסים, ועם הכפלה בגורם קבוע

$$F\{f(x,y)\} = F(u,v) \Rightarrow F\{f(ax, by)\} = \frac{F\left(\frac{u}{a}, \frac{v}{b}\right)}{|ab|}$$

8. **הזזה ואפנון:** הזזת תמונה גורמת לאיפנון בתדר, ואפנון במקום יוצר הזזה בתדר

$$\begin{aligned} F(u,v) &= F\{f(x,y)\} \\ \Rightarrow F\{f(x-X_0, y-Y_0)\} &= \exp\{j2\pi(uX_0 + vY_0)\} \cdot F(u,v) \\ \Rightarrow F\{\exp\{j2\pi(xU_0 + yV_0)\} f(x,y)\} &= F(u-U_0, v-V_0) \end{aligned}$$

9. **משפט הקונבולוציה:** קונבולוציה במקום מומרת להיות מכפלה בתדר, דהיינו,

$$\left. \begin{aligned} F_1(u,v) &= F\{f_1(x,y)\} \\ F_2(u,v) &= F\{f_2(x,y)\} \end{aligned} \right\} F\{f_1(x,y) \otimes f_2(x,y)\} = F_1(u,v) F_2(u,v)$$

באופן דומה, מכפלה של שני אותות דו-מימדיים במקום מותמרת לקונבולוציה בתדר, לפי הקשר

$$\left. \begin{aligned} F_1(u,v) &= F\{f_1(x,y)\} \\ F_2(u,v) &= F\{f_2(x,y)\} \end{aligned} \right\} F\{f_1(x,y) f_2(x,y)\} = F_1(u,v) \otimes F_2(u,v)$$

10. **משפט הקורלציה:** קורלציה בין שני אותות דו-מימדיים במקום, המוגדרת באופן הבא

$$f_1(x, y) * f_2(x, y) = \int_{\alpha} \int_{\beta} f_1(\alpha, \beta) f_2(\alpha + x, \beta + y) d\alpha d\beta$$

מומרת להיות מכפלה בתדר עם היפוך קואורדינטות

$$\left. \begin{aligned} F_1(u, v) &= F\{f_1(x, y)\} \\ F_2(u, v) &= F\{f_2(x, y)\} \end{aligned} \right\} F\{f_1(x, y) * f_2(x, y)\} = F_1(u, v) F_2(-u, -v)$$

באופן דומה, מכפלה של שני אותות דו-מימדיים במקום עם היפוך צירים מותמרת לקורלציה בתדר.

11. **שימור מכפלה פנימית:** מכפלה פנימית בין שתי תמונות היא בעלת אותו ערך בתדר ובמישור המקום, דהיינו

$$\begin{aligned} F_1(u, v) &= F\{f_1(x, y)\}, \quad F_2(u, v) = F\{f_2(x, y)\} \\ \Rightarrow \langle f_1(x, y), f_2(x, y) \rangle &= \int_x \int_y f_1(x, y) f_2^*(x, y) dx dy = \\ &= \langle F_1(u, v), F_2(u, v) \rangle = \int_u \int_v F_1(u, v) F_2^*(u, v) du dv \end{aligned}$$

כמקרה פרטי של תכונה זו מקבלים את משפט פרסבל לפיו האנרגיה של האות זהה במקום ובתדר

$$F(u, v) = F\{f(x, y)\} \Rightarrow \int_x \int_y |f(x, y)|^2 dx dy = \int_u \int_v |F(u, v)|^2 du dv$$

מדוע התמרת פוריה חשובה כל כך בדיון במערכות ליניאריות קבועות במקום? תשובה מסוימת לכך נרמזה בתכונות הנ"ל, ובעיקר תכונה 6 בהתייחס לתגובת התדר. בהינתן תמונה $f(x, y)$ ניתן להתייחס אליה כפונקציה הפרוסה על פני הבסיס הטרוויאללי של פונקציות דלתה (של דיראק) בכל מקום (x, y) . קבוצה זו של פונקציות מהווה בסיס אורתונורמלי הפורס את מרחב הפונקציות ב- L_2 . התמרת פוריה מציעה החלפת בסיס זה בקבוצת פונקציות אורתונורמליות אחרות הפורסות את מרחב הפונקציות – הפונקציות הרמוניות. יתרונו המשמעותי של הבסיס החדש הוא היותו עצמי למערכת ליניארית קבועה במקום. פירוש הדבר שבייצוג החדש, כל שעלינו לעשות על מנת לתאר את פעולת המערכת הוא לתאר כיצד "מוגברת" כל פונקציה הרמונית כזו, וזו בדיוק פונקציות תגובת התדר אותה פגשנו. ההסבר הנ"ל נשען על תחום במתמטיקה הקרוי אנליזה פונקציונלית. הסבר "נעים" יותר של התכונה הנ"ל יינתן בדיון באותות דיסקרטיים, ואז נקשור את הדברים לאלגברה ליניארית, הקלה יותר להבנה.

2.3 מערכות ליניאריות בדו-מימד – אותות דיסקרטיים

כבר ראינו כי ע"י דגימה עברנו מאות מעל הרצף $f(x,y)$ לאות דיסקרטי $f[m,n]$. בעקרון, כל שאמרנו בתחילת פרק זה תקף גם כאן עם שינוי סימון מתאימים. נחזור לפיכך בקצרה על עיקרי הדברים.

לאות $f[m,n]$, המוזן למערכת H , אנו מגדירים את מוצא המערכת ע"י $g[m,n] = H\{f[m,n]\}$. מערכת ליניארית אמ"ם $H\{\alpha f_1[m,n] + \beta f_2[m,n]\} = \alpha H\{f_1[m,n]\} + \beta H\{f_2[m,n]\}$. מערכת ליניארית H המוזנת בפונקצית הלם דו-מימדית $\delta[m - M_0, n - N_0]$ מניבה במוצאה את פונקצית התגובה להלם $h[m,n, M_0, N_0]$. בניגוד למקרה הרציף, פונקצית דלתה כאן פירושה '1' במקום השיא ואפס אחרת. מערכת ליניארית דיסקרטית מאופיינת לחלוטין ע"י תגובתה להלם. זאת רואים ע"י ייצוגה של התמונה כצירוף ליניארי ממושקל של פונקציות הלם דיסקרטיות

$$f[m,n] = \sum_k \sum_j f[k,j] \cdot \delta[m-k, n-j]$$

ובשל ליניאריות המערכת H , מתקבל מוצאה לאות $f[m,n]$ נתון ע"י הסכום

$$H\{f[m,n]\} = H\left\{ \sum_k \sum_j f[k,j] \delta[m-k, n-j] \right\} = \sum_k \sum_j f[k,j] h[m,n, k, j]$$

כאשר המערכת H קבועה במקום צריך להתקיים כי התגובה להלם היא פונקציה שאינה תלויה במקום, ולכן היא פונקציה של שני משתנים בלבד $h[m,n, k, j] = h[m-k, n-j]$. מוצא המערכת במקרה זה נתון ע"י $H\{f[m,n]\} = \sum_k \sum_j f[k,j] h[m-k, n-j] = f[m,n] \otimes h[m,n]$ וזוהי פעולת הקונבולוציה למקרה של מערכת קבועה במקום.

דוגמה 2.3: נדגים כיצד מיושמת קונבולוציה דו-מימדית דרך דוגמה. נניח לצורך פשטות קונבולוציה בין שני האותות הבאים:

$$f_1[m,n] = \begin{cases} 1 & -1 \leq m, n \leq 1 \\ 0 & \text{Otherwise} \end{cases}, \quad f_2[m,n] = \begin{cases} 1 & 0 \leq m, n \leq 1 \\ 0 & \text{Otherwise} \end{cases}$$

$$\sum_k \sum_j f_1[k,j] f_2[m-k, n-j] = f_1[m,n] \otimes f_2[m,n] = g[m,n]$$

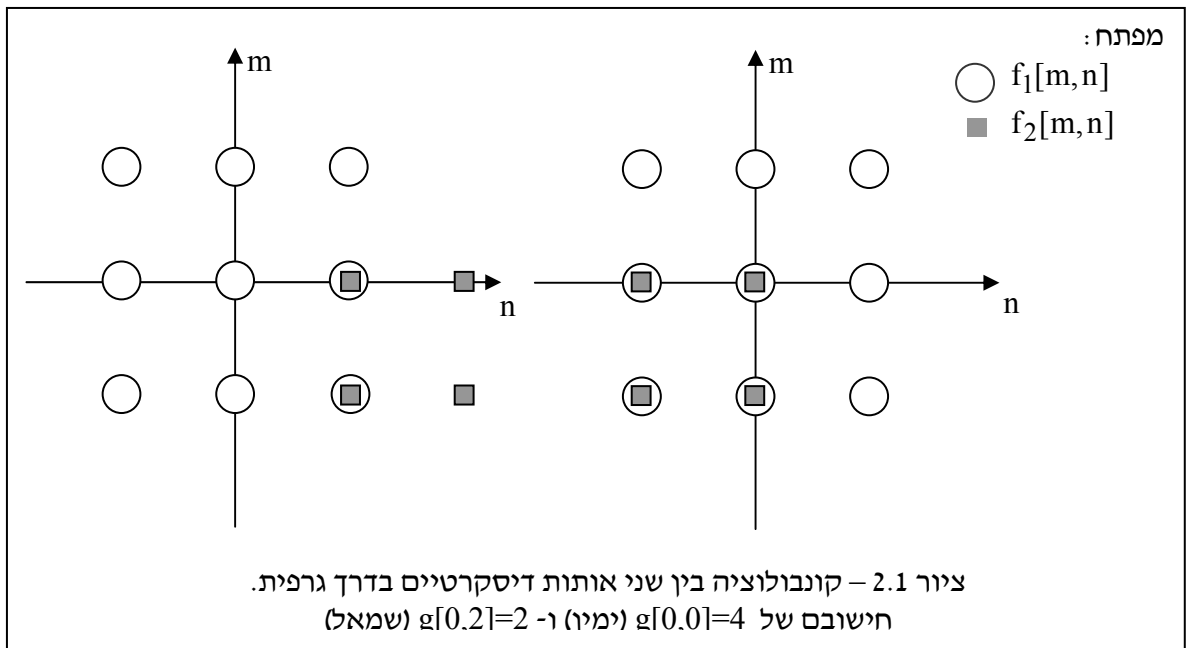
הפעולה הנ"ל יכולה להיעשות באופן גרפי ע"י ציור הפונקציה f_1 כמות שהיא, ועל-גביה ציור של הפונקציה f_2 כשהיא מהופכת בשני ציריה ומוזזת. גודל ההזזה קובע את מיקום המוצא. ערכו של המוצא

נקבע ע"י סיכום של מכפלות בין שני האותות המצויירים. למשל, כדי חישובם של $g[0,0]$ ו- $g[0,2]$ מתואר בציור 2.1. בגישה אלגברית נקבל

$$\begin{aligned}
 g[0,0] &= \sum_k \sum_j f_1[k,j]f_2[0-k,0-j] = \\
 &= f_1[-1,-1]f_2[1,1] + f_1[-1,0]f_2[1,0] + f_1[-1,1]f_2[1,-1] + \\
 &\quad f_1[0,-1]f_2[0,1] + f_1[0,0]f_2[0,0] + f_1[0,1]f_2[0,-1] + \\
 &\quad f_1[1,-1]f_2[-1,1] + f_1[1,0]f_2[-1,0] + f_1[1,1]f_2[-1,-1] = 1+1+1+1 = 4
 \end{aligned}$$

ובאופן דומה, $g[0,2]=2$.

מערכת ליניארית חסרת זיכרון תאופיין ע"י הקשר $g[m,n]=f[m,n]h[m,n]$. מערכת חסרת זיכרון וקבועה במקום תהיה פשוט הכפלת הכניסה בקבוע. ספרביליות של מערכת דו מימדית פירושה כי מתקיים $h[m,n] = h_1[m]h_2[n]$



דוגמה 2.4: נציג מספר מערכות דיסקרטיות ונדון בתכונותיהן. נעשה זאת בתאימות לדוגמה 2.1.

א. המערכת $H\{f[m,n]\} = f[m,n] + 3f[m-2,n+5]$ היא מערכת ליניארית. זוהי מערכת

קבועה במקום, עם זיכרון. תגובתה להלם היא $h[m,n] = \delta[m,n] + 3\delta[m-2,n+5]$.

ב. המערכת $H\{f[m,n]\} = f[m,n] + g[m,n]f[m-2,n+5]$ (עבור קבוע וידוע מראש)

היא ליניארית. זו מערכת שאינה קבועה במקום, וגם לה יש זיכרון. תגובתה להלם היא

$$h[m,n] = \delta[m,n] + g[m,n]\delta[m-2,n+5]$$

ג. המערכת $H\{f[m,n]\} = f[m,n] \cdot f[m-2,n+5]$ אינה ליניארית. זו מערכת עם זיכרון והיא

קבועה במקום.

- א. המערכת $H\{f[m,n]\} = f[m,n] + 1$ אינה ליניארית. זו מערכת חסרת זיכרון הקבועה במקום.
- ב. המערכת $H\{f[m,n]\} = \text{Log}\{f[m,n]\}$ אינה ליניארית. זו מערכת חסרת זיכרון הקבועה במקום.
- ג. המערכת $H\{f[m,n]\} = g[m,n] \cdot \text{Log}\{f[m,n]\}$ אינה ליניארית. בנוסף, זו מערכת תלויה מקום וחסרת זיכרון.
- ד. המערכת $H\{f[m,n]\} = f[m + 0.5, n + 0.5]$ אינה מוגדרת כלל, כי אין משמעות להזזה בחצי מרווח דגימה באותות דיסקרטיים.

מערכת תוגדר כמערכת FIR (Finite Impulse Response) אם תגובתה להלם בעלת תמך סופי. למערכות כאלה לא קיימת בעיית יציבות בשל היותן עם אפסים בלבד. מערכת שאינה FIR מוגדרת כמערכת IIR (Infinite Impulse Response). מערכת IIR כוללת גם קטבים ולכן מתעוררת בעיית יציבות. בעוד שבמערכות במימד אחד ניתן לחשב את הקטבים ע"י פקטוריאליזציה של פולינום, הרי שבדו-מימד זה בלתי אפשרי, והדבר מציב בעיה קשה. כפי שאנו נראה, בשל כך מסנני FIR מאוד פופולריים בתמונות, בעוד שמסנני IIR נדירים יותר.

2.4 התמרת פוריה דיסקרטית ותכונותיה

נניח כי מתוך האות $f(x,y)$ נוצר האות הדיסקרטי $f[m,n]$ ע"י פעולת דגימה במרווח יחידה בשני הצירים. נניח בנוסף כי דגימה זו התקבלה ע"י הכפלת התמונה הרציפה במסרק הלמים דו-מימדי. לכן

$$\tilde{f}[m,n] = f(x,y) \cdot \sum_m \sum_n \delta(m-x, n-y)$$

לפי רישום זה, כל דגימה היא פונקציה דלתה בגובה של ערך הפונקציה, ולכן הסימון השונה עם ה-tilde. התמרת פוריה הרציפה על אות זה תיתן

$$\begin{aligned} F\{\tilde{f}[m,n]\} &= \int_x \int_y f(x,y) \sum_m \sum_n \delta(m-x, n-y) \exp\{-j(ux + vy)\} dx dy = \\ &= \sum_m \sum_n \int_x \int_y f(m,n) \delta(m-x, n-y) \exp\{-j(ux + vy)\} dx dy = \\ &= \sum_m \sum_n f[m,n] \exp\{-j(um + vn)\} \end{aligned}$$

על בסיס זה, התמרת פוריה דו-מימדית לפונקציות דיסקרטיות $f[m,n]$ מוגדרת באופן הבא

$$F(\theta_1, \theta_2) = F_D \{f[m, n]\} = \sum_m \sum_n f[m, n] \exp\{-j(\theta_1 m + \theta_2 n)\}$$

המשתנים θ_1 ו- θ_2 הם תדרים מרחביים אופקי ואנכי, כפי שקיבלנו גם עבור המקרה הרציף. עם זאת קיים הבדל אחד מהותי – בעוד שבמקרה הרציף נעו תדרים אלו בתחום $[-\infty, +\infty]$, הרי שהפעם תדרים אלו מצויים בתחום $[-\pi, +\pi]$, ולכל ערך מחוץ לתחום זה, מתקיימת מחזוריות. ההתמרה ההפוכה נתונה ע"י

$$f[m, n] = F_D^{-1} \{F(\theta_1, \theta_2)\} = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} F(\theta_1, \theta_2) \cdot \exp\{+j(\theta_1 m + \theta_2 n)\} d\theta_1 d\theta_2$$

כשנדון במשפט הדגימה אנו נרחיב על הקשר הקיים בין הגדרות אלו ובין ההגדרות שניתנו להתמרת פוריה לאותות רציפים. לכל התכונות שהצגנו בסעיף קודם יש תואם בהתמרת הפוריה הנדונה כאן לאותות דיסקרטיים. נעבור על תכונות אלו בקצרה. בחרנו להשמיט את הדיון בתכונת ההקטנה/הגדלה, כיוון שבה לא מתקיימת דואליות פשוטה.

1. **יחידות:** לכל פונקציה $f[m, n]$ התמרה אחת ויחידה $F(\theta_1, \theta_2)$, ובביצוע ההתמרה ההפוכה, חוזרים לאותה פונקציה בדיוק. אין אובדן אינפורמציה בהתמרה קדימה או אחורה. צריך להיזהר ולומר כי תיתכנה שתי פונקציות שונות $f_1(x, y)$ ו- $f_2(x, y)$ אשר לאחר פעולת הדגימה הן זהות, ולכן התמרת הפוריה על האותות הדגומים יהיו זהים. אנו נכיר תופעה זו בהמשך, כשנדון בדגימת אותות דו-מימדיים.

2. **ליניאריות:** התמרת פוריה היא ליניארית, ולכן הפעלתה על צירוף ליניארי של תמונות זהה להפעלתה על כל תמונה בנפרד וצירוף ליניארי של התוצאות

$$F_D \{\alpha f_1[m, n] + \beta f_2[m, n]\} = \alpha F_D \{f_1[m, n]\} + \beta F_D \{f_2[m, n]\}$$

3. **התמרת תמונה צמודה:** הפעלת התמרת פוריה על תמונה צמודה מניבה התמרה צמודה ומהופכת בצירים, דהיינו

$$F_D \{f[m, n]\} = F(\theta_1, \theta_2) \Rightarrow F_D \{f^*[m, n]\} = F^*(-\theta_1, -\theta_2)$$

ולכן, עבור תמונה ממשית מתקבל כי התמרת הפוריה מקיימת $F(\theta_1, \theta_2) = F^*(-\theta_1, -\theta_2)$.

4. **היפוך צירים:** התמרת פוריה לתמונה עם היפוך צירים מניבה תוצאה הפוכת צירים

$$F\{f[m, n]\} = F(\theta_1, \theta_2) \Rightarrow F\{f[-m, -n]\} = F(-\theta_1, -\theta_2)$$

5. **ספרביליות ההתמרה:** ההתמרה ניתנת ליישום בשני שלבים בהם פועלים בחד-מימד. כמקודם, תכונה זו לא מניחה דבר על התמונה $f[m,n]$ עצמה. ניתן לראות זאת ישירות מההגדרה, לפי

$$F(\theta_1, \theta_2) = \sum_n \left[\sum_m f[m,n] \exp\{-j\theta_1 m\} \right] \exp\{-j\theta_2 n\}$$

בשל התכונה הנ"ל, כאשר אות הכניסה פריד מתקבל כי ההתמרה אף היא פרידה, והינה מכפלת שתי התמרות פוריה חד-מימדיות

$$F(\theta_1, \theta_2) = \left[\sum_m f_1[m] \exp\{-j\theta_1 m\} \right] \left[\sum_n f_2[n] \exp\{-j\theta_2 n\} \right]$$

6. **תגובת הלם:** למערכת ליניארית דיסקרטית קבועה במקום המאופיינת ע"י התגובה להלם $h[m,n]$, תגובת התדר של המערכת הליניארית נתונה ע"י $H(\theta_1, \theta_2) = F_D\{h[m,n]\}$. אם למערכת כזו מוכנסת תמונה הרמונית $f[m,n] = \exp\{j(mU_0 + nV_0)\}$, מוצא המערכת יהיה

$$g[m,n] = \sum_k \sum_i h[k,i] \exp\{j((m-k)U_0 + (n-i)V_0)\} = \exp\{j(mU_0 + nV_0)\} \cdot H(U_0, V_0)$$

משמעות תוצאה זו היא שלאות הכניסה המוצע, המוצא הינו אותו אות כשהוא מוכפל בתגובת התדר. פיזיקלית פירוש הדבר שלכניסה הרמונית בתדר מרחבי מסוים, המוצא ממערכת ליניארית זהה בתדרו ומשנה רק את הפאזה והאמפליטודה בהתאם לאופי המערכת.

7. **הזזה ואפנון:** הזזת תמונה גורמת לאיפנון בתדר, ואפנון במקום יוצר הזזה בתדר

$$\begin{aligned} F(\theta_1, \theta_2) &= F_D\{f[m,n]\} \\ \Rightarrow F_D\{f(m - M_0, n - N_0)\} &= \exp\{j(\theta_1 M_0 + \theta_2 N_0)\} \cdot F(\theta_1, \theta_2) \\ \Rightarrow F_D\{\exp\{j2\pi(mU_0 + nV_0)\} f[m,n]\} &= F(\theta_1 - U_0, \theta_2 - V_0) \end{aligned}$$

8. **משפט הקונבולוציה:** קונבולוציה במקום מומרת להיות מכפלה בתדר, דהיינו,

$$\left. \begin{aligned} F_1(\theta_1, \theta_2) &= F\{f_1[m,n]\} \\ F_2(\theta_1, \theta_2) &= F\{f_2[m,n]\} \end{aligned} \right\} F\{f_1[m,n] \otimes f_2[m,n]\} = F_1(\theta_1, \theta_2) F_2(\theta_1, \theta_2)$$

באופן דומה, מכפלה של שני אותות דו-מימדיים במקום מותמרת לקונבולוציה בתדר, לפי הקשר

$$\left. \begin{aligned} F_1(\theta_1, \theta_2) &= F\{f_1[m, n]\} \\ F_2(\theta_1, \theta_2) &= F\{f_2[m, n]\} \end{aligned} \right\} F\{f_1[m, n]f_2[m, n]\} = \frac{1}{4\pi^2} F_1(\theta_1, \theta_2) \otimes F_2(\theta_1, \theta_2)$$

9. **משפט הקורלציה:** קורלציה בין שני אותות דו-מימדיים במקום המוגדרת באופן הבא

$$f_1[m, n] * f_2[m, n] = \sum_k \sum_j f_1[k, j] f_2[m+k, n+j]$$

מומרת להיות מכפלה בתדר עם היפוך קואורדינטות

$$\left. \begin{aligned} F_1(\theta_1, \theta_2) &= F\{f_1[m, n]\} \\ F_2(\theta_1, \theta_2) &= F\{f_2[m, n]\} \end{aligned} \right\} F\{f_1[m, n] * f_2[m, n]\} = F_1(\theta_1, \theta_2) F_2(-\theta_1, -\theta_2)$$

באופן דומה, מכפלה של שני אותות דו-מימדיים במקום עם היפוך צירים מותמרת לקורלציה בתדר.

10. **שימור מכפלה פנימית:** מכפלה פנימית בין שתי תמונות היא בעלת אותו ערך בתדר ובמישור המקום, דהיינו

$$\begin{aligned} F_1(\theta_1, \theta_2) &= F_D\{f_1[m, n]\}, \quad F_2(\theta_1, \theta_2) = F_D\{f_2[m, n]\} \\ \Rightarrow \langle f_1[m, n], f_2[m, n] \rangle &= \sum_m \sum_n f_1[m, n] f_2[m, n] = \\ &= \langle F_1(\theta_1, \theta_2), F_2(\theta_1, \theta_2) \rangle = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} F_1(\theta_1, \theta_2) F_2^*(\theta_1, \theta_2) d\theta_1 d\theta_2 \end{aligned}$$

כמקרה פרטי של תכונה זו מקבלים את משפט פרסבל לפיו האנרגיה של האות זהה במקום ובתדר

$$F(\theta_1, \theta_2) = F_D\{f[m, n]\} \Rightarrow \sum_m \sum_n |f[m, n]|^2 = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} |F_1(\theta_1, \theta_2)|^2 d\theta_1 d\theta_2$$

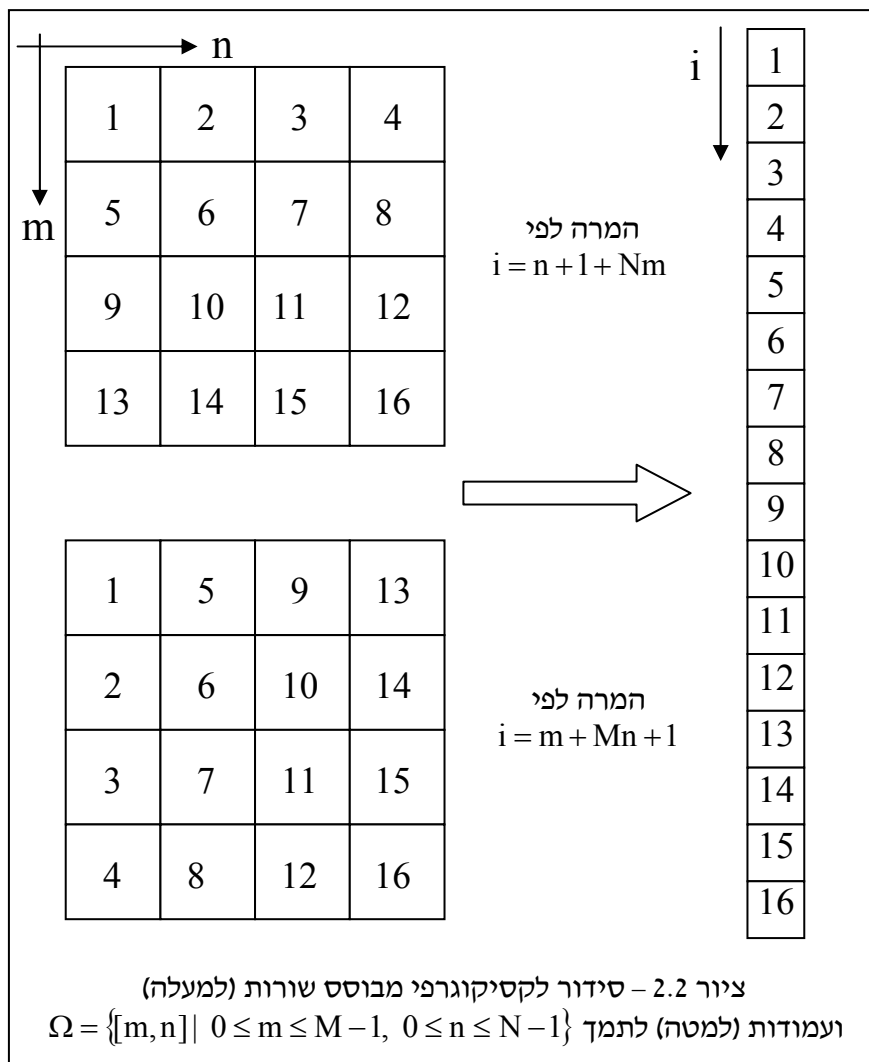
11. **מחזוריות:** כפי שכבר אמרנו קודם, הפונקציה $F(\theta_1, \theta_2)$ מחזורית ב- 2π בכל אחד מציריה, ולכן ערכיה מעניינים אותנו רק בריבוע $[-\pi, +\pi] \times [-\pi, +\pi]$.

2.5 אותות ומערכות ליניאריות בתמך סופי

עד כה עסקנו באותות מעל השריג הדגום $[m,n]$, אך לא הגבלנו את תמך התמונה. נניח כעת כי התמונה עליה רצוננו לפעול נתונה לנו כ- $f[m,n]$ ומוגדרת מעל השריג $\Omega = \{[m,n] \mid 0 \leq m \leq M-1, 0 \leq n \leq N-1\}$ הכולל NM נקודות בדיוק. נניח כי נתונה לנו מערכת ליניארית דיסקרטית H המאופיינת ע"י התגובה להלם $h[m,n,k,j]$. מוצא המערכת נתון ע"י

$$g[m,n] = H\{f[m,n]\} = \sum_k \sum_j f[k,j]h[m,n,k,j]$$

עד כאן לא ראינו כל שוני ביחס לנאמר בסעיף קודם. כיוון שהאות $f[m,n]$ כולל כמות סופית של דגימות, נוכל לסדרו כוקטור באורך MN . זאת נוכל לעשות בין היתר גם ע"י שרשור השורות או שרשור העמודות בזו אחר זו. תהליך זה קרוי column/row stack או סידור לקסיקוגרפי (lexicographic ordering) עפ"י שורות/עמודות. ציור 2.2 ממחיש תהליך סידור זה.



היתרון הגדול בדיון על תמונות דיסקרטיות עם תמך סופי הוא היכולת לעשות שימוש בכלים מאלגברה ליניארית, כפי שנראה מיד. נטען כאת את הטענה הבאה: כל מערכת ליניארית H (אשר כזכור מאופיינת ע"י תגובתה להלם $h[m,n,k,j]$) ניתנת לתיאור כמטריצה. בהינתן תמונת כניסה בגודל $M_1 \times N_1$, מומרת תמונה זו לוקטור בגודל $M_1 N_1$. על תמונה זו מופעלת מערכת ליניארית ומוצאה היא תמונה בגודל $M_2 \times N_2$, המומרת לוקטור בגודל $M_2 N_2$. המטריצה המתארת את פעולת המערכת תהיה בגודל של $M_2 N_2$ שורות ו- $M_1 N_1$ עמודות. תוכן מטריצה זו נקבע באופן חד-חד ערכי ע"י תגובת ההלם של המערכת. רעיון זה של ייצוגה של מערכת ליניארית ע"י מטריצה אינו חדש ונדון כבר באותות ומערכות חד-מימדיים. כאן נרחיב את הדיון לאותות דו-מימדיים.

טענה זו אינה כה דרמטית כפי שהיא נשמעת! אם נסמן ב- \underline{f} את התמונה לאחר הסידור הלקסיקוגרפי, וב- \underline{H} את מטריצת המערכת, המכפלה $\underline{H}\underline{f}$ יוצרת וקטור חדש שיסומן $\underline{g} = \underline{H}\underline{f}$. דבר ראשון בו נבחין הוא שלמערכת נכנסה תמונה ויצאה תמונה, כששתיהן בגדלים הנכונים, וששתיהן מסודרות בסידור לקסיקוגרפי. האיבר ה- i בוקטור \underline{g} בנוי כמכפלה של השורה ה- i מהמטריצה \underline{H} בוקטור \underline{f} , לפי

$$g_i = \sum_{k=1}^{M_1 N_1} H_{i,k} f_k$$

ובכך בפועל מתקבל שכל פיקסל בתמונת המוצא הוא צירוף ליניארי של פיקסלים מתמונת הכניסה. מה באשר לתוכן של המטריצה \underline{H} ? כיצד נקבע אותו? נענה על שאלה זו דרך דוגמה בסיסית.

דוגמה 2.5: נניח כי אנו עוסקים בתמונות $f[m,n]$ ו- $g[m,n]$ מעל השריג $\Omega = \{[m,n] \mid 0 \leq m \leq 1, 0 \leq n \leq 1\}$ (כלומר, תמונות של 2 על 2 פיקסלים). ונניח אופרטור כללי

$$g[m,n] = H\{f[m,n]\} = \sum_k \sum_j f[k,j] h[m,n,k,j]$$

נניח כי אנו עוסקים בסידור לקסיקוגרפי לפי עמודות. נמחיש תחילה כיצד תיראה השורה הראשונה. שורה זו בונה את האיבר ה- $[0,0]$ של $g[m,n]$. לפי הקשר הנ"ל מתקבל

$$\begin{aligned}
 g[0,0] &= \sum_k \sum_j f[k,j]h[0,0,k,j] = \\
 &= f[0,0]h[0,0,0,0] + f[1,0]h[0,0,1,0] + f[0,1]h[0,0,0,1] + f[1,1]h[0,0,1,1] = \\
 &= \begin{bmatrix} h[0,0,0,0] & h[0,0,1,0] & h[0,0,0,1] & h[0,0,1,1] \end{bmatrix} \begin{bmatrix} f[0,0] \\ f[1,0] \\ f[0,1] \\ f[1,1] \end{bmatrix}
 \end{aligned}$$

לא קשה להשתכנע שהמטריצה \mathbf{H} בכללה נותנת את הקשר הבא

$$\begin{bmatrix} g[0,0] \\ g[1,0] \\ g[0,1] \\ g[1,1] \end{bmatrix} = \begin{bmatrix} h[0,0,0,0] & h[0,0,1,0] & h[0,0,0,1] & h[0,0,1,1] \\ h[1,0,0,0] & h[1,0,1,0] & h[1,0,0,1] & h[1,0,1,1] \\ h[0,1,0,0] & h[0,1,1,0] & h[0,1,0,1] & h[0,1,1,1] \\ h[1,1,0,0] & h[1,1,1,0] & h[1,1,0,1] & h[1,1,1,1] \end{bmatrix} \begin{bmatrix} f[0,0] \\ f[1,0] \\ f[0,1] \\ f[1,1] \end{bmatrix}$$

כחמשך של דוגמה זו, נניח כעת כי את תמונת המוצא אנו רוצים לקבל מעל תמך שונה, בו יש שישה ערכי מוצא במקום ארבעה - $\Omega_g = \{[m,n] \mid 0 \leq m \leq 2, 0 \leq n \leq 1\}$. חזרה על התהליך הנ"ל תיתן

$$\begin{bmatrix} g[0,0] \\ g[1,0] \\ g[2,0] \\ g[0,1] \\ g[1,1] \\ g[2,1] \end{bmatrix} = \begin{bmatrix} h[0,0,0,0] & h[0,0,1,0] & h[0,0,0,1] & h[0,0,1,1] \\ h[1,0,0,0] & h[1,0,1,0] & h[1,0,0,1] & h[1,0,1,1] \\ h[2,0,0,0] & h[2,0,1,0] & h[2,0,0,1] & h[2,0,1,1] \\ h[0,1,0,0] & h[0,1,1,0] & h[0,1,0,1] & h[0,1,1,1] \\ h[1,1,0,0] & h[1,1,1,0] & h[1,1,0,1] & h[1,1,1,1] \\ h[2,1,0,0] & h[2,1,1,0] & h[2,1,0,1] & h[2,1,1,1] \end{bmatrix} \begin{bmatrix} f[0,0] \\ f[1,0] \\ f[0,1] \\ f[1,1] \end{bmatrix}$$

כשניתן לראות כי ארבע מהשורות במשוואה מטריצית זו חוזרות בדיוק על קשרים מהמשוואה הקודמת.

מתוך הדוגמה הנ"ל גם ברור מה יקרה בתמונות גדולות יותר. $H[m,n,k,j]$ עבור $[m,n]$ קבועים ימלאו את השורה המתאימה לאיבר ה- $[m,n]$ במוצא. מספרה של שורה זו יהיה במקרה הכללי $i = m + 1 + Mn$. הערכים בתוך השורה יסודרו לפי הסידור הלקסיקוגרפי של האינדקסים של תמונת הכניסה.

תכונה מעניינת שנובעת מהתיאור הנ"ל היא נושא הצירוף בין מערכות. נניח כי נתונות לנו שתי מערכות ליניאריות (בתיאור מטריצי שלהן ולגודל זהה של תמונת כניסה ומוצא) \mathbf{H}_1 ו- \mathbf{H}_2 . הפעלת שתי המערכות על אותה כניסה וסיכום המוצאים ייתן

$$\underline{g}_1 = \mathbf{H}_1 \underline{f} \quad , \quad \underline{g}_2 = \mathbf{H}_2 \underline{f} \quad \Rightarrow \quad \underline{g}_1 + \underline{g}_2 = \mathbf{H}_1 \underline{f} + \mathbf{H}_2 \underline{f} = [\mathbf{H}_1 + \mathbf{H}_2] \underline{f}$$

כך שסיכום מוצאי המערכות כרוך בעצם בסיכום המטריצות. זה לא מפתיע כי סיכום המערכות כמוהו כמו סיכום פונקציות התגובה להלם בשל הליניאריות.

באופן דומה, עבור מערכות ריבועיות, דהיינו, מערכות שתמונת הכניסה באותו גודל של תמונת היציאה, שרשרת המערכות בזו אחר זו נותן

$$\underline{g}_1 = \mathbf{H}_1 \underline{f} \quad , \quad \underline{g}_2 = \mathbf{H}_2 \underline{g}_1 = \mathbf{H}_2 \mathbf{H}_1 \underline{f}$$

והפעם קיבלנו כי יש להכפיל את המטריצות זו בזו כדי לקבל מערכת אפקטיבית אחת עם אותה פעולה. תכונה ידועה באלגברה ליניארית היא שמטריצות ריבועיות בד"כ לא מתחלפות במכפלה, כלומר $\mathbf{H}_1 \mathbf{H}_2 \neq \mathbf{H}_2 \mathbf{H}_1$. אמירה זו מיתרגמת לכך שלא ניתן במקרה הכללי להחליף סדר הפעלתן של מערכות ליניאריות – החלפה כזו צפויה לתת תוצאה שונה. עם זאת, הזכרנו קודם כי מערכות ליניאריות קבועות במקום ניתנות להחלפה ללא פגיעה בתוצאה. זה מעורר את השאלה מה מיוחד במטריצות המייצגות מערכת ליניארית קבועה במקום? כזכור, מערכת קבועה במקום מאופיינת בקשר $h[m,n,k,j]=h[m-k,n-j]$. נחזור לדוגמה הקודמת ונראה כיצד נראית מטריצה זו לתמונה קטנה.

דוגמה 2.6: בהמשך לדוגמה 2.5, נניח כעת שהמערכת קבועה במקום ולכן מתקיים הקשר שהוזכר למעלה: $h[m,n,k,j]=h[m-k,n-j]$. הצבת תכונה זו במטריצה שהתקבלה (לקלט ופלט מעל אותו תמך) תיתן את המטריצה הבאה

$$\begin{bmatrix} g[0,0] \\ g[1,0] \\ g[0,1] \\ g[1,1] \end{bmatrix} = \begin{bmatrix} h[0,0,0,0] & h[0,0,1,0] & h[0,0,0,1] & h[0,0,1,1] \\ h[1,0,0,0] & h[1,0,1,0] & h[1,0,0,1] & h[1,0,1,1] \\ h[0,1,0,0] & h[0,1,1,0] & h[0,1,0,1] & h[0,1,1,1] \\ h[1,1,0,0] & h[1,1,1,0] & h[1,1,0,1] & h[1,1,1,1] \end{bmatrix} \begin{bmatrix} f[0,0] \\ f[1,0] \\ f[0,1] \\ f[1,1] \end{bmatrix} = \begin{bmatrix} h[0,0] & h[-1,0] & h[0,-1] & h[-1,-1] \\ h[1,0] & h[0,0] & h[1,-1] & h[0,-1] \\ h[0,1] & h[-1,1] & h[0,0] & h[-1,0] \\ h[1,1] & h[0,1] & h[1,0] & h[0,0] \end{bmatrix} \begin{bmatrix} f[0,0] \\ f[1,0] \\ f[0,1] \\ f[1,1] \end{bmatrix}$$

עבור החלק השני של דוגמה 2.5 נקבל:

$$\begin{bmatrix} g[0,0] \\ g[1,0] \\ g[2,0] \\ g[0,1] \\ g[1,1] \\ g[2,1] \end{bmatrix} = \begin{bmatrix} h[0,0] & h[-1,0] & h[0,-1] & h[-1,-1] \\ h[1,0] & h[0,0] & h[1,-1] & h[0,-1] \\ h[2,0] & h[1,0] & h[2,-1] & h[1,-1] \\ h[0,1] & h[-1,1] & h[0,0] & h[-1,0] \\ h[1,1] & h[0,1] & h[1,0] & h[0,0] \\ h[2,1] & h[1,1] & h[2,0] & h[1,0] \end{bmatrix} \begin{bmatrix} f[0,0] \\ f[1,0] \\ f[0,1] \\ f[1,1] \end{bmatrix}$$

הייחוד במטריצות הנ"ל ניכר אם מסתכלים על ארבעת חלקיה בגודל 2 על 2 במקרה הראשון, ובלוקים בגודל 3 (שורות) על 2 במקרה השני. כל תת-בלוק כזה הוא בעל מבנה טופליץ, בו ערכי האיברים על האלכסונים הראשיים זהים, כמתואר בציור 2.3. המטריצה אותה קיבלנו בדוגמה היא בעלת מבנה בלוק-טופליץ. מבנה זה מתקבל אם כל בלוק במטריצה מהווה מטריצת טופליץ, ואם בנוסף, הבלוקים עצמם ערוכים בצורת טופליץ (שלאורך אלכסונים ראשיים הבלוקים זהים).

במקרה הכללי, אנו צפויים לקבל כי מערכת ליניארית קבועה במקום הפועלת על תמונה בגודל $M_1 \times N_1$ ומניבה תמונה בגודל $M_2 \times N_2$ תתואר כמטריצה בגודל $M_2 N_2$ שורות על $M_1 N_1$ עמודות. מטריצה זו תחולק ל- $N_2 \times N_1$ בלוקים כשכל אחד בגודל $M_2 \times M_1$ ובעל מבנה טופליץ, והסדר ביניהם הוא כשל מבנה טופליץ.

$$\begin{bmatrix} a & b & c & d & e \\ f & a & b & c & d \\ g & f & a & b & c \\ h & g & f & a & b \\ i & h & g & f & a \end{bmatrix}$$

ציור 2.3 – מבנה כללי של מטריצת טופליץ בגודל 5 על 5. רק 9 ערכים מאפיינים את כל 25 כניסותיה של המטריצה. אם כל איבר במטריצה הנ"ל הוא עצמו מטריצת טופליץ, הרי שהמבנה המתקבל הוא מבנה של מטריצת בלוק-טופליץ.

לסיכום ביניים: המטריצה המתארת מערכת ליניארית קבועה במקום הינה בעלת מבנה בלוק טופליץ. האם מבנה זה מספק את תכונת ההחלפה – התשובה היא לצערנו עדיין לא. נראה זאת דרך הדוגמה הבאה.

דוגמה 2.7: ניקח שתי מטריצות טופליץ ונכפול אותן זו בזו בשני הכיוונים.

$$\begin{pmatrix} 3.0 & 1.0 & 0.8 & 0.0 \\ 1.0 & 3.0 & 1.0 & 0.8 \\ 0.8 & 1.0 & 3.0 & 1.0 \\ 0.0 & 0.8 & 1.0 & 3.0 \end{pmatrix} \begin{pmatrix} 2.0 & 1.0 & 0.5 & 0.0 \\ 1.0 & 2.0 & 1.0 & 0.5 \\ 0.5 & 1.0 & 2.0 & 1.0 \\ 0.0 & 0.5 & 1.0 & 2.0 \end{pmatrix} = \begin{pmatrix} 7.4 & 5.8 & 4.1 & 1.3 \\ 5.5 & 8.4 & 6.3 & 4.1 \\ 4.1 & 6.3 & 8.4 & 5.5 \\ 1.3 & 4.1 & 5.8 & 7.4 \end{pmatrix}$$

$$\begin{pmatrix} 2.0 & 1.0 & 0.5 & 0.0 \\ 1.0 & 2.0 & 1.0 & 0.5 \\ 0.5 & 1.0 & 2.0 & 1.0 \\ 0.0 & 0.5 & 1.0 & 2.0 \end{pmatrix} \begin{pmatrix} 3.0 & 1.0 & 0.8 & 0.0 \\ 1.0 & 3.0 & 1.0 & 0.8 \\ 0.8 & 1.0 & 3.0 & 1.0 \\ 0.0 & 0.8 & 1.0 & 3.0 \end{pmatrix} = \begin{pmatrix} 7.4 & 5.5 & 4.1 & 1.3 \\ 5.8 & 8.4 & 6.3 & 4.1 \\ 4.1 & 6.3 & 8.4 & 5.8 \\ 1.3 & 4.1 & 5.5 & 7.4 \end{pmatrix}$$

כפי שאנו רואים, המכפלה קרובה מאוד אך לא זהה. בהמשך נבין מהיכן נובע שינוי זה, ומדוע לא התקבלה יכולת החלפה כמקובל באותות בעלי תמך אינסופי.

נמשיך בקו אותו בנינו עד כה, ונדון כעת במקרה בו המערכת הליניארית פרידה. במקרה זה מתקיים $h[m,n] = h_1[m]h_2[n]$. נחזור שוב על הדוגמה הנדונה ונראה כיצד משתנה המטריצה במקרה זה.

דוגמה 2.8: בהמשך לדוגמה 2.6 (החלק הראשון), נניח כעת שהמערכת גם פרידה. המטריצה תהיה

$$\begin{bmatrix} g[0,0] \\ g[1,0] \\ g[0,1] \\ g[1,1] \end{bmatrix} = \begin{bmatrix} h_1[0]h_2[0] & h_1[-1]h_2[0] & h_1[0]h_2[-1] & h_1[-1]h_2[-1] \\ h_1[1]h_2[0] & h_1[0]h_2[0] & h_1[1]h_2[-1] & h_1[0]h_2[-1] \\ h_1[0]h_2[1] & h_1[-1]h_2[1] & h_1[0]h_2[0] & h_1[-1]h_2[0] \\ h_1[1]h_2[1] & h_1[0]h_2[1] & h_1[1]h_2[0] & h_1[0]h_2[0] \end{bmatrix} \begin{bmatrix} f[0,0] \\ f[1,0] \\ f[0,1] \\ f[1,1] \end{bmatrix}$$

מה מיוחד במבנה הני"ל? קשה לראות זאת אך ישנה כאן מכפלת קרונקר (Kronecker) בין שתי מטריצות בגודל 2 על 2 בצורה

$$\begin{aligned} & \begin{bmatrix} h_1[0]h_2[0] & h_1[-1]h_2[0] & h_1[0]h_2[-1] & h_1[-1]h_2[-1] \\ h_1[1]h_2[0] & h_1[0]h_2[0] & h_1[1]h_2[-1] & h_1[0]h_2[-1] \\ h_1[0]h_2[1] & h_1[-1]h_2[1] & h_1[0]h_2[0] & h_1[-1]h_2[0] \\ h_1[1]h_2[1] & h_1[0]h_2[1] & h_1[1]h_2[0] & h_1[0]h_2[0] \end{bmatrix} = \\ & = \begin{bmatrix} h_2[0] & h_2[-1] \\ h_2[1] & h_2[0] \end{bmatrix} \otimes \begin{bmatrix} h_1[0] & h_1[-1] \\ h_1[1] & h_1[0] \end{bmatrix} = \\ & = \begin{bmatrix} h_2[0] \cdot \begin{bmatrix} h_1[0] & h_1[-1] \\ h_1[1] & h_1[0] \end{bmatrix} & h_2[-1] \cdot \begin{bmatrix} h_1[0] & h_1[-1] \\ h_1[1] & h_1[0] \end{bmatrix} \\ h_2[1] \cdot \begin{bmatrix} h_1[0] & h_1[-1] \\ h_1[1] & h_1[0] \end{bmatrix} & h_2[0] \cdot \begin{bmatrix} h_1[0] & h_1[-1] \\ h_1[1] & h_1[0] \end{bmatrix} \end{bmatrix} \end{aligned}$$

באופן דומה, עבור החלק השני של הדוגמה יתקבל

$$\begin{bmatrix} h_1[0]h_2[0] & h_1[-1]h_2[0] & h_1[0]h_2[-1] & h_1[-1]h_2[-1] \\ h_1[1]h_2[0] & h_1[0]h_2[0] & h_1[1]h_2[-1] & h_1[0]h_2[-1] \\ h_1[2]h_2[0] & h_1[1]h_2[0] & h_1[2]h_2[-1] & h_1[1]h_2[-1] \\ h_1[0]h_2[1] & h_1[-1]h_2[1] & h_1[0]h_2[0] & h_1[-1]h_2[0] \\ h_1[1]h_2[1] & h_1[0]h_2[1] & h_1[1]h_2[0] & h_1[0]h_2[0] \\ h_1[2]h_2[1] & h_1[1]h_2[1] & h_1[2]h_2[0] & h_1[1]h_2[0] \end{bmatrix} =$$

$$= \begin{bmatrix} h_2[0] & h_2[-1] \\ h_2[1] & h_2[0] \end{bmatrix} \otimes \begin{bmatrix} h_1[0] & h_1[-1] \\ h_1[1] & h_1[0] \\ h_1[2] & h_1[1] \end{bmatrix}$$

(יש לשים לב ולהיזהר מכך שבחרנו באותו סימון לקונבולוציה ולמכפלת קרונקר).

במקרה הכללי, אנו צפויים לקבל כי מערכת ליניארית קבועה במקום ופרידה הפועלת על תמונה בגודל $M_1 \times N_1$ ומניבה תמונה בגודל $M_2 \times N_2$ תתואר כמטריצה בגודל $M_2 N_2$ שורות על $M_1 N_1$ עמודות. מטריצה זו תיבנה כמכפלת קרונקר של מטריצות בגודל $N_2 \times N_1$ ו- $M_2 \times M_1$ כשכל אחת מהן בעלת מבנה טופליץ. ברור כי למטריצה הכוללת יהיה מבנה בלוק טופליץ, בשל דרך בנייתה.

הפירוק הנ"ל אינו מקרי. אם נחזור ליסודות, הרי שקונבולוציה עם תגובת הלם פרידה פירושה הקשר הבא

$$g[m, n] = H\{f[m, n]\} = \sum_k \sum_j f[k, j] h[m - k, n - j] = \sum_k h_1[m - k] \sum_j f[k, j] h_2[n - j]$$

וכבר נאמר בפרק זה כי קשר זה משמעו פעולה נפרדת על עמודות ושורות. כיצד תכונה זו באה לידי ביטוי בייצוג מטריצי? ובכן, שוב דרך דוגמה, נמחיש כיצד הדבר פועל.

דוגמה 2.9: ניקח את תמונת הכניסה מדוגמה קודמת, $f[m, n]$, והפעם נחזיק אותה כמטריצה בגודל 2 על 2, לפי

$$f = \begin{bmatrix} f[0,0] & f[0,1] \\ f[1,0] & f[1,1] \end{bmatrix}$$

נניח כי רצוננו להפעיל את המערכת הליניארית הקבועה במקום ופרידה כפי שתואר בדוגמה הקודמת, לשם יצירת תמונת מוצא בגודל 3 שורות על 2 עמודות. הקשר הבא מתאר כיצד תפעל מערכת זו, וניתן, ע"י פתיחת המכפלות לוודא את נכונותה של המשוואה

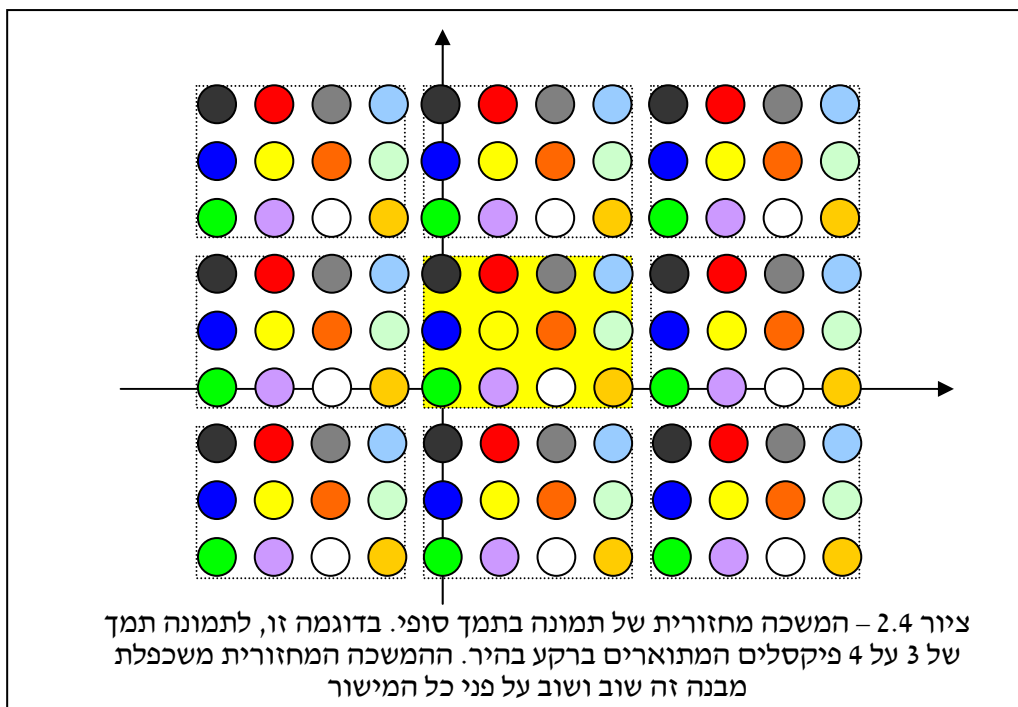
$$\begin{bmatrix} h_1[0] & h_1[-1] \\ h_1[1] & h_1[0] \\ h_1[2] & h_1[1] \end{bmatrix} \begin{bmatrix} f[0,0] & f[0,1] \\ f[1,0] & f[1,1] \end{bmatrix} \begin{bmatrix} h_2[0] & h_2[-1] \\ h_2[1] & h_2[0] \end{bmatrix} = \begin{bmatrix} g[0,0] & g[0,1] \\ g[1,0] & g[1,1] \\ g[2,0] & g[2,1] \end{bmatrix}$$

המטריצה כופלת את התמונה מצד ימין (בגודל 2 על 2) מבצעת פעולת סינון שורות לתמונה – ניתן לראות זאת ע"י פתיחת המכפלה, ואבחנה כי נוצרו צירופים ליניאריים של פיקסלים באותה שורה. באופן דומה, המכפלה מצד שמאל מבצעת סינון עמודות.

נושא אחרון בו נדון במסגרת סעיף זה היא פעולת הקונבולוציה הציקלית. עד כה, כשפלשה תגובת ההלם מהתמך, הנחנו כי מחוצה לו ישנם אפסים. כעת מציע גישה שונה הקרויה המשכה מחזורית של התמונה. בהינתן תמונה המוגדרת על התמך $\Omega = \{[m, n] \mid 0 \leq m \leq M-1, 0 \leq n \leq N-1\}$, נגדיר את ערכיה בכל מקום במישור באופן הבא:

$$f[m, n] = \begin{cases} f[m, n] & 0 \leq m \leq M-1, 0 \leq n \leq N-1 \\ f[m \bmod M, n \bmod N] & \text{otherwise} \end{cases}$$

ציור 2.4 מתאר תהליך זה לתמונה בגודל 3 על 4.



פעולת הקונבולוציה בין תמונה כזו (לאחר המשכה מחזורית) ובין תגובת הלם (ואנו מניחים כאן מערכת לינארית קבועה במקום), תתקבל ע"י הביטוי הבא

$$g_c[m, n] = \sum_{k=0}^{M-1} \sum_{j=0}^{N-1} f[k, j] h[(m-k) \bmod M, (n-j) \bmod N]$$

הקרוי קונבולוציה ציקלית. הפעולה \bmod נותנת שכפול לפי הקשר $m \bmod M = m - \lfloor m/M \rfloor M$, ופעולת ה- $\lfloor * \rfloor$ היא פעולת הערך השלם הגדול ביותר מלמטה לערך. כדוגמה, עבור m בתחום 0 עד $M-1$, הערך לא משתנה כלל. באופן דומה, $M \bmod M = 0$, $(M+1) \bmod M = 1$. לערכים שליליים, $M \bmod M = -1$, $(M-1) \bmod M = M-1$, וכיוצא שצא שכאשר התגובה להלם פולשת מצד אחד של התמך, היא בפועל כופלת איברים מצידו האחר. נחזור על דוגמאות קודמות, כשהפעם אנו מבצעים קונבולוציה ציקלית

דוגמה 2.10: אנו מתייחסים לתמונת כניסה $f[m,n]$ מעל תמך $\Omega = \{[m,n] \mid 0 \leq m \leq 1, 0 \leq n \leq 1\}$, ונתעניין בערכי המוצא $g[m,n]$ מעל השריג $\Omega_g = \{[m,n] \mid 0 \leq m \leq 2, 0 \leq n \leq 1\}$. נניח כי אנו עוסקים בסידור לקסיקוגרפי לפי עמודות. נמחיש תחילה כיצד תיראה השורה הראשונה. שורה זו בונה את האיבר ה- $[0,0]$ של $g[m,n]$. לפי הקשר הנ"ל מתקבל

$$\begin{aligned} g[0,0] &= \sum_k \sum_j f[k,j] h[-k \bmod 2, -j \bmod 2] = \\ &= f[0,0]h[0,0] + f[1,0]h[1,0] + f[0,1]h[0,1] + f[1,1]h[1,1] = \\ &= \begin{bmatrix} h[0,0] & h[1,0] & h[0,1] & h[1,1] \end{bmatrix} \begin{bmatrix} f[0,0] \\ f[1,0] \\ f[0,1] \\ f[1,1] \end{bmatrix} \end{aligned}$$

והמטריצה H בכללה נותנת את הקשר הבא

$$\begin{bmatrix} g[0,0] \\ g[1,0] \\ g[2,0] \\ g[0,1] \\ g[1,1] \\ g[2,1] \end{bmatrix} = \begin{bmatrix} h[0,0] & h[1,0] & h[0,1] & h[1,1] \\ h[1,0] & h[0,0] & h[1,1] & h[0,1] \\ h[0,0] & h[1,0] & h[0,1] & h[1,1] \\ h[0,1] & h[1,1] & h[0,0] & h[1,0] \\ h[1,1] & h[0,1] & h[1,0] & h[0,0] \\ h[0,1] & h[1,1] & h[0,0] & h[1,0] \end{bmatrix} \begin{bmatrix} f[0,0] \\ f[1,0] \\ f[0,1] \\ f[1,1] \end{bmatrix}$$

כפי שרואים, גם הפעם למטריצה מבנה בלוק טופליץ, כמקודם. למעשה, המבנה אפילו יותר ייחודי מאשר בלוק טופליץ כללי – המבנה כעת בלוק סיבובי (Block-Circulant). במבנה זה, שורה ראשונה של המטריצה מכתובה את כולה, ע"י סיבוב משמאל לימין. ערך שנפלט ימינה נכנס משמאל. ציור 2.5 ממחיש דרך בנייתה של מטריצה כזו. גם הפעם, מטריצה בעלת מבנה בלוק-סיבובי מתקבלת ע"י בלוקים שכל אחד מהם הוא בלוק-סיבובי, וסדר הבלוקים הוא בלוק-סיבובי.

נמשיך את הדיון בדוגמה, כשאנו מניחים כי במערכת פרידה. במקרה זה מתקבל

$$\begin{bmatrix} g[0,0] \\ g[1,0] \\ g[2,0] \\ g[0,1] \\ g[1,1] \\ g[2,1] \end{bmatrix} = \begin{bmatrix} h_1[0]h_2[0] & h_1[1]h_2[0] & h_1[0]h_2[1] & h_1[1]h_2[1] \\ h_1[1]h_2[0] & h_1[0]h_2[0] & h_1[1]h_2[1] & h_1[0]h_2[1] \\ h_1[0]h_2[0] & h_1[1]h_2[0] & h_1[0]h_2[1] & h_1[1]h_2[1] \\ h_1[0]h_2[1] & h_1[1]h_2[1] & h_1[0]h_2[0] & h_1[1]h_2[0] \\ h_1[1]h_2[1] & h_1[0]h_2[1] & h_1[1]h_2[0] & h_1[0]h_2[0] \\ h_1[0]h_2[1] & h_1[1]h_2[1] & h_1[0]h_2[0] & h_1[1]h_2[0] \end{bmatrix} \begin{bmatrix} f[0,0] \\ f[1,0] \\ f[0,1] \\ f[1,1] \end{bmatrix} =$$

$$= \begin{bmatrix} h_2[0] & h_2[1] \\ h_2[1] & h_2[0] \\ h_2[0] & h_2[1] \end{bmatrix} \otimes \begin{bmatrix} h_1[0] & h_1[1] \\ h_1[1] & h_1[0] \end{bmatrix} \begin{bmatrix} f[0,0] \\ f[1,0] \\ f[0,1] \\ f[1,1] \end{bmatrix}$$

$$\begin{bmatrix} a & b & c & d & e \\ f & a & b & c & d \\ g & f & a & b & c \\ h & g & f & a & b \\ i & h & g & f & a \end{bmatrix} \quad \begin{bmatrix} a & b & c & d & e \\ e & a & b & c & d \\ d & e & a & b & c \\ c & d & e & a & b \\ b & c & d & e & a \end{bmatrix}$$

ציור 2.5 – מבנה כללי של מטריצה סיבובית בגודל 5 על 5. אם בטופליץ (צד שמאל) איפיינו 9 ערכים את המטריצה, הרי שהפעם (צד ימין) 5 בלבד מאפיינים אותה. אם כל איבר במטריצה הנ"ל הוא עצמו מטריצה בלוק-סיבובית, הרי שהמבנה המתקבל הוא מבנה של מטריצת בלוק-סיבובית.

והפעם יצא כי המטריצות הקטנות המתארות פעולות חד-מימדיות הינן סיבוביות בעצמן. מטריצות סיבוביות או מטריצות בלוק-סיבוביות מקיימות את תכונת ההחלפה. הדוגמה הבאה ממחישה תכונה זו.

דוגמה 2.11: ניקח שתי מטריצות סיבוביות ונכפול אותן זו בזו בשני הכיוונים.

$$\begin{pmatrix} 3.0 & 1.0 & 0.8 & 0.0 \\ 0.0 & 3.0 & 1.0 & 0.8 \\ 0.8 & 0.0 & 3.0 & 1.0 \\ 1.0 & 0.8 & 0.0 & 3.0 \end{pmatrix} \begin{pmatrix} 2.0 & 1.0 & 0.5 & 0.0 \\ 0.0 & 2.0 & 1.0 & 0.5 \\ 0.5 & 0.0 & 2.0 & 1.0 \\ 1.0 & 0.5 & 0.0 & 2.0 \end{pmatrix} = \begin{pmatrix} 6.4 & 5.0 & 4.1 & 1.3 \\ 1.3 & 6.4 & 5.0 & 4.1 \\ 4.1 & 6.3 & 6.4 & 5.0 \\ 5.0 & 4.1 & 1.3 & 6.4 \end{pmatrix}$$

$$\begin{pmatrix} 2.0 & 1.0 & 0.5 & 0.0 \\ 0.0 & 2.0 & 1.0 & 0.5 \\ 0.5 & 0.0 & 2.0 & 1.0 \\ 1.0 & 0.5 & 0.0 & 2.0 \end{pmatrix} \begin{pmatrix} 3.0 & 1.0 & 0.8 & 0.0 \\ 0.0 & 3.0 & 1.0 & 0.8 \\ 0.8 & 0.0 & 3.0 & 1.0 \\ 1.0 & 0.8 & 0.0 & 3.0 \end{pmatrix} = \begin{pmatrix} 6.4 & 5.0 & 4.1 & 1.3 \\ 1.3 & 6.4 & 5.0 & 4.1 \\ 4.1 & 6.3 & 6.4 & 5.0 \\ 5.0 & 4.1 & 1.3 & 6.4 \end{pmatrix}$$

כפי שאנו רואים, התוצאה זהה. יתרה מכך, מתקבל כי מטריצת התוצאה אף היא סיבובית. עובדה זו מתלכדת עם כך ששרשור מערכות המתארות קונבולוציה ציקלית הינן יחדיו קונבולוציה ציקלית חדשה.

2.6 התמרת פוריה דיסקרטית (DFT) לאותות בתמך סופי

כהמשך ישיר לתיאור התמרת הפוריה הרציפה, וניוונה לטיפול באותות דיסקרטיים, נעבור כעת לדון באותות דיסקרטיים בעלי תמך סופי. נתונה לנו תמונה $f[m,n]$ מעל התמך $\Omega = \{[m,n] | 0 \leq m \leq M-1, 0 \leq n \leq N-1\}$. שימוש ישיר בהגדרת התמרת פוריה לאותות דיסקרטיים, וניצול העובדה שהתמך סופי, מניבה

$$F(\theta_1, \theta_2) = F_D \{f[m,n]\} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f[m,n] \exp\{-j(m\theta_1 + n\theta_2)\}$$

וההתמרה ההפוכה נתונה ע"י

$$f[m,n] = F_D^{-1} \{F(\theta_1, \theta_2)\} = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} F(\theta_1, \theta_2) \exp\{j(m\theta_1 + n\theta_2)\} d\theta_1 d\theta_2$$

ובכל הנאמר לעיל אין כל חדש, זוהי חזרה על טיפול באותות דיסקרטיים ללא תמך סופי. אבל, נוצר כאן תהליך מעניין שניתן לניצול: בתחילה, לאותות רציפים נוצרה התמרה רציפה. ברור כי אות דיסקרטי מכיל פחות מידע מאות רציף, ולכן אך טבעי הוא שנקבל פשוט מה בהתמרת הפוריה שלו. ואמנם, התמרת פוריה של אות דיסקרטי נמצאה להיות מחזורית, כך שעלינו להתעניין בה רק בריבוע $[-\pi, +\pi] \times [-\pi, +\pi]$. כעת, עברנו לדבר על אותות דיסקרטיים בעלי תמך סופי שפירושו כי ישנם מספר סופי וידוע של דגימות המגדירות לחלוטין את האות, ולכן גם את ההתמרה שלו. עם זאת, ההתמרה נותרה בינתיים כפונקציה רציפה מעל הריבוע $[-\pi, +\pi] \times [-\pi, +\pi]$, וברור כי ישנו כאן בזבוז. אם נגדיר את ההתמרה במקרה זה כך שיתקבלו MN ערכים בלבד בתדר, נראה כי אז שימרנו את עושרו של האות. ואמנם, מגדירים התמרת פוריה הדיסקרטית (Discrete Fourier Transform – DFT) באופן הבא

$$F[k, \ell] = F_{DFT} \{f[m,n]\} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f[m,n] \exp\left\{-j2\pi\left(\frac{mk}{M} + \frac{n\ell}{N}\right)\right\}$$

ההגדרה הנ"ל בוחרת להתעניין בערכי $F(\theta_1, \theta_2)$ רק בנקודות $[\theta_1 = 2\pi k/M, \theta_2 = 2\pi \ell/N]$. כפי שניתן לראות, לערכים $[k, \ell]$ מחוץ לתחום התמך מתקבלת מחזוריות, כך שלעניינינו, יוצא כי רק MN ערכים מגדירים במלואה את ההתמרה. ההתמרה ההפוכה נתונה ע"י

$$f[m,n] = F_{DFT}^{-1} \{F[k, \ell]\} = \frac{1}{NM} \sum_{k=0}^{M-1} \sum_{\ell=0}^{N-1} F[k, \ell] \exp\left\{j2\pi\left(\frac{mk}{M} + \frac{n\ell}{N}\right)\right\}$$

כשדנים ב-DFT מקובל להשתמש בסימון $W_N^k = \exp\left\{j \frac{2\pi k}{N}\right\}$. אז ההתמרה וההתמרה ההפוכה נתונות

ע"י

$$F[k, \ell] = F_{\text{DFT}}\{f[m, n]\} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f[m, n] W_M^{-km} W_N^{-n\ell}$$

$$f[m, n] = F_{\text{DFT}}^{-1}\{F[k, \ell]\} = \frac{1}{NM} \sum_{k=0}^{M-1} \sum_{\ell=0}^{N-1} F[k, \ell] W_M^{km} W_N^{n\ell}$$

מדוע MN הערכים המוצעים מספיקים להגדרת ההתמרה? מסתבר כי ההתמרה המוצעת הינה

אורתונורמלית (עד כדי קבוע), דהיינו, ברישום מטריצי מהצורה

$$\begin{bmatrix} F[0,0] \\ F[1,0] \\ \vdots \\ F[M-1, N-1] \end{bmatrix} = \begin{bmatrix} W_{MN}^0 & W_{MN}^0 & \dots & W_{MN}^0 \\ W_{MN}^0 & W_{MN}^1 & \dots & W_{MN}^{N-1} \\ \vdots & \vdots & \ddots & \vdots \\ W_{MN}^0 & W_{MN}^{M-1} & \dots & W_{MN}^{(M-1)(N-1)} \end{bmatrix} \begin{bmatrix} f[0,0] \\ f[1,0] \\ \vdots \\ f[M-1, N-1] \end{bmatrix}$$

$$\Rightarrow \underline{F} = \underline{W} \underline{f}$$

מטריצת ההתמרה הינה מטריצה יוניטרית (שוב – חסר הקבוע). לכן, היפוכה קיים ופשוט (נתון ע"י שחלוף וצמוד). לכן, סידרת MN ערכי ההתמרה מתארים את MN דגימות הכניסה באופן מלא. מטריצה זו בגודל MN על MN מלכסנת מטריצות בלוק סיבוביות, כלומר, אם \underline{H} מטריצה בלוק סיבובית, נקבל כי $\underline{H} \underline{W} \underline{W}^*$ היא מטריצה אלכסונית. מדוע זה מעניין? ובכן, נניח כי \underline{H} מייצגת מטריצת תגובה להלם של מערכת ליניארית וקבועה במקום הפועלת כקונבולוציה ציקלית. אזי, מוצא המערכת לתמונת כניסה \underline{f} נתונה ע"י $\underline{g} = \underline{H} \underline{f}$. כיוון שהמטריצה \underline{H} בלוק-סיבובית, נקבל כי מתקיים הקשר הבא

$$\underline{g} = \underline{H} \underline{f} \Rightarrow [\underline{H} \underline{W} \underline{W}^*][\underline{W} \underline{f}] = \underline{W} \underline{g} \Rightarrow \Delta_H \underline{F} = \underline{G}$$

כאשר המטריצה Δ_H מטריצה אלכסונית. לכן, כל BIN בהתמרת התדר בין כניסה ליציאה קשורים ביניהם ע"י מכפלה בסקלר שהוא תגובת התדר בנקודה זו – בדיוק כפי שקיבלנו במקרה הרציף.

המטריצה \underline{W} המבצעת התמרת פוריה ניתנת לכתיבה כמכפלת קרונקר של שתי מטריצות קטנות יותר, כשכל אחת מבצעת התמרת פוריה חד-מימדית. בזה אין כל חדש כי אנו יודעים כי אפילו התמרת הפוריה הרציפה הינה פרידה. הייחוד הפעם הוא בכך שפרידות זו מיתרגמת למכפלת קרונקר, כיוון שלראשונה אנו עוסקים בייצוג מטריצי להתמרה. בשלב זה נעצור את הדיון. ניתן כמובן להמשיך ולתאר את תכונות ה-DFT, ייצוגו כפעולה מטריצית, ועוד. אנו נשוב לדון בנושאים אלו בפרק הדן בהתמרות דיסקרטיות בדו-מימד – פרק 4.

פרק 3 – דגימה ושחזור של תמונות

41	<u>פרק 3 – דגימה ושחזור של תמונות</u>
41	3.1 דגימה אחידה של אותות דו-מימדיים
44	3.2 שחזור מדגימה אחידה
46	3.3 תופעת הקיפול בתמונות
51	3.4 משפטי דגימה מוכללים לדו-מימד

בפרק זה נציג בהרחבה את הנושאים דגימה ושחזור לאותות דו-מימדיים. כזכור מפרק 1, ע"י דגימה אנו עוברים מאות רציף המחייב שימוש באינסוף סיביות לשם ייצוג מלא, לאות דיסקרטי. חשוב להבהיר כי יש לבצע על תוצאת הדגימה פעולת קוונטיזציה על מנת לקבל אות המיוצג בכמות סופית (אך גדולה בד"כ) של סיביות. נושא הקוונטיזציה ייסקר במסגרת הפרק הבא.

3.1 דגימה אחידה של אותות דו-מימדיים

נתונה לנו תמונה המיוצגת כאות דו מימדי $f(x,y)$. רצוננו לדגום תמונה זו במרווחים D_x ו- D_y

ולקבל תמונה דגומה

$$f[m,n] = f(x,y) \Big|_{\substack{x = mD_x \\ y = nD_y}} = f(mD_x, nD_y)$$

השאלה המרכזית בה נעסוק במסגרת סעיף זה היא אילו תנאים יש להציב על המרווחים D_x ו- D_y כך שניתן יהיה לשחזר את $f(x,y)$ מתוך דגימותיו $f[m,n]$? לצורך מענה לשאלה זו, נעשה שימוש בתיאור מתמטי נוח יותר לפעולת הדגימה – דגימה ע"י מכפלה במסרק הלמים. נקבע כי האות נדגם ע"י המכפלה

$$\tilde{f}[m, n] = f(x, y) \cdot \sum_m \sum_n \delta(mD_x - x, nD_y - y)$$

בפרק קודם עשינו שימוש בביטוי זה בדיוק (עם קביעה שרירותית של מרווחי הדגימה להיות 1) בבואנו להסביר את המקור להגדרת התמרת הפוריה לאותות דיסקרטיים. כעת נבצע מניפולציה פשוטה אחרת – כיוון שהאות בנוי ממכפלה של שני אותות רציפים בעקרון, התמרת הפוריה שלו היא קונבולוציה של ההתמרות. לכן

$$\begin{aligned} F\{\tilde{f}[m, n]\} &= F\left\{f(x, y) \cdot \sum_m \sum_n \delta(mD_x - x, nD_y - y)\right\} = \\ &= F\{f(x, y)\} \otimes F\left\{\sum_m \sum_n \delta(mD_x - x, nD_y - y)\right\} = \\ &= \frac{1}{D_x D_y} F(u, v) \otimes \sum_m \sum_n \delta\left(u - \frac{m}{D_x}, v - \frac{n}{D_y}\right) = \\ &= \frac{1}{D_x D_y} \sum_m \sum_n F\left(u - \frac{m}{D_x}, v - \frac{n}{D_y}\right) \end{aligned}$$

ההתמרה על סידרת ההלמים מוכרת בגרסתה החד-מימדית בשם נוסחת פואסון. מסרק ההלמים החד-מימדי יוצר את המסרק הדו-מימדי ע"י מכפלה ספרבילית:

$$\begin{aligned} \text{Comb}(x, y, D_x, D_y) &= \sum_m \sum_n \delta(mD_x - x, nD_y - y) = \text{Comb}(x, D_x) \cdot \text{Comb}(y, D_y) \\ \Rightarrow F\{\text{Comb}(x, y, D_x, D_y)\} &= \frac{1}{D_x D_y} \text{Comb}(u, v, D_x^{-1}, D_y^{-1}) \end{aligned}$$

קיבלנו כי הפונקציה $F(u, v)$ משוכפלת שוב ושוב במרווחים קצובים בשני הצירים - $(1/D_x, 1/D_y)$. ציור 3.1 ממחיש תופעה זו. מצויר זה גם ניכרת התכונה הבאה – כאשר $F(u, v)$ חסום פס לתחום המלבן המשוכפל, מתקבל כי השכפולים לא דורכים זה על זה. נניח כי אמנם $F(u, v)$ מקיים

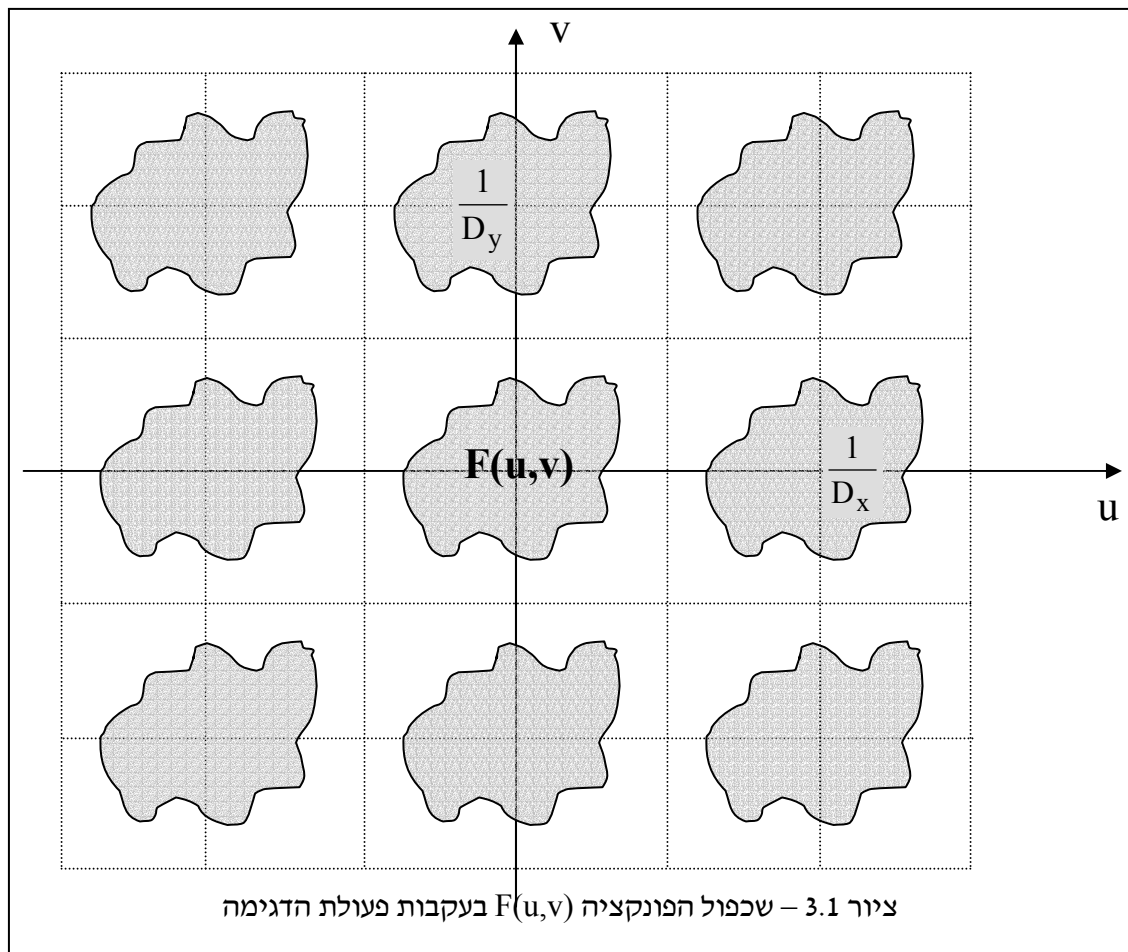
$$F(u, v) = \begin{cases} F(u, v) & |u| \leq U_{\max}, |v| \leq V_{\max} \\ 0 & \text{otherwise} \end{cases}$$

אזי, על מנת שהשכפולים לא יעלו זה על זה נדרוש

$$U_{\max} < \frac{1}{2D_x} \Rightarrow D_x < \frac{1}{2U_{\max}}$$

$$V_{\max} < \frac{1}{2D_y} \Rightarrow D_y < \frac{1}{2V_{\max}}$$

הדרישה הזו אינה זרה לנו – פירושה הוא שעלינו לדגום מספיק צפוף כדי לכלול כמות מספקת של דגימות לייצוג מלא של האות הדו-מימדי.



מדוע חשוב לנו למנוע מהשכפולים לעלות זה על זה? ובכן, בהינתן האות הדגום $\tilde{f}[m,n]$ (שכל דגימה בו היא פונקציית הלם), זהו אות רציף הניתן לסינון. ע"י הפעלת מסנן ריבועי LPF מהצורה

$$H(u, v) = \begin{cases} D_x D_y & |u| \leq \frac{1}{2D_x}, |v| \leq \frac{1}{2D_y} \\ 0 & \text{otherwise} \end{cases}$$

על אות זה נקבל בדיוק את אות המקור, כיוון שהסינון הינו מכפלה בתדר, וע"י ההכפלה במסנן הנ"ל אנו מבטלים את כל השכפולים למעט זו הראשית, וכיוון שהיא בדיוק ההתמרה של אות המקור, הרי שביצענו בשלמות פעולה הפוכה לפעולת הדגימה.

מבחינה מעשית, לא נתון לנו האות $\tilde{f}[m,n]$, אלא אות דגום $f[m,n]$. לכן השאלה המתבקשת

היא כיצד בפועל מבצעים את הסינון המשחזר? נענה לשאלה זו ע"י כיתוב מפורש של פעולת השחזור

$$\begin{aligned} \hat{f}(x,y) &= F^{-1}\{H(u,v)F\{\tilde{f}[m,n]\}\} = \tilde{f}[m,n] \otimes h(x,y) = \\ &= \int_{\alpha} \int_{\beta} \sum_m \sum_n f[m,n] \delta(\alpha - mD_x, \beta - nD_y) h(x - \alpha, y - \beta) d\alpha d\beta = \\ &= \sum_m \sum_n f[m,n] h(x - mD_x, y - nD_y) = \sum_m \sum_n f[m,n] \operatorname{sinc}\left[\frac{x - mD_x}{D_x}\right] \operatorname{sinc}\left[\frac{y - nD_y}{D_y}\right] \end{aligned}$$

כלומר, השחזור נעשה ע"י צירוף ליניארי של פונקציות sinc דו-מימדיות, המוגדרות ע"י

$$\operatorname{sinc}(x) = \frac{\sin(\pi x)}{\pi x}$$

הניתוח שהוצג לעיל אינו אלא הרחבה טבעית של משפט הדגימה לדו-מימד, לפיו אות העובר דגימה אחידה בשני הצירים, ובמקביל להם, מרווחי הדגימה הנדרשים צריכים להיות קטנים מהיפוכו של פעמיים התדר המירבי בכל ציר.

3.2 שחזור מדגימה אחידה

פעולת השחזור נועדה לקחת אות דגום ולחזור אל הרצף. פעולה זו נדירה למדי שכן מעט מאוד יכול להיעשות על תמונה ברצף (אפילו לא הצגתה). עם זאת, ניתן להתייחס לפעולת השחזור כניסיון להגדיל רזולוציה, דהיינו, לקחת תמונה דגומה ולייצר דגימות נוספות כך ששריג הדגימה הסופי יהיה צפוף יותר וייתן תחושה של רצף. תהליך של יצירת תמונה גדולה יותר אך בעלת מספר גדול יותר של דגימות תיקרא אינטרפולציה.

שחזור מדגימה בנקודה כלשהי עפ"י נוסחת השחזור שהתקבלה קודם לכן מחייב ביצוע צירוף ליניארי המערב את כל אינסוף הדגימות הנתונות לנו. בפועל, כשמשחזרים חייבים לקרב פעולה זו ע"י שימוש בצירוף ליניארי בתמך מוגבל וסופי. קיימות מספר אפשרויות מעשיות, בדומה לנלמד עבור אותות חד-מימדיים.

השחזור הפשוט ביותר מציע החלפת כל דגימה בפונקציה מלבנית בגודל $D_x \times D_y$ ממורכזת

סביב הראשית, בעלת גובה קבוע ששווה לערך הדגימה. זהו ה-Zero Order Hold, אך עבור תמונות אין דרישה לסיבתיים ולכן הפונקציה סימטרית סביב הראשית. בפועל, האות הדיסקרטי (כשהוא מיוצג כמטריצת הלמים בגבהים שונים) עובר קונבולוציה עם פונקציה זו. כשתהליך זה מבוצע לאינטרפולציה לשם הגדלת התמונה פי 2 בכל ציר, נקבל כי פשוט עלינו לשכפל כל שורה וכל עמודה פעמיים – תהליך הקרוי Pixel Replication, או Nearest Neighbor, כיוון שהערכים החסרים ממולאים ע"י העתקת השכן הקרוב ביותר. ציור 3.2 ממחיש כיצד נראית תוצאת תהליך זה עבור תמונת סינוס דו-מימדי בגודל 21 על 21 שעברה הגדלה פי 10 בכל ציר.

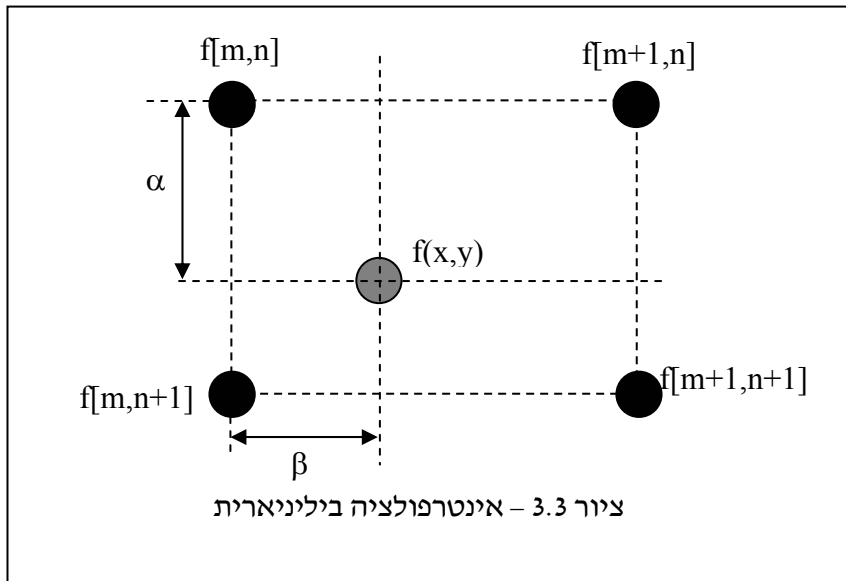


אפשרות טובה יותר היא הפונקציה המשולשת הדו-מימדית, הנבנית ממכפלת שתי פונקציות משולשות חד-מימדיות סימטריות סביב הראשית, הכופלות זו את זו ליצירת פונקציה דו-מימדית ספרבילית. הפונקציה המשולשת עבור הציר האופקי מקבלת גובה 1 בראשית ו-0 בנקודות $\{-D_x, +D_x\}$. שימוש בפונקציה זו לאינטרפולציה קרוי Bilinear Interpolation, והשימוש בו מאוד פופולרי. עבור 4 דגימות סמוכות $\{f[m,n], f[m+1,n], f[m,n+1], f[m+1,n+1]\}$ ניתן להראות כי ערך המצוי בתוך הריבוע נתון ע"י

$$f(x,y) = \begin{bmatrix} 1-\alpha & \alpha \end{bmatrix} \begin{bmatrix} f[m,n] & f[m,n+1] \\ f[m+1,n] & f[m+1,n+1] \end{bmatrix} \begin{bmatrix} 1-\beta \\ \beta \end{bmatrix}$$

כאשר α, β הם ערכים בין 0 ל-1 הנותנים את מיקום ה-x וה-y יחסית לאורך הדגימה, כפי שמתואר בציור 3.3.

תוצאת הגדלת תמונה ע"י אינטרפולציה ביליניארית נתונה בציור 3.4 על אותה דוגמה מציור 3.2. כפי שניתן לראות השיפור הוא עצום. התוצאה עבור דוגמה זו מצוינת כיוון שהתמונה המקורית חלקה מאוד, ואינטרפולציה ביליניארית ממלאת את החסר בקווים אלכסוניים הנותנים תחושה של רצף.

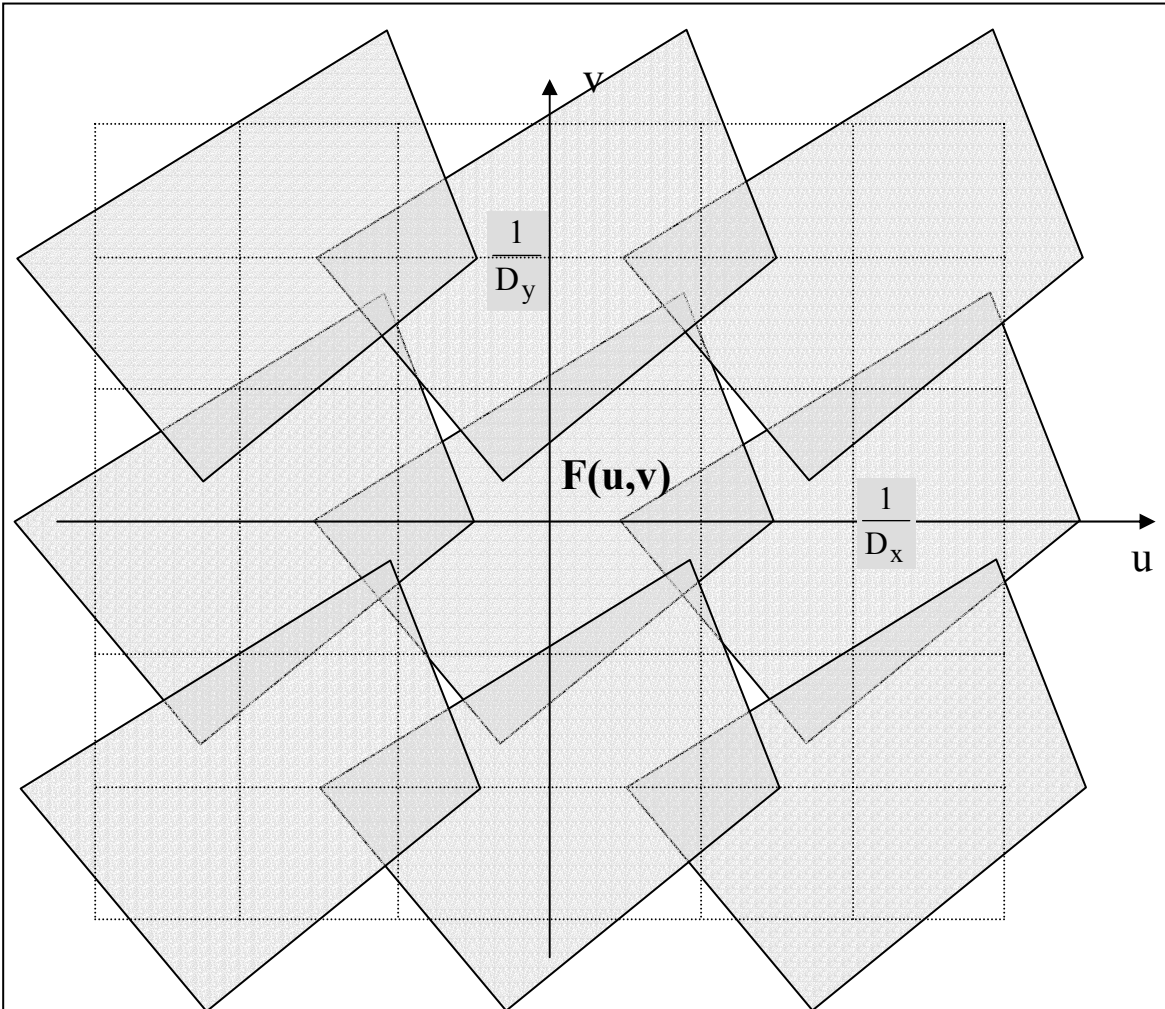


ניתן להמשיך ולתאר עקומים מסדר גבוה יותר לאינטרפולציה, אך אנו נעצור את הדיון בשלב זה. המשך ישיר של השיטות שהוזכרו כאן יובא בהמשך כשנשאל כיצד להגדיל תמונה תוך שימור שפות ופרטים.

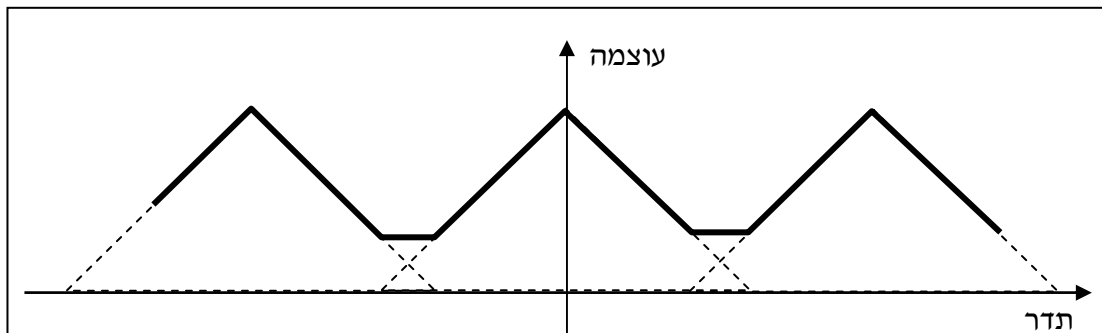
3.3 תופעת הקיפול בתמונות

תופעת הקיפול – Aliasing - הינה פלישה של השכפולים בתדר כתוצאה מדגימה, זה לתוך זה, וברור כי במקרה כזה לא ניתן לשחזר את האות המקורי ללא מידע נוסף. תופעת הקיפול מתרחשת כאשר מרווח הדגימה גדול מידי, ואינו מקיים את דרישת הסף שהוזכרה בסעיף קודם. למשל, כפי שנראה בהמשך, אות בעל תמך סופי במקום אינו יכול להיות חסום תדר. לכן, דגימתו בכל מרווח שהוא חייבת להניב את תופעת הקיפול. ציור 3.4 ממחיש מקרה עם תופעת קיפול.

באופן טבעי נטפל בתמונות ובכאלה בעלי תמך סופי, ותמיד נרצה לדגום אותן! האם פירוש הדבר שנסבול מתופעת הקיפול? ואם כן, כיצד תתבטא אותה תופעה בתמונות אלו אחר דגימה? כמה גדול הנזק שייגרם (אם בכלל)? ובכן, לאותות מעשיים לא ניתן להניח התנהגות חסומת תדר. אולם, בהחלט ניתן להניח כי העוצמה של התמרת הפוריה דועכת ואפילו מהר כשמתרחקים מראשית הצירים (ה-DC). לכן, גם אם חלה פלישה של השכפולים אין פירוש הדבר נזק גדול בהכרח, כיוון שפלישתו של זנב ההתמרה תשנה רק במקצת, כפי שמתואר למקרה חד-מימדי בציור 3.5.

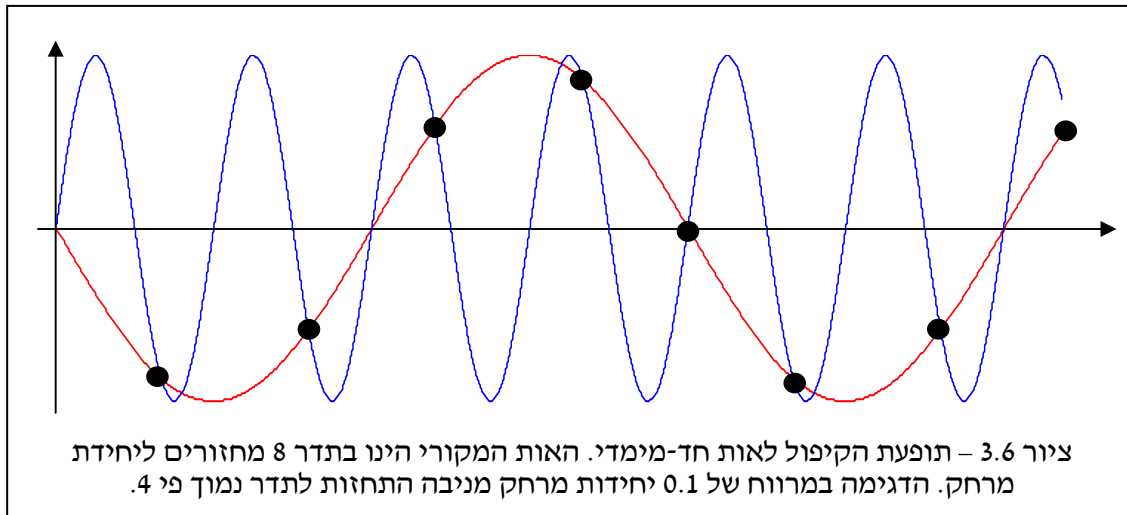


ציור 3.4 – דגימה עם מרווח לא מספק ותופעת הקיפול



ציור 3.5 – תופעת הקיפול כתוספת בזנבות ההתמרה

אם בכל זאת הזנב אינו כה חלש אנרגטית, השפעתו תתבטא בתמונה הדגומה, ותתקבלנה תופעות המוכרות שהיבטים שונים של תופעת הקיפול. תופעת הקיפול הינה במהותה תופעת התחזות בה תדר מרחבי מעל הסף האפשרי מתחזה לתדר נמוך המצוי בתמונת ראי לו. לשם דוגמה, אם הדגימה נעשית במרווח של 0.1 (ביחידות אורך כלשהן), פירוש הדבר שהתדר המירבי המותר לאות הוא 5 מחזורים ליחידת אורך. נניח כי האות הנדגם כלל גם מרכיב בתדר 8 מחזורים בשניה. תדר זה יופיע בדגימה כשהוא מתחזה ל- 2 מחזורים לשניה (תמונת הראי סביב 5 לערך 8). ציור 3.6 ממחיש תהליך זה.



בציור 3.7 נתונה תמונה דיסקרטית אשר נדגמה בצפיפות מספקת (כלומר, אין בה תופעת קיפול), ולצידה תת-דגימות שלה ע"י דילול פשוט ביחסים שלמים. לדוגמה, לקיחת כל פיקסל שני מכל שורה שניה יוצר תמונה קטנה פי ארבע המתייחס למרווח דגימה כפול. התמונות השונות הינן בגודל שונה, אך מוצגות בגודל אחיד ע"י מתיחה פשוטה הגורמת לשכפול פיקסלים. כפי שניתן לראות על פניו של הבית בצד שמאל של התמונה, נוצר תדר שונה לכל מרווח – אפקט קיפול קלאסי. אפקט נוסף ניתן לאבחנה בגדר, שם ניכר כי איזורים יוצאים במרווח מסויים מטושטשים ובמרווחים גדולים יותר חוזרים להיות פעילים – גם אפקט של קיפול. מלבד שתי תופעות לאה ניכרת גם תופעת שינון של שפות חדות המצויות בזוית אלכסונית כמו במגדלור ובגנות.

תופעת הקיפול היא תופעה מטרידה כשהיא בעוצמה חזקה. על מנת להימנע ממנה בתהליך הדגימה, נדרש להפעיל מסנן מעביר נמוכים (LPF) אשר יהפוך את התמונה הנתונה לחסומת תדר באופן אפקטיבי, ורק לאחר מכן לבצע את הדילול. ציור 3.8 ממחיש אפקט זה. הפעם חזרנו על הניסוי של ציור 3.7, כשלפני פעולת הדילול ביצענו סינון. במסגרת זו לא נאמר מהו המסנן שהופעל – דיון בנושא תכנון מסנן נאות יינתן במסגרת פרק אחר. נאמר רק כי ברור כי על מסנן זה להעביר טוב ככל האפשר את התדרים הנמוכים עד סף מסוים, ולחסום טוב ככל האפשר מתדר סף זה. כמו כן ברור כי בשל אופייה של הדגימה, המסנן הנדרש ניתן לתכנון כמסנן ספרבילי, כשתדר הסף בכל ציר עשוי להיות שונה. כפי שניתן לראות כי כל התופעות שהוזכרו נעלמו או דוכאו בצורה משמעותית.



ציור 3.7 – המחשת תופעת הקיפול: למעלה בצד ימין – תמונת מקור. למעלה בצד שמאל – דגימה במרווח כפול. למטה בצד שמאל – דגימה במרווח גדול פי 3. למטה בצד ימין – דגימה במרווח גדול פי ארבע.



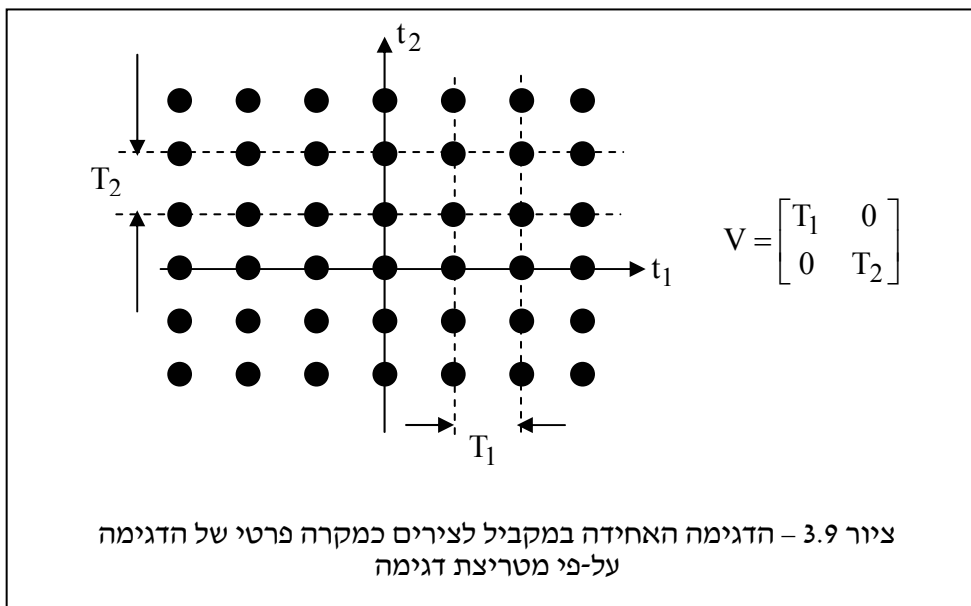
ציור 3.8 – חזרה על ציור 3.7 כשלפני פעולת הדילול מבוצע מסנן LPF נאות לקיצוץ התדרים הגבוהים

3.4 משפטי דגימה מוכללים לדו-מימד

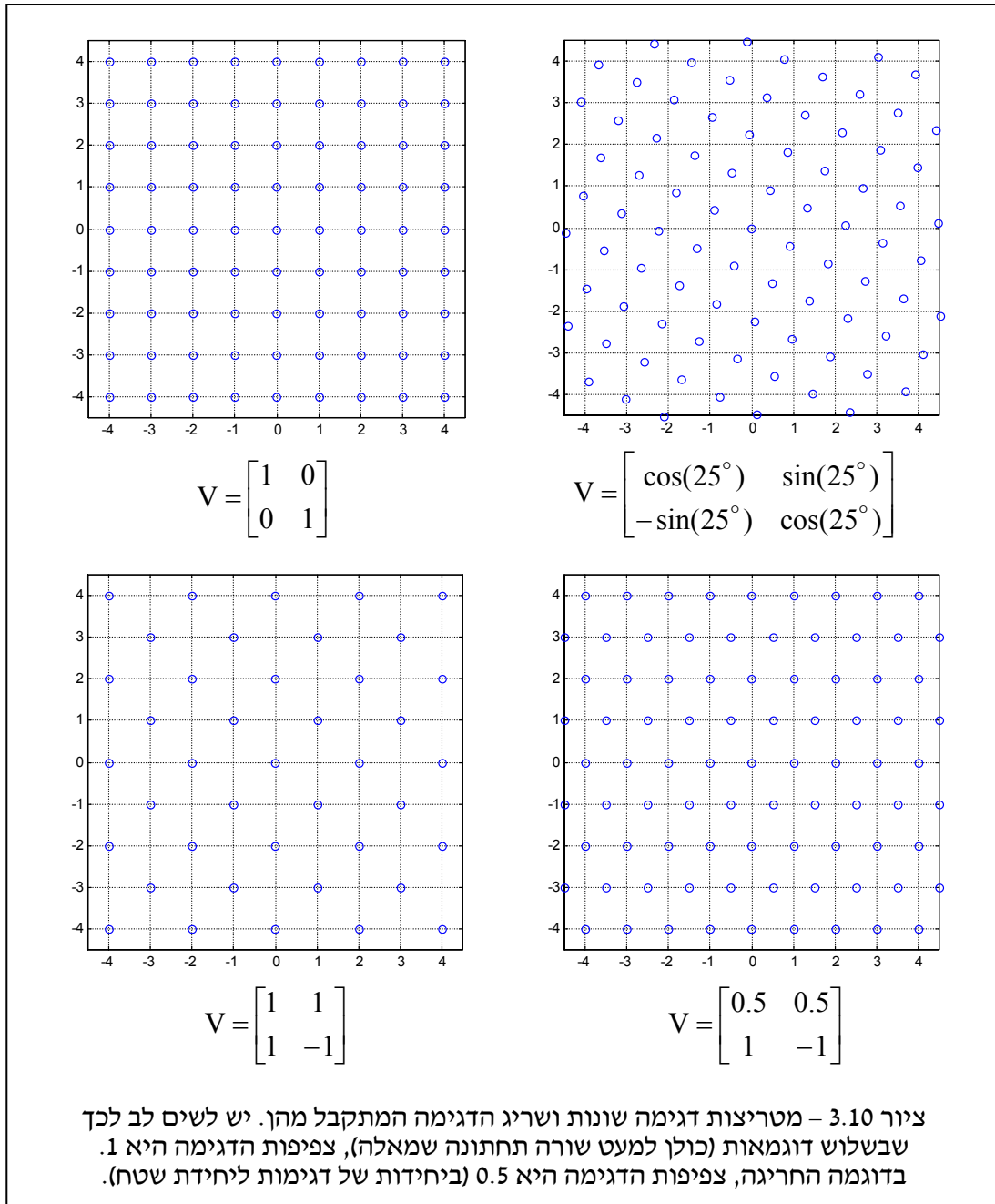
הכללה טבעית וישירה למשפט הדגימה שתואר בסעיפים קודמים מוצעת באופן הבא: ניקח שני וקטורים בלתי תלויים ליניארית v_1 ו- v_2 , במימד 2. משני הוקטורים נבנית מטריצה V כך שהוקטורים הנ"ל הם עמודותיה. נציע לדגום את המרחב בנקודות

$$\begin{bmatrix} t_1 \\ t_2 \end{bmatrix} = \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix} \begin{bmatrix} n_1 \\ n_2 \end{bmatrix} \quad n_1, n_2 \in \mathbf{N}$$

כלומר, השריג הרגיל של המספרים הטבעיים מומר ליניארית לנקודות $[t_1, t_2]$. שריג הדגימה המתקבל מסודר במרווחים אחידים אך עם סיבוב ביחס למישור ועובר עיוות בין הצירים. המקרה הטריטוריאל של דגימה אחידה מתואר בציור 3.9. מספר דוגמאות של המטריצה V והשריג הנובע ממנה מוצגים בציור 3.10 ניתן להראות כי צפיפות הדגימה נתונה ע"י היפוך ערכו המוחלט של דטרמיננט המטריצה V . זאת כיוון שהוקטורים v_1 ו- v_2 יוצרים מעויין במישור אשר שטחו הוא בדיוק הדטרמיננט. במקרה של דגימה אחידה ומקבילה לצירים מתקבל כי הדטרמיננט הוא $T_1 T_2$ ופירוש הדבר הוא שיש דגימה אחת לכל תא שטח של $T_1 T_2$. בכל מקרה אחר בוא הדטרמיננט הוא $|V|$, יש דגימה אחת לכל תא שטח בגודל $|V|$.



במקרה הכללי של מטריצה V , מתקבל שריג הלמים בעל מבנה מסודר המתאר את הדגימה. בדומה לתהליך שבוצע עבור הדגימה האחידה, אם נייצג את האות הדגום כמכפלה בשריג מתאים של הלמים נקבל



$$\tilde{f}[\underline{t}] = f(x, y) \cdot \sum_{\underline{n}} \delta(\underline{V}\underline{n}) = f(x, y) \cdot \text{Comb}(x, y, \underline{V})$$

כל שעשינו הוא להכליל את מושג המסרק. לכן, כמקודם, כל שעלינו לדעת היא ההתמרה של המסרק הכללי הנ"ל. מתקבל (ואנו לא נביא את ההוכחה)

$$\begin{aligned} F\{\tilde{f}[\underline{m}, \underline{n}]\} &= F\{f(x, y) \cdot \text{Comb}(x, y, \underline{V})\} = F(u, v) \otimes F\{\text{Comb}(x, y, \underline{V})\} = \\ &= \frac{1}{|\underline{V}|} F(u, v) \otimes \text{Comb}(u, v, \underline{V}^{-T}) = F(\underline{\Omega}) \end{aligned}$$

וכך עשינו שימוש בהכללה של נוסחת פואסון. קל לראות כי המקרה הפרטי המוכר לנו של דגימה אחידה ומקבילה לצירים מסתדר היטב עם התוצאה הכללית שניתנה. במקרה הכללי תתקבל מחזוריות מהצורה

$$F(V^T \Omega) = F(V^T [\Omega + U \underline{k}])$$

וכבעבר, יש לדאוג לצפיפות דגימה כזו שתבטיח כי השכפולים השונים לא עולים זה על זה, למניעת תופעת הקיפול. השחזור פועל בדרך דומה – יש לבדוד החוצה את השכפול הראשי המצוי סביב הראשית. צורת המסנן במקרה הקודם הייתה מלבנית – עובדה שהביאה לפונקציית שחזור מבוססת sinc. הפעם, צורת המסנן תלויה בצורת השריג (ולא באופן פשוט).

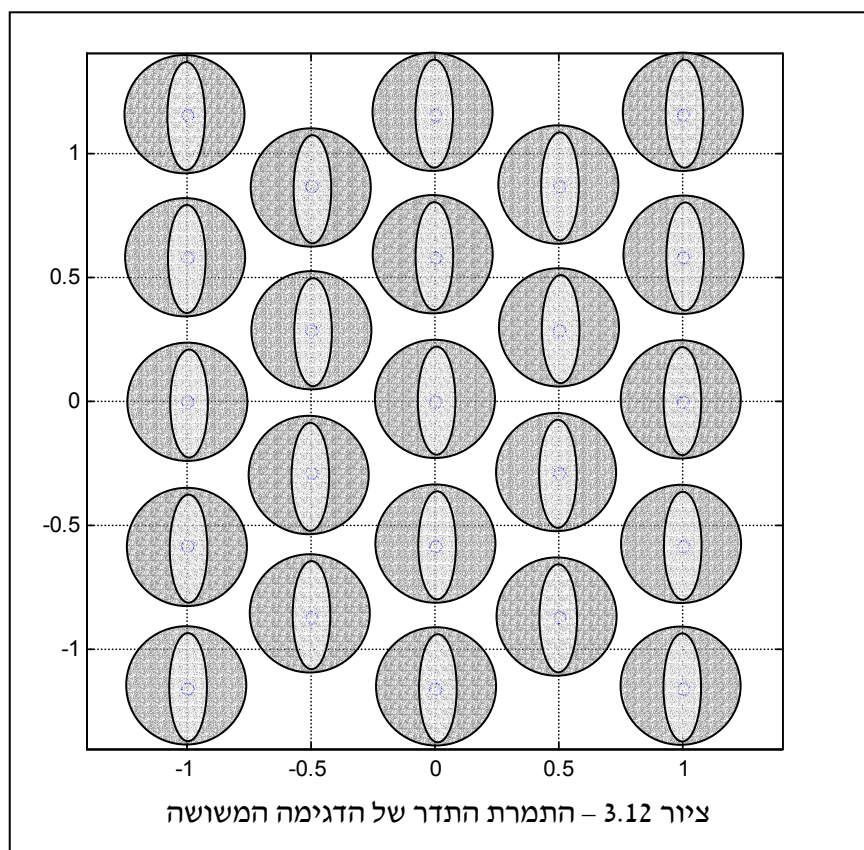
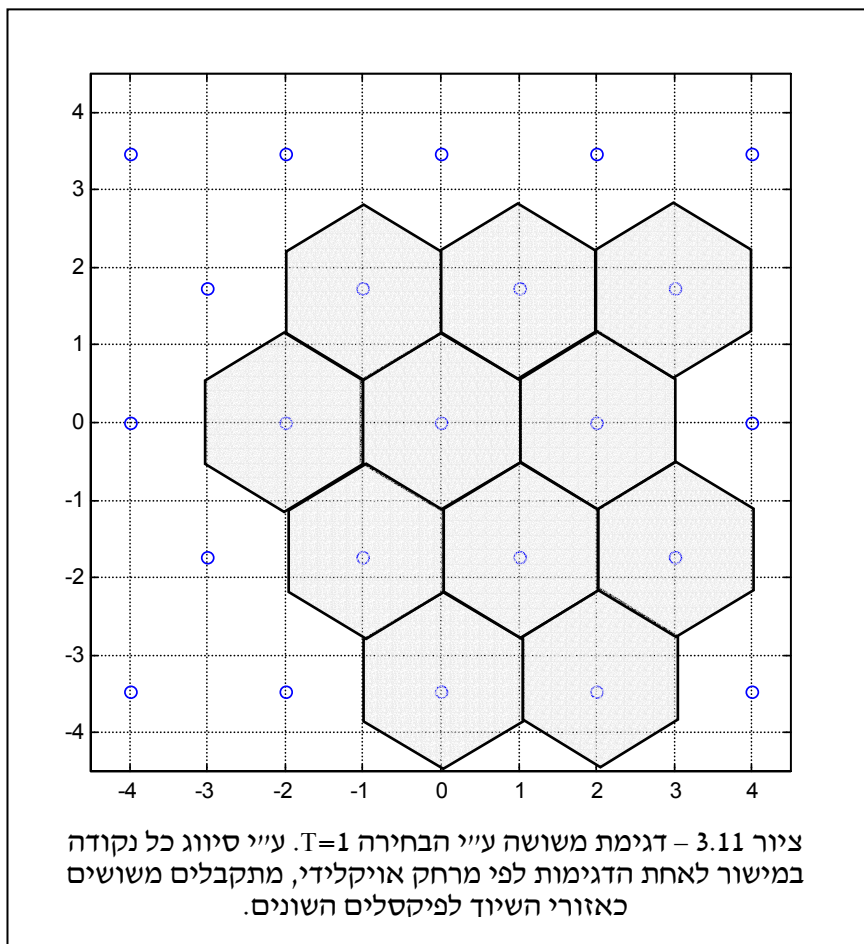
מדוע חשוב להכיר את ההכללה שניתנה כאן? ובכן, חשוב להבהיר כי יש אינסוף מטריצות דגימה אפשריות. ע"י עיון בהתמרת הפוריה של האות הנתון ניתן להתאים מטריצת דגימה אופטימלית. מטריצה זו תאופיין בצפיפות הדגימה המינימלית, וכך ייצא שחסכנו בכמות הסיביות הנחוצות לאיפיונו המלא של האות. בפועל, דגימה כללית מעין זו המתוארת כאן אינה קלה לביצוע, והדגימות היחידות המקובלות הן דגימה אחידה מקבילה לצירים או דגימה משושה. דגימת משושה מתקבלת ע"י הבחירה

$$V = \begin{bmatrix} T & T \\ \sqrt{3}T & -\sqrt{3}T \end{bmatrix}$$

ומבנה שריג הדגימה המתאים מתואר בציור 3.11 עבור הבחירה $T=1$, ולכן, הצפיפות היא דגימה אחת לשטח של 3.461. שריג מסרק ההלמים בתדר מתואר בציור 3.12, ומתקבל ע"י המטריצה

$$V^{-T} = \begin{bmatrix} 0.5 & 0.5 \\ 0.2887 & -0.2887 \end{bmatrix}$$

והצפיפות של שריג זה הפוכה, דהיינו -0.2887 . בציור 3.12 אנו מניחים צורה כלשהי להתמרת התדר של האות המקורי $f(x,y)$ (מעגל ובתוכו אליפסה), ולכן מצליחים לתאר את האות הדגום בתדר במלואו. כפי שניתן לראות אין תופעת קיפול, כיוון שאין חפיפה בין השכפולים השונים. עבור אותו אות רציף, דגימה אחידה במקום מחייבת מרווח של 0.6 (בתדר) בין השכפולים – הן בציר x , והן בציר y . לכן, הצפיפות בתדר נדרשת להיות $0.6^2 = 0.36$, ואז נקבל שהשכפולים נוגעים זה בזה. במקום נקבל כי נדרשת דגימה אחת ל-2.8 יחידות שטח. כפי שאנו רואים דרך דוגמה זו, לאות בעל תמך מעגלי, דגימה ע"י שריג משושה יעילה יותר.



פרק 4 – קוונטיזציה לתמונות

55 <u>פרק 4 – קוונטיזציה לתמונות</u>
55 4.1 קוונטיזציה סקלרית - בסיס
59 4.2 קוונטיזר מקס-לויד
67 4.3 קוונטיזציה עם שיקולים חזותיים
71 4.4 קוונטיזציה וקטורית וקוונטיזציית צבע

בהינתן תמונה ברצף, הראינו בפרק הקודם כיצד יש לבצע עליה דגימה באופן נאות. לאחר פעולת הדגימה באה בד"כ פעולה נוספת בה ערכי הדגימה הממשיים מומרים למקבץ סופי ומוגדר מראש של ערכים אפשריים – פעולה זו קרויה קוונטיזציה. במסגרת זו אנו נעשה אבחנה בין קוונטיזציה סקלרית וקוונטיזציה וקטורית. הקוונטיזציה הסקלרית נחוצה לטיפול בתמונות שחור-לבן, והוקטורית לתמונות צבע.

חשוב לציין כי פעולת הקוונטיזציה (סקלרית או וקטורית) אינה נחלתו הבלעדית של נושא קליטת התמונה הרציפה בו דנים בהמרת התמונה הנתונה לקובץ מחשב. כפי שנראה בפרקים מתקדמים, לקוונטיזציה קשר הדוק לדחיסת אותות, וכן להמרתם לצורך הדפסה. במסגרת פרק זה נציג את העקרונות תוך שימת דגש על יישום אחד – קליטת תמונה והמרתה מרצף.

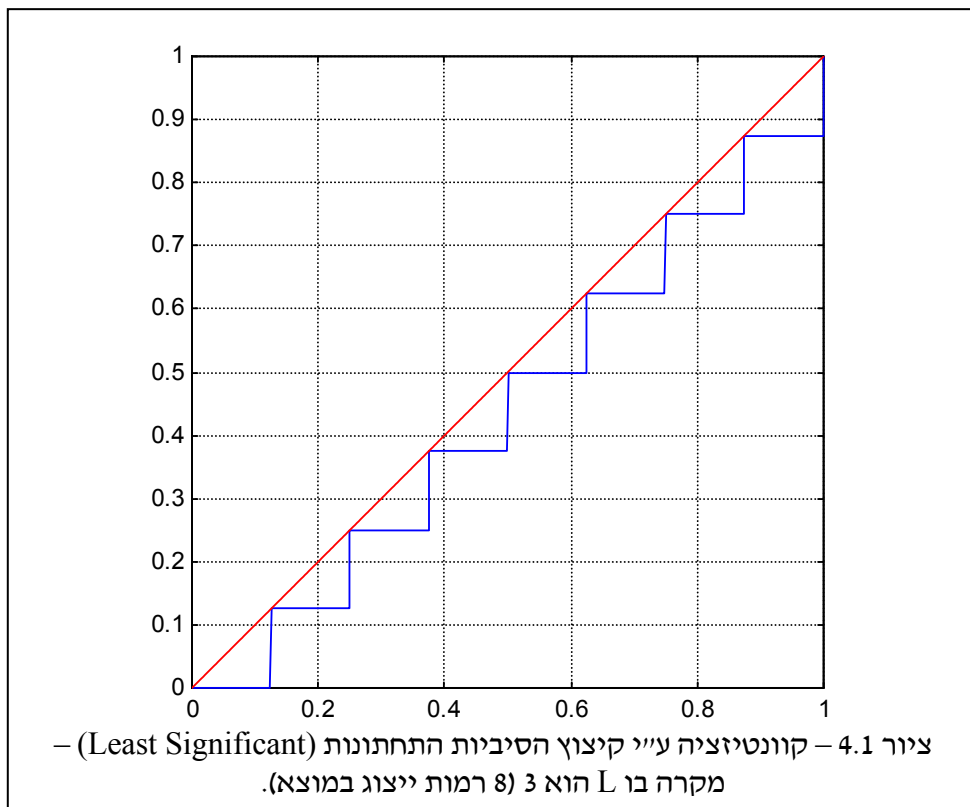
4.1 קוונטיזציה סקלרית - בסיס

נניח כי נתונה לנו תמונה דגומה אשר תסומן כ- $f[m,n]$. ערכי הדגימות הינם מעל הממשיים וטווח הערכים מצוי באינטרוול $[0,1]$. כזכור, הערך 0 מתייחס לגוון שחור מוחלט, והערך 1 לגוון לבן מוחלט. ברור כי לשם ייצוג נאות של ערכי רמות האפור נידרש לאינסוף סיביות. מוצע אם כך להשתמש ב- L סיביות (L הוא ערך קבוע מראש וערכיו הסבירים יהיו בתחום 1-16). השאלה המתבקשת היא כמובן כיצד להמיר את הערכים הרציפים לערכים הדיסקרטיים. כיוון ש- L סיביות מייצגות 2^L ערכים שונים, אנו נדרשים לבחור 2^L רמות ייצוג שונות.

אחת הפעולות הפשוטות ביותר אותן ניתן להציע היא פעולת קיצוץ לסיביות הנמוכות. בגישה זו, עלינו לייצג את רמת האפור בקבוצת L הסיביות הגבוהות ביותר (Most Significant). מתמטית, פעולה זו נעשית ע"י המשוואה

$$Q(f) = \frac{\lfloor 2^L \cdot f \rfloor}{2^L}$$

לדוגמה, עבור $L=2$, נקבל כי הערכים בתחום $[0, 0.25)$ ימופו לאפס, הערכים באינטרוול $[0.25, 0.5)$ ימופו לגובה 0.25 וכו'. באופן דומה ניתן להראות כי במוצא הקוונטיזר ייתכנו רק 4 ערכים אפשריים – 0, 0.25, 0.5 ו-0.75. לערכי הכניסה בהם חל שינוי במוצא נקרא רמות ההחלטה, ולערכי המוצא האפשריים נקרא רמות הייצוג. ציור 4.1 ממחיש את הקשר בין הכניסה למוצא של הקוונטיזר המוצע עבור $L=3$. רמות הייצוג הם מיקומי הקווים האופקיים בפונקציית המדרגות על הציר האנכי, ורמות ההחלטה הם מיקומי הקווים האנכיים על הציר האופקי. כפי שאנו רואים, הקשר מאופיין ע"י פונקציית מדרגות.

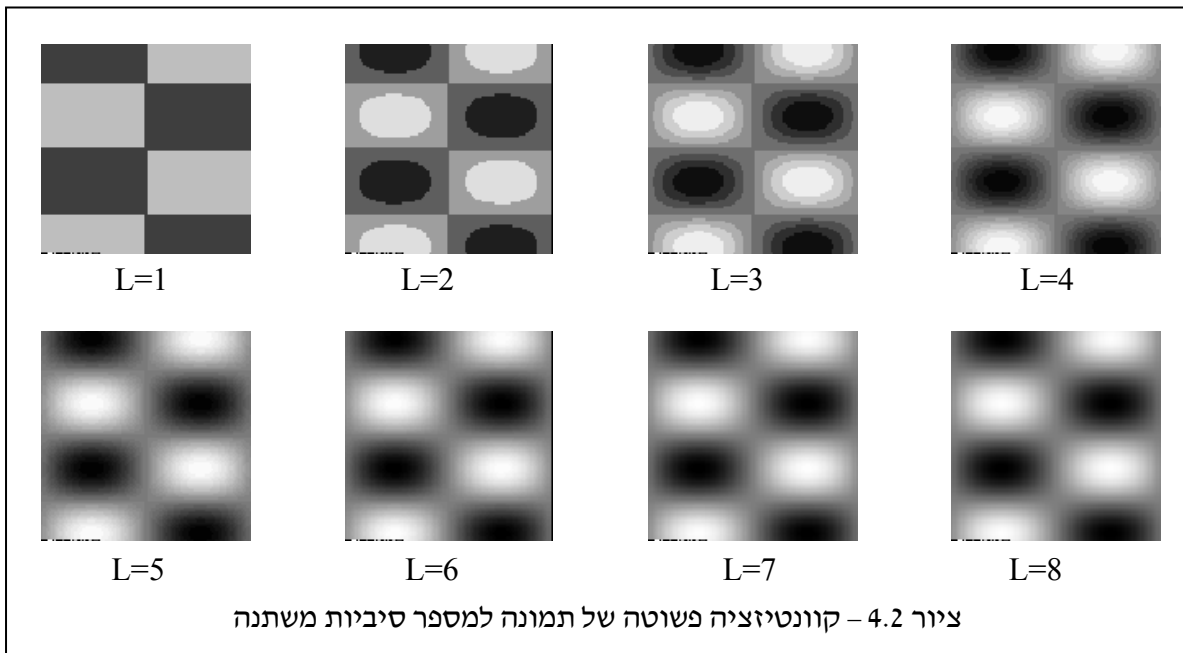


קוונטיזר מושלם היה קוונטיזר בעל מוצא זהה לכניסה, כיוון שאז אין שגיאה. ברור כי שימוש במספר סופי של סיביות לא מאפשר קוונטיזר כזה. כפי שניתן לראות, בגישה המוצעת תמיד נקבל כי מוצא הקוונטיזר נמוך מכניסתו. השגיאה בין הכניסה ליציאה תהיה אפס בנקודות המפגש בין הקו המדרג לקו האלכסוני (המייצג את מערכת היחידה), ומקסימלי בנקודות הברך שביניהן. ניתן לשפר את הפעולה הנ"ל ולהקטין את השגיאה ע"י הכנסת תוספת למוצא:

$$Q(f) = \frac{\lfloor 2^L \cdot f \rfloor}{2^L} + 2^{-(L+1)} = \frac{\lfloor 2^L \cdot f \rfloor + 0.5}{2^L}$$

התוספת תגרום לגרף המדרגות לנוע מעלה ולהתמקם סביב מערכת היחידה (הקו האלכסוני), ואמנם, השגיאה בין הכניסה ליציאה תהיה קטנה יותר.

הגישה שתוארה פשוטה למדי ולכן פופולרית. בתמונות אופייניות, ביצוע קוונטיזציה כזו ליותר מ-5 סיביות (32 רמות ייצוג) מניבה על פי רוב תוצאה בה הקוונטיזציה לא מורגשת לעיין אנושית. כשיורדים מ-5 סיביות ניתן לראות בפועל את פעולת הקוונטיזציה, ונוצרות תופעות צורמות. אנו נדון בדרכים לתיקון תופעות אלו בהמשך. ציור 4.2 מראה תמונה פשוטה (תמונת הסינוס הדו-מימדי), עם קוונטיזציה למספר סיביות משתנה, וממחישה תופעה זו. הקוונטיזציה שבוצעה משתמשת בנוסחה עם העדכנית (הכוללת את התוספת להזזת הגרף כלפי מעלה).

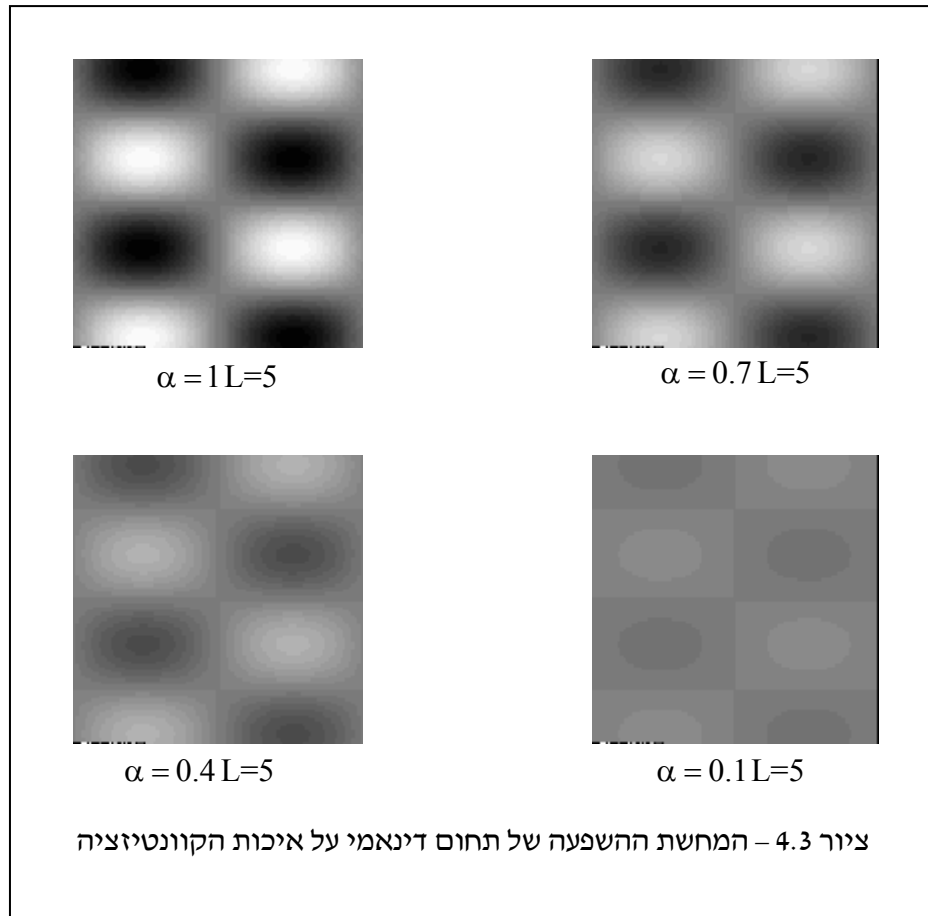


נציע כעת ניסוי מעניין. ניקח את אותה תמונה הממלאת ערכים בתחום $[0,1]$ ונכווץ את התחום הדינאמי שלה ע"י הקשר

$$\hat{f}[m,n] = \alpha[f[m,n] - 0.5] + 0.5$$

כאשר α בתחום $[0,1]$. עבור $\alpha = 0.5$ יתקבל כי התמונה כעת מקבלת ערכי רמות אפור בתחום $[0.25, 0.75]$, כלומר התחום הדינאמי של התמונה קטן. ביצוע קוונטיזציה עם כמות סיביות נאותה (נניח 5 סיביות) בגישה שתוארה לעיל תיצור בעיה. קוונטיזציה על התמונה המקורית תיתן איכות נאותה, אך

עם כיווץ התחום הדינאמי נקבל כי הקוונטיזציה מניבה תוצאה צורמת. ציור 4.3 מראה תוצאה זו עבור שימוש ב- 5 סיביות ושינוי α .



ניכר שוב כי עבור ערכי α קטנים, 5 סיביות אינן מספיקות. הסיבה לכך ברורה – עבור תחום דינאמי קטן, רק חלק מ- 32 רמות הייצוג האפשריות באות לידי ביטוי, וכל האחרות מתבזבזות! רעיון נאות יהיה, לפיכך, ביצוע קוונטיזציה שתלויה בתוכן התמונה, כך שבכל התמונות הנ"ל ינוצלו כל רמות הייצוג באופן מיטבי. זהו בדיוק הרעיון מאחורי קוונטיזר מקס-לוייד שיוצג בסעיף הבא.

הקוונטיזציה שתוארה לעיל (על שתי גרסאותיה) קרויה אחידה (יוניפורמית) כיוון שהמרווחים בין רמות הייצוג השונות או בין רמות ההחלטה השונות קבועים. נשאלת השאלה: מדוע לא לאמץ תפיסה בה הפונקציה המדרגת כללית ומאופיינת ע"י רמות החלטה וייצוג כלשהם? ואם כן, כיצד נחליט מה תהיינה הרמות הנאותות? במקרה הכללי, פעולת הקוונטיזציה תהיה פונקציה מדרגת הממפה את האינטרוול $[0,1]$ לסידרה של $N = 2^L$ גבהים שונים באופן המתואר בציור 4.4. את רמות הייצוג ורמות ההחלטה נסמן ב- $\{f_k\}_{k=0}^N, \{f_k\}_{k=1}^N$, כמתואר בציור.



4.2 קוונטיזר מקס-לויז

קוונטיזר מקס-לויז (הקרוי על שםם של שני חוקרים שפיתחוהו בנפרד ובאופן בלתי תלוי) מציע קוונטיזציה התלויה בתמונה המוזנת לה מחד, וקוונטיזציה שתרשה פונקציות מדרגות לא יוניפורמיות מאידך. הרעיון המרכזי הוא להציע את הקוונטיזר הכי טוב במובן זה שהמרחק בין כניסתו למוצאו יהיה הקטן ביותר האפשרי, בכפוף לכך שמופעלת פונקציות מדרגות על הכניסה. המרחק נמדד כשגיאה ריבועית ממוצעת – מדד נוח מתמטית שלא תמיד מתלכד עם תכונות מערכת הראיה. הנעלמים בבעיה הם הפרמטרים של פונקציות המדרגות, דהיינו, רמות ההחלטה ורמות הייצוג. בניסוח מתמטי זה ייראה כך:

$$\text{Minimize}_{\{f_k, r_k\}_k} \frac{1}{|\Omega|} \sum_{m, n \in \Omega} (f[m, n] - Q\{f[m, n]\})^2$$

בניסוח המתואר קשה להאמין שהבעיה בת פתרון מעשי. עם זאת ע"י מניפולציות פשוטות יחסית אנו נראה כי הבעיה פשוטה לאין ערוך מהנדמה. נגדיר את הקבוצה Ω_k כאוסף כל הנקודות $[m, n]$ בהן רמת האפור היא בתחום $[r_{k-1}, r_k)$, דהיינו:

$$\forall k = 1, 2, 3, \dots, N: \Omega_k = \{[m, n] \mid r_{k-1} \leq f[m, n] < r_k\}$$

נוכל לפרק את הסכימה בביטוי שהוגדר ל- N נתחים באופן הבא :

$$\sum_{m,n \in \Omega} (f[m,n] - Q\{f[m,n]\})^2 = \sum_{k=1}^N \sum_{m,n \in \Omega_k} (f[m,n] - Q\{f[m,n]\})^2$$

אבל, ברור כי כל ערכי הקוונטיזציה של $f[m,n]$ השייכים ל- Ω_k יהיו הערך הקבוע f_k , ולכן הביטוי הופך להיות :

$$\sum_{m,n \in \Omega} (f[m,n] - Q\{f[m,n]\})^2 = \sum_{k=1}^N \sum_{m,n \in \Omega_k} (f[m,n] - f_k)^2$$

נתבונן בביטוי המופיע בסכימה הפנימית. ערכי $f[m,n]$ בו ינועו בתחום (r_{k-1}, r_k) . לערך מסויים באינטרוול זה, r^* , כל הנקודות $[m,n]$ המקבלות ערך זה יניבו אותו ערך בדיוק - $(r^* - f_k)^2$, וערך זה יוכפל במספר הנקודות $[m,n]$ שערכן הוא r^* - מספר זה יסומן ב- $P(r^*)$. לכן, ניתן לסכום את הביטוי הפנימי באופן אחר, בו אנו נעים על כל הערכים בתחום (r_{k-1}, r_k) בקפיצות אינפיטיסימליות, וכנגד כל ערך לחבר את הביטוי הנ"ל כשהוא מוכפל בכמות הנקודות הנאותה. כיוון שנעים על פני ערכים ממשיים בקפיצות אינפיטיסימליות, הסכימה הינה אינטגרציה מהצורה :

$$\sum_{m,n \in \Omega_k} (f[m,n] - f_k)^2 = |\Omega| \cdot \int_{r_{k-1}}^{r_k} (r - f_k)^2 P(r) dr$$

ההכפלה ב- dr פשוט מנרמלת את הביטוי הכולל, ולכן נדרשת ההכפלה בגודל התמך. בעצם, הפונקציה $P(r)$ היא פונקציה ההסתברות לקבלת רמת אפור כלשהי, ולכן עליה לקיים

$$\forall f, P(r) \geq 0, \quad \int_{-1}^1 P(r) dr = 1$$

הביטוי המתקבל אם כך הוא

$$\text{Minimize} \quad \sum_{k=0}^N \int_{r_k}^{r_{k+1}} [r - f_k]^2 P(r) dr$$

$\{r_k\}_{k=1}^{N-1}, \{f_k\}_{k=1}^N$

וזהו הביטוי אותו נרצה להביא למינימום. נשים לב לכך ש- r_0 ו- r_N ידועים כיוון שאלה ערכי הקצה. כדי להשיג את המינימום עלינו לגזור את הביטוי ביחס לנעלמים (רמות ההחלטה ורמות הייצוג) ולהשוות לאפס. גזירה זו מניבה:

$$\frac{\partial}{\partial f_k} \sum_{k=0}^N \int_{r_k}^{r_{k+1}} [r - f_k]^2 P(r) dr = \int_{r_k}^{r_{k+1}} [f_k - r] P(r) dr = 0$$

$$\frac{\partial}{\partial r_k} \sum_{k=0}^N \int_{r_k}^{r_{k+1}} [r - f_k]^2 P(r) dr = -[r_k - f_{k+1}]^2 P(r_k) + [r_k - f_k]^2 P(r_k) = 0$$

בשתי הנגזרות, הסכימה החיצונית נופלת כיוון שהנעלם נתייחס לגורם אחד (במקרה הראשון) או שניים (במקרה השני), וכל היתר מהווים לעניין הגזירה קבוע. הגזירה לפי f_k נכנסת פנימה לאינטגרל ומתבצעת בתוכו. הגזירה לפי r_k מניבה שני איברים כיוון שנעלם זה מהווה פעם גבול תחתון ופעם גבול עליון באינטגרלים. גזירה לפי גבול האינטגרל הנה גובהה של הפונקציה הפנימית לאינטגרל בנקודה זו, וכך מתקבלת התוצאה. אנו עושים שימוש בנוסחאות הבאות:

$$\frac{\partial}{\partial y} \int_{-\infty}^y f(x) dx = f(y)$$

$$\frac{\partial}{\partial y} \int_y^{\infty} f(x) dx = -\frac{\partial}{\partial y} \left[\int_{-\infty}^{\infty} f(x) dx - \int_{-\infty}^y f(x) dx \right] = -f(y)$$

$$\frac{\partial}{\partial y} \int_a^b f(x, y) dx = \int_a^b \frac{\partial}{\partial y} f(x, y) dx$$

ביצוע מספר פעולות אלגבריות מתקבל:

$$k = 1, 2, \dots, N: \quad f_k = \frac{\int_{r_k}^{r_{k+1}} r P(r) dr}{\int_{r_k}^{r_{k+1}} P(r) dr}$$

&

$$k = 1, 2, 3, \dots, N-1 \quad r_k = \frac{f_{k+1} + f_k}{2}$$

משמעות תוצאה זו היא שבהינתן רמות הייצוג, רמות ההחלטה מתקבלות ע"י מיצוע פשוט, כך שבין כל שתי רמות ייצוג עוקבות קיימת באמצע בדיוק רמת החלטה. הביטוי האחר (מצד שמאל) קובע כי

בהינתן רמות ההחלטה, רמות הייצוג מתקבלות ע"י מרכז מאסה של פונקצית הפילוג P באינטרוול $[r_k, r_{k+1}]$.

ברור כי משוואות אלה סתומות, ולא ניתן לחלץ ביטויים סגורים עבור הנעלמים השונים. באופן מעשי, מציאת פתרון נעשית באופן איטרטיבי, בו מתחילים בניחוש מוצלח של רמות ההחלטה (או רמות הייצוג – זה לא משנה), ומעדכנים את רמות הייצוג לפי המשוואות הנ"ל. עם הערכים המעודכנים חוזרים ומשנים את רמות ההחלטה, וחוזר חלילה. אתחול סביר לתהליך הוא בחירת רמות אחידות כפי שתואר קודם, וביצוע איטרציות משם. התהליך האיטרטיבי המוצע מובטח להקטין את השגיאה הכוללת (ערך הפונקציה המובאת למינימום) בכל איטרציה, מובטח להתכנס לערכים יציבים כלשהם, אך בפירוש לא מובטח שהפתרון המתקבל יהיה המינימום הגלובלי. בהחלט יתכן שהפתרון ייתקע על מינימום לוקלי, ולכן חשוב לבצע את התהליך מספר פעמים עם איתחולים שונים ולבחור את הטוב שבין הפתרונות.

בעיית המינימום הלוקלית תלויה באופן חזק באופייה של פונקצית ההסתברות $P(r)$. כאשר פונקציה זו מאוד לא חלקה, ישנן מינימום לוקליות רבות, ובדרך כלל ההתכנסות תהיה לפתרון סמוך לערכי האיתחול עם שיפור עדין. פתרון מעשי לבעיה זו היא החלקת הפונקציה $P(r)$ בטרם עשיית שימוש בה.

שאלה חשובה שנזנחה עד כה היא מקורה של הפונקציה $P(r)$. כזכור, התהליך מתחיל עבורנו בקבלת תמונה $f[m,n]$ שערכיה ממשיים ובתחום $[0,1]$. נניח כי כמות הפיקסלים סופית ונתונה ע"י $|\Omega|$. אזי, הפונקציה $P(r)$ נתונה כסכום של $|\Omega|$ פונקציות הלם חד-מימדיות הממוקמות בערכי רמות האפור של כל פיקסל, דהיינו:

$$P(r) = \frac{1}{|\Omega|} \sum_{[m,n] \in \Omega} \delta(r - f[m,n])$$

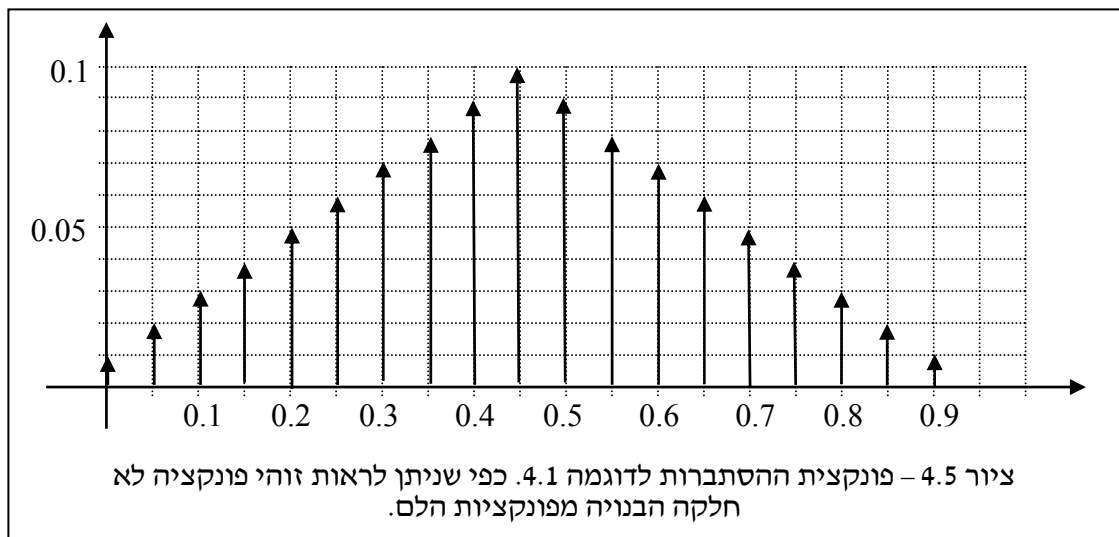
החלוקה ב- $|\Omega|$ נעשית לשם נרמול הפונקציה. וכפי שאנו רואים, במקרה של תמונה דגומה, סביר למדי שפונקציה זו תהיה מאוד לא חלקה. החלקתה אפשרית ע"י ביצוע קונבולוציה עם גאוסיאן ממורכז ברוחב σ , גודל הקובע את עוצמת ההחלקה, ואז:

$$P_G(r) = \frac{1}{|\Omega|} \sum_{[m,n] \in \Omega} G(r - f[m,n], \sigma)$$

דוגמה 4.1: נניח כי נתונה תמונה בת 10 על 10 פיקסלים מעל התמך $[0,9] \times [0,9]$. רמת האפור בתמונה בנקודה $[m,n]$ נתונה ע"י $f[m,n] = 0.05 \cdot (m+n)$

$$f[m,n] = \frac{1}{20} \begin{bmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\ 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 \\ 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 \\ 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 \\ 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 \\ 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 & 17 \\ 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 & 17 & 18 \end{bmatrix}$$

נבנה עבור תמונה זו את פונקציית ההסתברות. רמת אפור 0 תתקבל רק במקום אחד – $[0,0]$, רמת אפור 0.05 תתקבל בשני פיקסלים – $[1,0]$ ו- $[0,1]$, וכך הלאה עד רמת 0.45 שתתקבל בעשרה פיקסלים. הערך 0.5 יתקבל בתשעה פיקסלים וכו', ועד 0.9 שיתקבל בפיקסל אחד. תיאור גרפי של הפונקציה $P(r)$ נתון בצור 4.5.

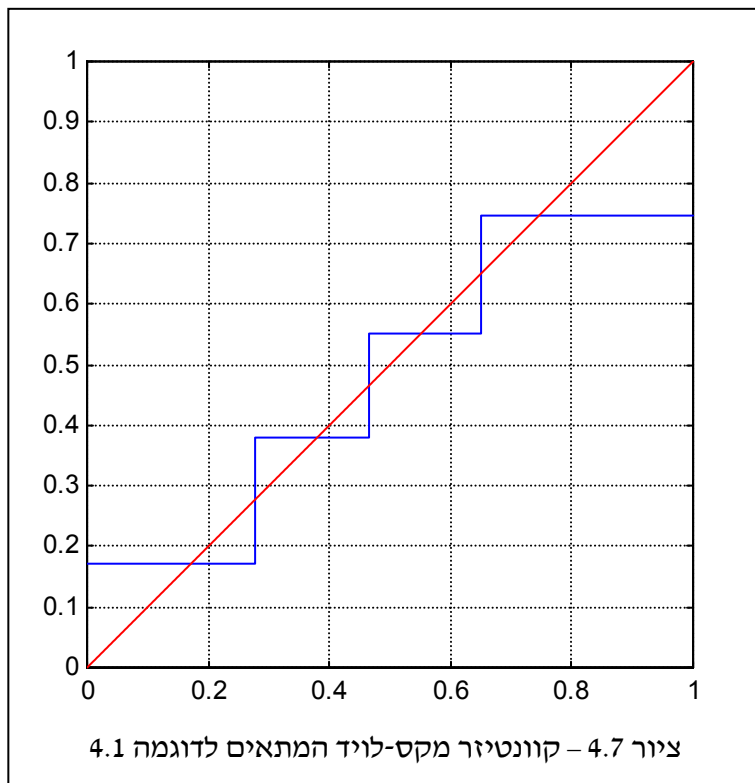
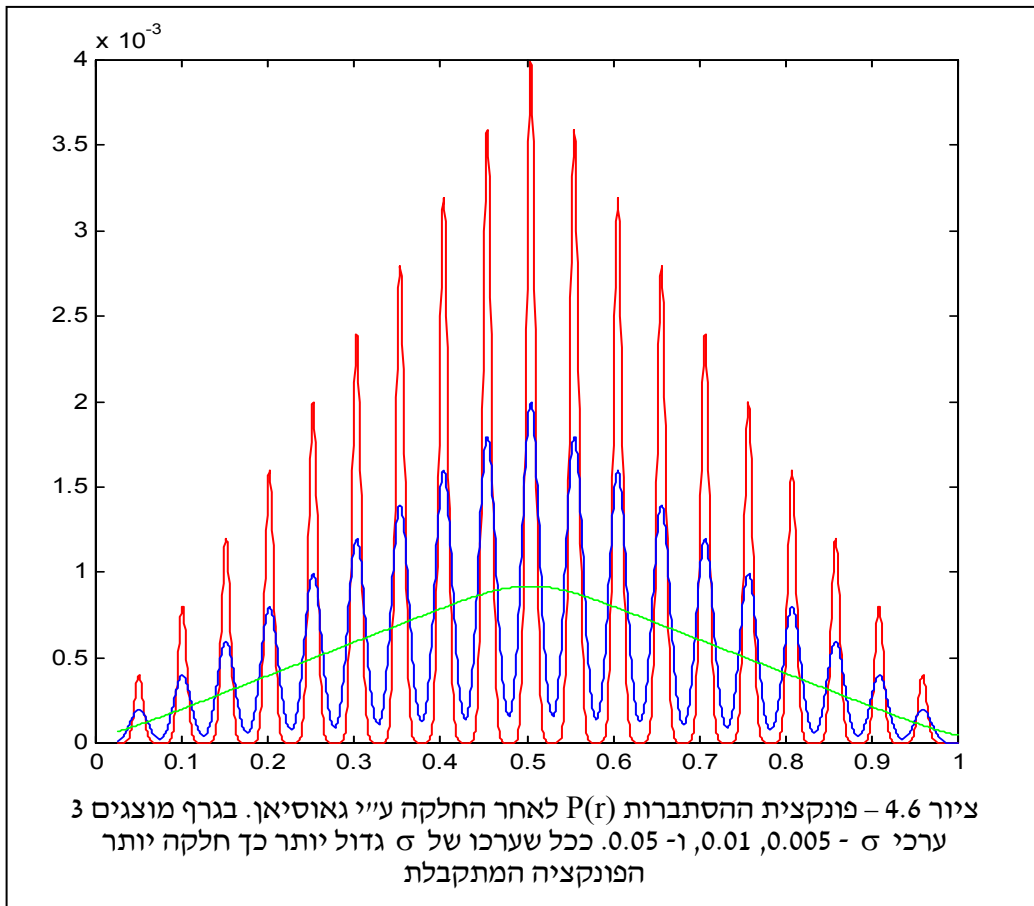


צור 4.6 מציג תוצאת MATLAB לפונקציה זו תוך שימוש בגאוסיאן עם ערכים שונים של השונות σ , ואמנם ניתן לראות שהחלקה מצליחה יפה (תופעה מעניינת היא שעם הגדלת ה- σ גובה הפונקציה הכללי יורד, וזאת כיוון ששך השטח מתחת הגרף צריך להישאר 1).

נניח כעת כי רצוננו בקוונטיזר בין ארבע רמות ייצוג (כלומר 2 סיביות לייצוג כל רמה). תוך שימוש בקירוב המוחלק של פונקציית רמות האפור, מהו הקוונטיזר האופטימלי? תוכנית לחישוב איטרטיבי כמתואר לעיל הניבה את התוצאה המתוארת בצור 4.7. ערכי רמות החלטה והייצוג לאחר 10 איטרציות הם

$$r: [0 \quad 0.2761 \quad 0.4656 \quad 0.6487 \quad 1.0000]$$

$$f: [0.1728 \quad 0.3794 \quad 0.5519 \quad 0.7455]$$



ניתן לראות מספר תופעות מעניינות בתוצאה שהתקבלה. תופעה ראשונה היא שהקוונטיזר חותר להיות סימטרי סביב ערך הכניסה 0.45 – תוצאה צפויה כיוון שההסתברות סימטרית סביב ערך זה.

תופעה שניה היא שהמרווחים בין רמות ההחלטה באמצע התחום קטנים משמעותית מהמרווחים בקצוות. זאת כיוון שבתחום סביב 0.45 יש הרבה פיקסלים בעלי רמות אפור אלה, ולכן כדאי להעניק להם רמת ייצוג מוצלחת יותר אשר תקטין את השגיאה. לעומת זאת, בקצוות ישנה כמות מועטה של פיקסלים ושגיאה בקוונטיזציה עבורם פחות כואבת.

דוגמה 4.2: נטפל כעת בתמונה שלה פונקצית הסתברות אחידה בתחום $[0.2, 0.6]$:

$$P(r) = \begin{cases} 2.5 & 0.2 \leq r \leq 0.6 \\ 0 & \text{Otherwise} \end{cases}$$

מהו הקוונטיזר האופטימלי? מסתבר כי במקרה זה (ובעצם בכל מקרה של פילוג אחיד מעל אינטרוול כלשהו), הקוונטיזר האופטימלי הוא גם הקוונטיזר היוניפורמי שפגשנו בתחילת פרק זה. לכן, במקרה זה, הקוונטיזר האופטימלי יהיה בעל רמות החלטה $[0 \ 0.3 \ 0.4 \ 0.5 \ 1]$ ורמות הייצוג יהיו $[0.25 \ 0.35 \ 0.45 \ 0.55]$. מדוע זה נכון? ובכן, אם נציב לתוך נוסחאות התנאים ההכרחיים את העובדה שהפילוג אחיד נקבל כי באיזור התמך של הפילוג:

$$f_k = \frac{\int_{r_k}^{r_{k+1}} rP(r)dr}{\int_{r_k}^{r_{k+1}} P(r)dr} = \frac{\int_{r_k}^{r_{k+1}} r dr}{\int_{r_k}^{r_{k+1}} dr} = \frac{r_{k+1}^2 - r_k^2}{2(r_{k+1} - r_k)} = \frac{r_{k+1} + r_k}{2}$$

כלומר, גם רמות הייצוג הינן ממוצע של רמות ההחלטה (כזכור, קיים כבר קשר הפוך שקובע כי רמות ההחלטה הן ממוצע של רמות הייצוג).

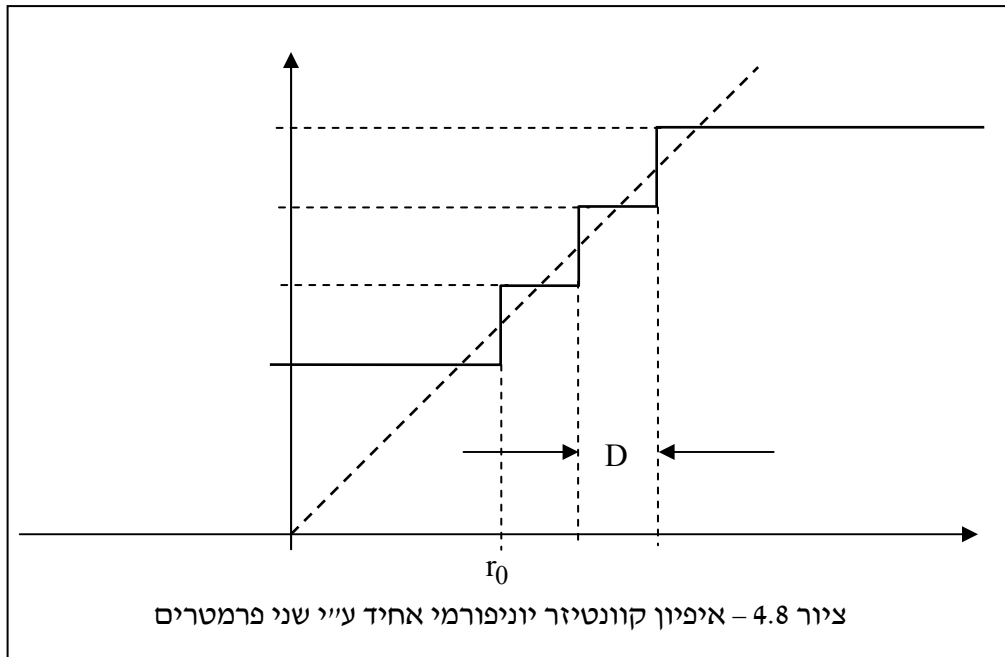
נושא אחרון במסגרת הדיון בקוונטיזרים אופטימליים הוא השימוש בקוונטיזר יוניפורמי אופטימלי. קוונטיזר יוניפורמי קל מאוד למימוש, ולכן מפתה מאוד לעשות בו שימוש, גם כשהוא אינו מתאים לפילוג הנתון. במקום $2N-1$ ערכים חופשיים (רמות ההחלטה ורמות הייצוג), יהיו בידינו כעת רק 2 פרמטרים – מיקום המדרגה הראשונה וגודל המרווחים. השגיאה הריבועית תהיה:

$$\int_{-\infty}^{r_0} \left(r - \left[r_0 - \frac{D}{2} \right] \right)^2 P(r) dr +$$

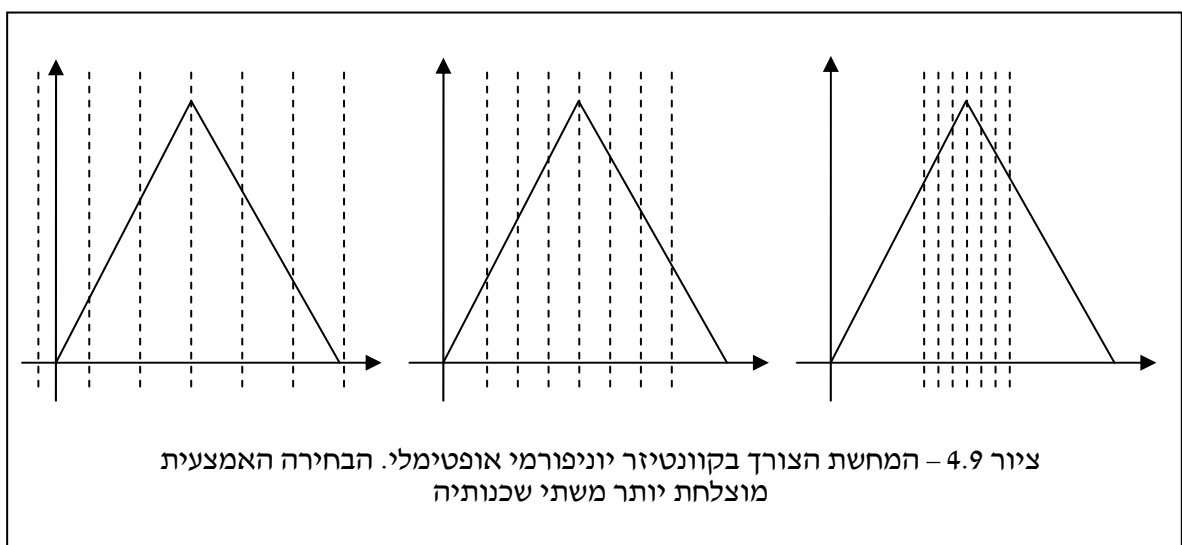
$$+ \sum_{k=1}^{N-2} \int_{r_0+(k-1)D}^{r_0+kD} \left(r - \left[r_0 + kD - \frac{D}{2} \right] \right)^2 P(r) dr +$$

$$+ \int_{r_0+(N-2)D}^{\infty} \left(r - \left[r_0 + (N-2)D + \frac{D}{2} \right] \right)^2 P(r) dr$$

כל יתר הקוונטיזר מוגדר היטב מתוך שני ערכים אלה. בעיית אופטימיזציה מעניינת היא למצוא את הערכים r_0 , D לקבלת שגיאה מינימלית. ציור 4.8 מראה כיצד נראה קוונטיזר כזה לארבע רמות ייצוג.



ע"י גזירת הביטוי הנ"ל ניתן לקבוע באופן דומה תנאים הכרחיים על שני הנעלמים הנ"ל. ציור 4.9 ממחיש את השימוש בקוונטיזר יוניפורמי אופטימלי. הרעיון הבסיסי הוא לשחק עם ה-Scale של פונקציית הקוונטיזר ומיקומה היחסי לשם השגת שגיאה טובה יותר. בציור מתואר פילוג כלשהו ועליו את המיקומים של מדרגות הקוונטיזציה. ברור כי הבחירה בציור המרכזי תניב שגיאה קטנה יותר, בהשוואה לשתי האפשרויות האחרות.



4.3 קוונטיזציה עם שיקולים חזותיים

בתחילת פרק זה ראינו כי בקוונטיזציה עם מספר סיביות נמוך מתקבלת תופעת Contouring, בה רואים עקומים המתחמים איזורים שווי רמת אפור. קוונטיזר מקס-לויד יכול לסייע בפתרונה של הבעיה זו, אך מחייב תכנון של קוונטיזר לכל תמונה מחדש, דבר מטריד למדי. בפועל, מקובל להשתמש בקוונטיזר אחיד ושרירותי שאינו תלוי תמונה, ולפיכך נחוצים קרוב ל-8 סיביות לקבלת איכות טובה. עם זאת, במצבים בהם לא ניתן להקצות 8 סיביות (אם מסיבות חומרה, זכרון או אילוצי תצוגה), הבעיה מתעוררת ומטרידה.

במסגרת סעיף זה נדגים כיצד ניתן לפתור (לפחות חלקית) את בעיית ה-Contouring ללא שימוש בקוונטיזר מקס-לויד, וללא שימוש בפעולות תלויות תמונה, תוך שימוש בפחות סיביות. הרעיון המרכזי הוא לנצל את תכונות מערכת הראיה האנושית, כדי להשיג את השיפור המיוחל. תופעת ה-Contouring בולטת כל-כך כיוון שמערכת הראיה האנושית רגישה לשפות ואפילו עדינות. העקומים המתקבלים הינם שפות לכל דבר, ולכן הם צורמים כל-כך. הפתרון, לפיכך, הוא שבירת שפות אלו, וזה מושג ע"י הוספת רעש לתמונה לפני הקוונטיזציה.

הרעש הנוסף נדרש להיות רעש המפולג אחיד וסימטרי, חסר קורלציה מרחבית (דהיינו – דגימותיו הן תהליך iid – מפולגים באופן זהה וחסר תלות). פילוג אחיד נתון ע"י:

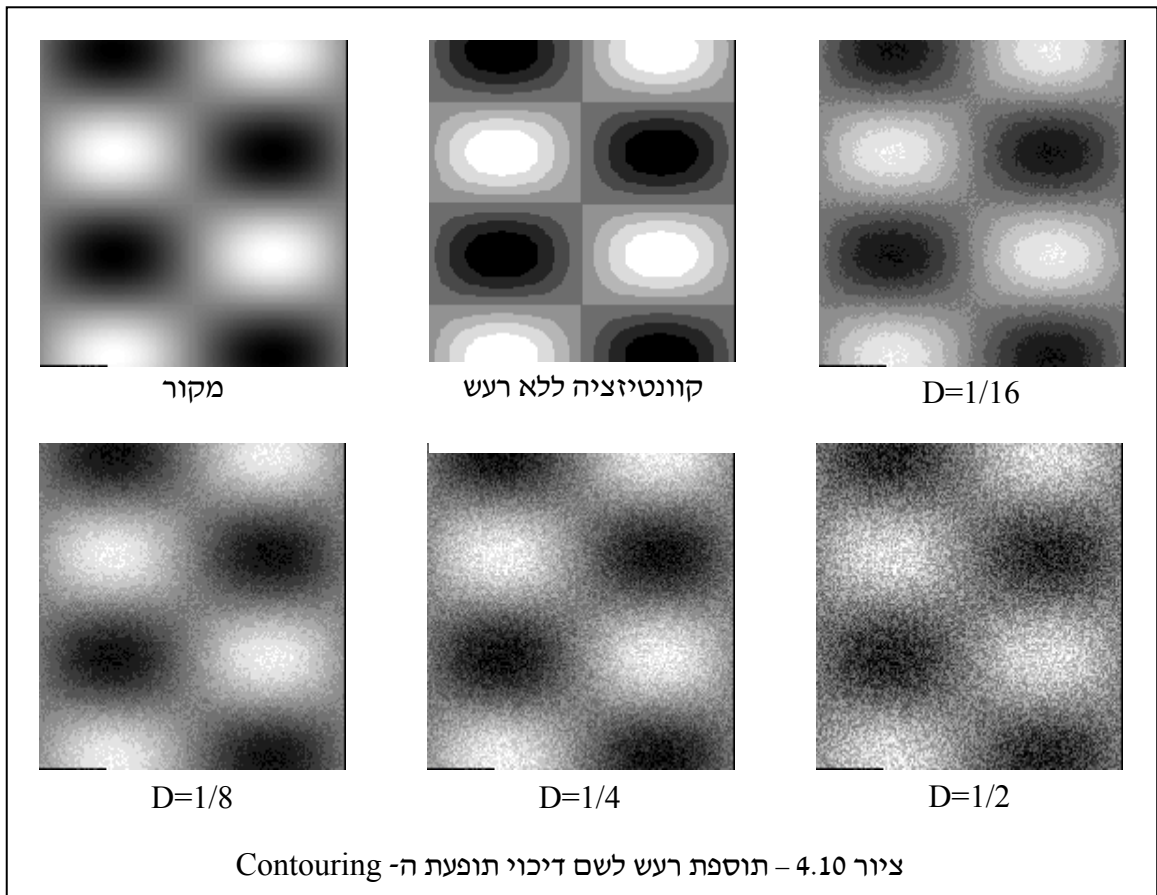
$$P(n) = \begin{cases} \frac{1}{2D} & -D \leq n \leq D \\ 0 & \text{Otherwise} \end{cases}$$

כאשר D היא אמפליטודת הרעש. אם התמונה עוברת קוונטיזציה ל- R סיביות, האמפליטודה המומלצת לשימוש תהיה 2^{-R-1} . אם נסמן תמונת רעש בעלת פילוג אחיד וסימטרי סביב הראשית באמפליטודה 0.5 ע"י $g[m,n]$, פעולת הקוונטיזציה המוצעת היא:

$$Qf[m,n] = \frac{\lfloor (f[m,n] + g[m,n] \cdot 2^{-R}) \cdot 2^R \rfloor}{2^R} + 2^{-(R+1)} = \frac{\lfloor 2^R f[m,n] + g[m,n] \rfloor}{2^R} + 2^{-(R+1)}$$

בציור 4.10 מתוארת תוצאה של תמונת הסינוס הדו-מימדי לאחר קוונטיזציה ל-3 סיביות עם עוצמות שונות של הרעש (רעש המקור בעל אמפליטודה של 0.5), וניכר שעבור עוצמת רעש כמומלץ התוצאה נראית הרבה יותר טוב מללא רעש כלל.

תוצאה מעניינת נוספת היא שאם עושים שימוש ברעש פסבדו-אקראי (כלומר, רעש מיוצר מחשב, הניתן לבניה מחודשת בדיוק מוחלט) לפי המאפיינים שתוארו, ניתן להחסיר את הרעש ממוצא הקוונטיזר

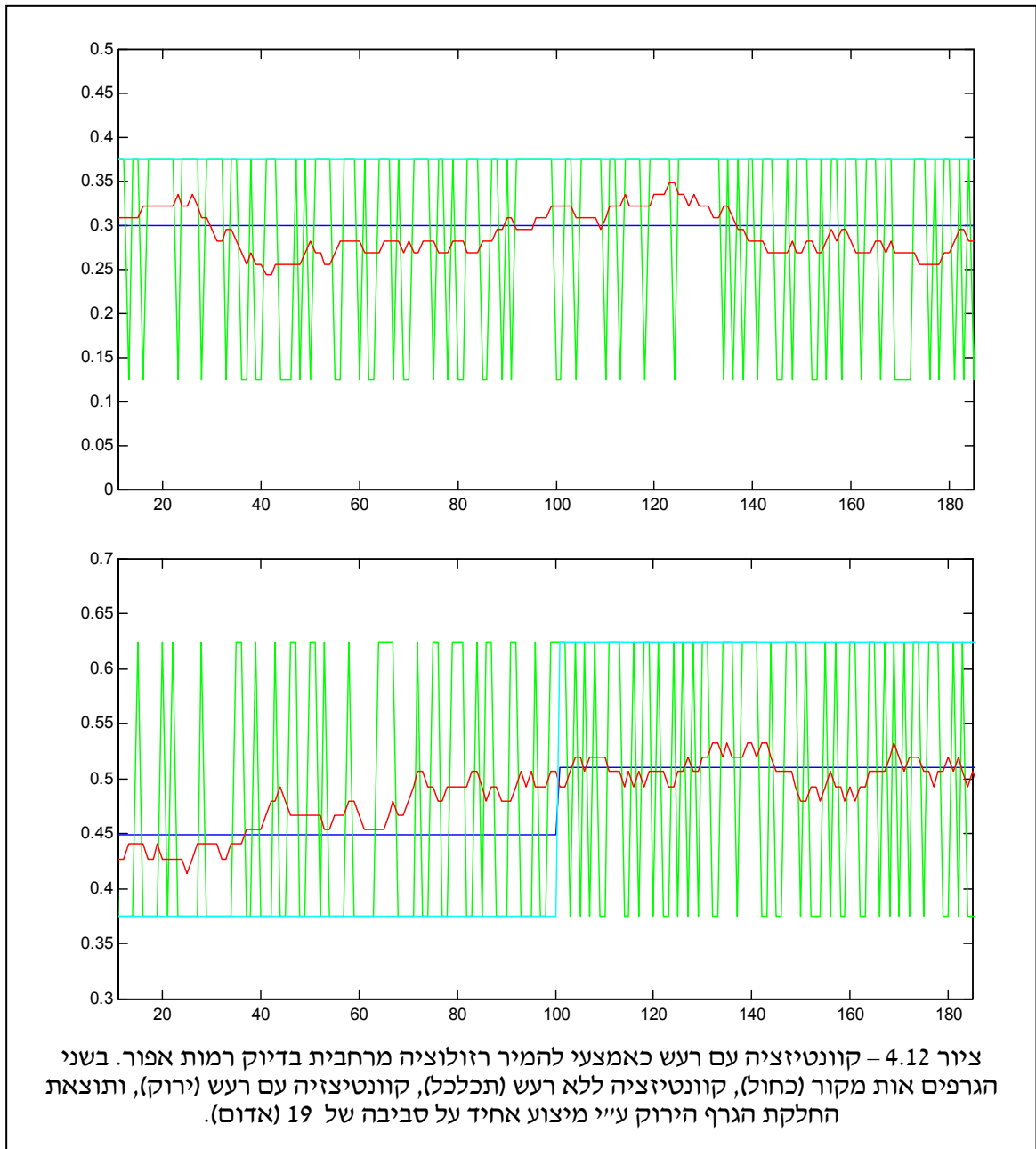


ולקבל אות קרוב למדי למקור, בטרם הקוונטיזציה. תופעה זו ניתנת לניצול באופן הבא (וזוהי רק דוגמה אחת אפשרית): נניח כי נתונה לנו תמונה ואנו צריכים להדפיסה ולהעבירה לגורם אחר. נניח כי המדפסת יכולה להציג רק עם R סיביות (מדפסת אופיינית יכולה עם סיבית אחת). נבצע את הקוונטיזציה עם רעש פסבדו-אקראי ונקבל הדפסה שלה. מקבל התמונה יוכל לסורקה ולקבל קובץ תמונה לאחר קוונטיזציה. הוא יוכל לשחזר במדויק את הרעש ולהחסירו מהתמונה, וכך לקבל תוצאה קרובה לתמונה המקורית. צויר 4.11 ממחיש את התמונות בתהליך זה לשימוש בסיבית אחת.

מדוע הרעש מסייע? כיצד נוצלה כאן מערכת הראיה? כפי שנאמר קודם, המרכיב הצורם ביותר הם קווי השפה בין אזורים שווי רמת אפור, וקווים אלו חלקים למדי, ולכן קופצים לעין. הוספת רעש אקראי עם ערכים חיוביים ושליליים בעוצמה של מרווח אחד שלם בקוונטיזציה המופעל גורמת לקוונטיזציה לקפוץ רמות מעלה מטה באופן אקראי מסביב לשפה. קפיצות אלו שוברות את קו התפר, ולכן מעלימות אותו. מאידך, כיוון שהרעש סימטרי סביב האפס, הגוון הממוצע נותר כבמקור. למעשה, תכונה נוספת של מערכת הראיה באה לידי ביטוי. העין מחליקה את הרעש ע"י דיכוי התדרים הגבוהים. החלקה זו נותנת כי איננו רואים בפועל את הקפיצות המהירות מרמה לרמה, אלא את הגוון הסביבתי הכולל, וגוון זה נאמן יותר למקור.

דוגמה 4.2: נתונה תמונה עם רמת אפור קבועה ושווה 0.3, ולתמונה זו מבוצעת קוונטיזציה ל-4 רמות ללא רעש. קוונטיזציה זה ממפה את כל הערכים בתחום $[0.25, 0.5]$ לערך 0.375. לכן נקבל במוצא הקוונטיזציה תמונה קבועה שרמתה היא 0.375, והשגיאה בין המקור לתוצאה קבועה ושווה 0.075. כאשר נוסיף רעש אקראי בעוצמה של 0.125, ערכי התמונה ינועו באקראי בתחום $[0.175, 0.425]$. הערכים

מיקום המדרגה מתואר בצורה מדוייקת ע"י הקוונטיזר ללא הרעש, אך עם סטיות רמה גדולות. הקוונטיזציה שעושה שימוש ברעש כמעט ולא נותנת ביטוי למדרגה שהיתה.

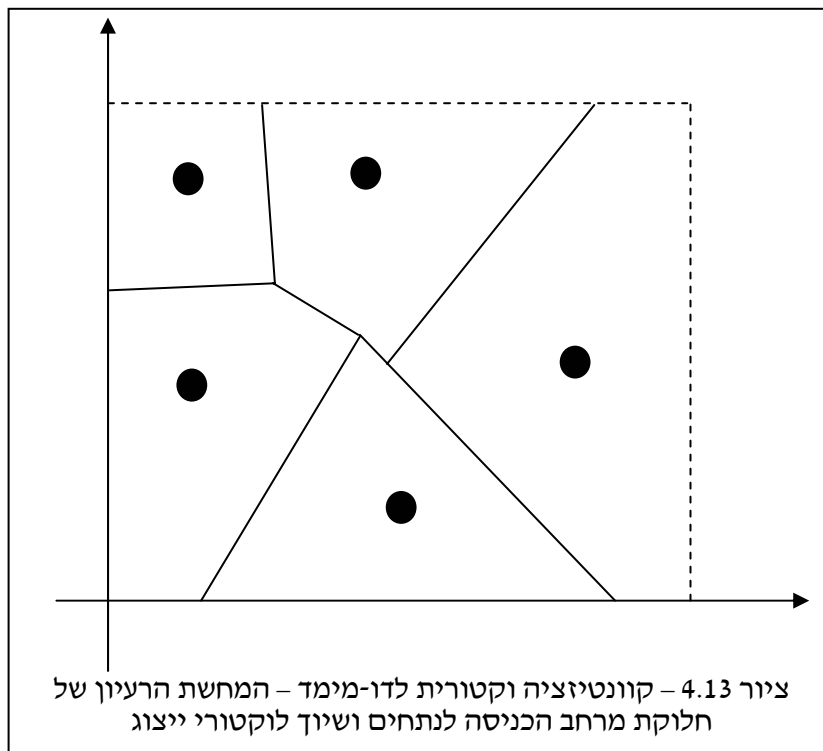


כפי שנראה במסגרת פרקים שיבואו בהמשך, למתואר כאן קשר ישיר לתהליך ה-Half-Tonning הנחוץ במדפסות. גם שם, שימוש בתכונות מערכת הראיה מניב ירידה כמות הסיביות הנחוצות לייצוגה של התמונה.

4.4 קוונטיזציה וקטורית וקוונטיזציית צבע

עד כה עסקנו בתמונות שחור-לבן, בהן הקוונטיזר מוזן בערכים סקלריים - רמות האפור. תמונת צבע, לעומת זאת, מאופיינת בדרך כלל בכל מקום $[m,n]$ ע"י שלושה ערכים של רמות אפור - G, R ו- B , כשכל אחד מהם מספר ממשי בתחום $[0,1]$. כשאין בעיה של כמות סיביות, מקובל לבצע קוונטיזציה יוניפורמית לכל שכבת צבע בנפרד ל-8 סיביות, וכך לקבל כי 24 סיביות מייצגות את מרחב הצבעים. לקוונטיזציית צבע חשיבות מעשית רבה. במגוון גדול של מסכי מחשב תומך כרטיס המסך במספר צבעים מצומצם (בד"כ 256 צבעים שונים - 8 סיביות בלבד). לכן, ביצוע קוונטיזציה סקלרית לכל שכבת צבע בנפרד ל-8 סיביות אינה מתאימה.

קוונטיזציית צבע היא קוונטיזציה וקטורית, בה הקוונטיזר מוזן בוקטור בעל שלושה מרכיבים, ומניב וקטור אלטרנטיבי בעל שלושה מרכיבים. הרעיון הבסיסי בקוונטיזציה וקטורית הוא חלוקת מרחב הכניסה ל- N נתחים שונים, ולבחור לכל נתח כזה "רמת ייצוג" (ובעצם - וקטור ייצוג) אחד. בהינתן כניסה כלשהי, המרתה לוקטור הייצוג תיקבע לפי הקבוצה אליה היא שייכת. ציור 4.13 ממחיש רעיון זה לוקטורים דו-מימדיים. בציור, כל וקטור דו-מימדי מומר לאחד מתוך 5 וקטורי ייצוג אפשריים, ואלו הם ה-"צבעים" האפשריים (מפת השיוך המתקבלת נקראית דיאגרמת Voronoi).



לקוונטיזציה וקטורית ככלל יש יישומים רבים גם לצורך דחיסה וגם לצורך הדפסה, בדומה לנאמר על קוונטיזציה סקלרית - במסגרת פרק זה נתמקד בנושא הקוונטיזציה כתהליך שבא אחרי דגימה.

הפתרון הפשוט ביותר לקוונטיזציה וקטורית הוא הפרדת הטיפול בין שלושת שכבות הצבע השונות, וביצוע קוונטיזציה יוניפורמית לכל אחת מהן בנפרד. כך למשל, אם נדרשים 256 צבעי ייצוג, נוכל לקחת את התמונה ולבצע קוונטיזציה לשתי שכבות ממנה ל- 6 תחומים, ולשכבת הצבע השלישית ל- 7 תחומים. קוביית היחידה התלת-מימדית $[0,1] \times [0,1] \times [0,1]$ תחולק בצורה כזו ל- $6 \cdot 6 \cdot 7 = 252$ מלבנים תלת-מימדיים שווי נפח הממלאים את הקוביה. ארבעת הצבעים הנותרים לא חייבים להיות מבוזבזים – בד"כ במסכי מחשב, הצבעים הנותרים משמשים כצבעים שמורים לרקע, חלונות וכו'.

הגישה המתוארת מקבילה להצעתנו לשימוש בקוונטיזר יוניפורמי לתמונות שחור-לבן. חסרון מרכזי של גישה זו הוא הטיפול גרוע בתמונות בהן ישנו ריכוז של צבעים מסוימים. בדומה לדיון על קוונטיזציה סקלרית, עבור תמונה שרמות האפור בה (בשלושת השכבות) מרוכזות בתת-אינטרוול קטן, נקבל ייצוג גרוע של התמונה. היתרון הבולט של שיטה זו הוא פשטותה, והיותה בלתי תלויה בתמונה עליה היא מופעלת.

על מנת לשפר את השיטה, מוצע לאמץ קוונטיזר שאינו יוניפורמי, ולעשות זאת תוך התייחסות לתוכן התמונה. גישה זו מכלילה את רעיון קוונטיזר מקס-לויד לטיפול בוקטורים. אם ניזכר בדרך פעולתו של קוונטיזר מקס-לויד, הרי שתהליך בנייתו כרוך בשני צעדים המבוצעים איטרטיבית. נזכיר מהם צעדים אלו ונראה כיצד להכלילם למימד גבוה מ- 1:

1. בהינתן N רמות הייצוג, יש לקבוע את רמות ההחלטה. במקרה הסקלרי עשינו זאת ע"י קביעת רמת ההחלטה כאמצע הדרך בין רמות הייצוג הסמוכות, לפי הקשר

$$r_k = \frac{f_{k+1} + f_k}{2}$$

כתחליף לגישה זו, אם נתייחס לרמות ההחלטה כגבולות בין האינטרוולים השונים, הרי שציר הממשיים מחולק לאזורים לפי שיטת השכן הקרוב, בה לכל נקודה על הציר נבדק מי רמת הייצוג הקרובה ביותר ואליה היא משויכת.

בהכללה לדו-מימד נאמר כי נתונים לנו N וקטורי ייצוג המסומנים ב- $f_k \in \mathbb{R}^S$. אנו לוקחים את מרחב הכניסה ה- S מימדי במקרה הכללי ($S=3$ לתמונות צבע), וכל נקודה בו משויכת לוקטור הייצוג הקרוב ביותר.

אם במקרה הסקלרי בדקנו מרחק ע"י ריבוע ההפרש של הערכים, הרי שכאן יש לנו חופש מוחלט בבחירת המטריקה. ניתן להשתמש במטריקה אויקלידית (סכום ריבועי הפרשי הקואורדינטות של הוקטורים). לעיתים ניתן להציע מטריקה ייחודית המחשבת את המרחק בין הצבעים לפי תפיסתה של מערכת הראייה, וזה כמובן מוצלח יותר.

2. בהינתן $N+1$ רמות ההחלטה, יש לקבוע את רמות הייצוג העדכניות. במקרה הסקלרי זה נעשה ע"י החישוב

$$f_k = \frac{\int_{r_k}^{r_{k+1}} rP(r)dr}{\int_{r_k}^{r_{k+1}} P(r)dr}$$

משמעות הביטוי – חישוב מרכז המסה של כל קבוצה המשויכת לאחת מ- N רמות הייצוג. מרכז המסה כאן בנוי באינטגרל על פני כל ערכי הרצף ושימוש בפונקציה ההסתברות של רמות האפור. כאשר עוסקים בתמונה דגומה $f[m,n]$, נניח כי קבוצת הפיקסלים Ω_k משויכת לרמת הייצוג f_k , ונניח כי כמות הפיקסלים בקבוצה זו סופית (הנחה סבירה למדי, כיוון שסך כל הפיקסלים בתמונה היא כמות סופית). אזי, תוך שימוש בעובדה שפונקציה ההסתברות היא סדרת הלמים בערכי רמות האפור הנתונות, ניתן להראות כי מרכז המסה נתון ע"י:

$$f_k = \frac{\int_{r \in \Omega_k} \sum_{[m,n] \in \Omega} \delta(r - f[m,n]) r dr}{\int_{r \in \Omega_k} \sum_{[m,n] \in \Omega} \delta(r - f[m,n]) dr} = \frac{1}{|\Omega_k|} \sum_{[m,n] \in \Omega_k} f[m,n]$$

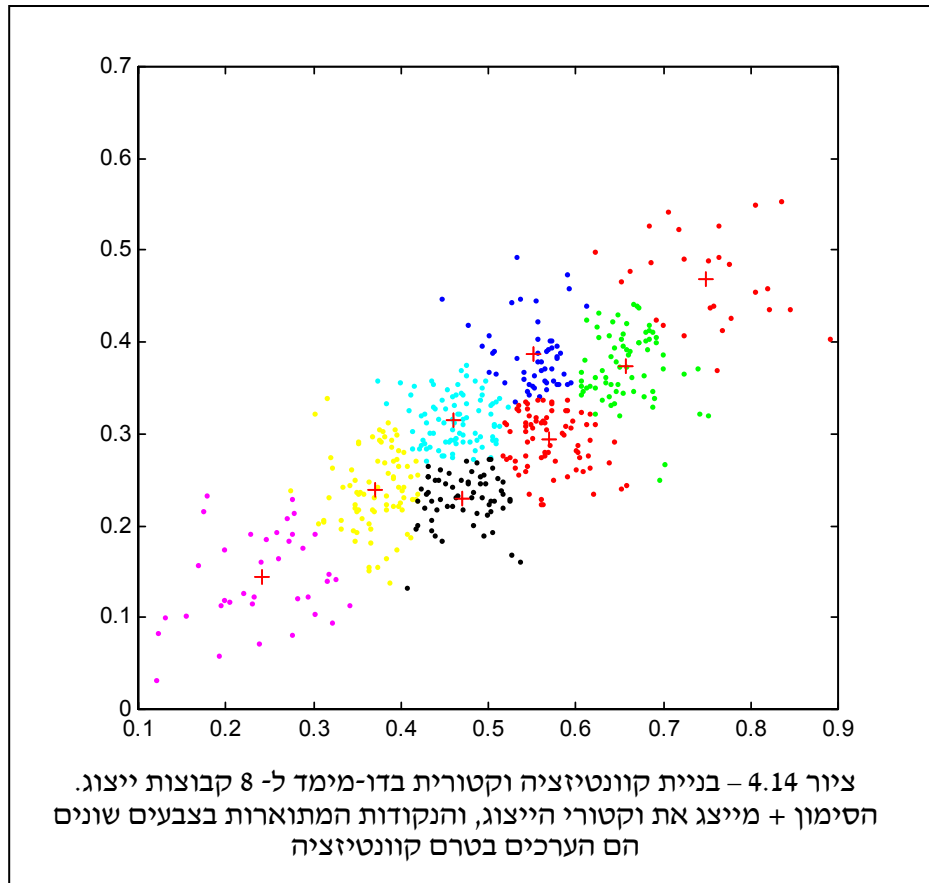
וכך יצא כי חישוב זה אינו אלה חישוב ממוצע הוקטורים בקבוצה. בהכללה לוקטורים, נרצה לחשב את מרכז המסה ע"י חישוב ממוצע דומה של כל הצבעים השייכים לקבוצה.

למעשה, האלגוריתם שתואר מביא למינימום את השגיאה הבאה:

$$\varepsilon^2 \left(\{f_k\}_{k=1}^N \right) = \sum_{[m,n] \in \Omega} \text{Min}_{1 \leq k \leq N} d\{f_k, f[m,n]\}$$

כאשר הפונקציה d היא המטריקה המיושמת, וכפי שרואים, יש דמיון רב לנעשה במקרה הסקלרי.

דוגמה 4.3: נמחיש את האלגוריתם שתואר עבור מקרה דו-מימדי. נניח כי נתונות לנו נקודות בדו-מימד בפיזור כמתואר בציור 4.14. רצוננו למצוא קוונטיזר ל- 8 וקטורי ייצוג דו-מימדיים. נפעיל את האלגוריתם המתואר. אתחול האלגוריתם נעשה ע"י בחירה אקראית של 8 נקודות במישור בקרבת מרכז המסה (הממוצע) של הנקודות הנתונות. ביצוע של 10 איטרציות של האלגוריתם שתואר הניב את התוצאה המתוארת בציור 4.14. לשם נוחות נצבעה כל קבוצה בצבע שונה.



פרק 5 - עיבודים בסיסיים לתמונות

75	פרק 5 – עיבודים בסיסיים לתמונות
76	5.1 פעולות נקודה והיסטוגרמת תמונה
83	5.2 פעולות עיצוב היסטוגרמה
96	5.3 פעולות מרחביות בסיסיות
107	5.4 פעולות מרחביות מסתגלות

בפרק זה נתחיל את הדיון בפעולות עיבוד על תמונות. העיבודים בהם נעסוק בפרק זה יתמקדו בניסיון לשפר את מראיה של התמונה הנתונה. חשוב להבהיר כבר כעת כי בעיבוד תמונות קיימים עיבודים בעלי מטרה שונה, כגון דחיסה, ניתוח תוכן התמונה ועוד, אך אלה יידחקו לפרקים מאוחרים יותר.

מדוע לשפר תמונה? ובכן, במקרים רבים מתקבלת תמונה שאינה באיכות הרצויה, אם בשל דרך השגתה, או בשל קלקולים שעברה בתהליכי עיבוד קודמים. בעיות אופייניות הן תמונה מרוחה הדורשת חידוד, תמונה בה נדרש למתוח את הקונטרסט, תמונה בה העצמים החשובים אינם נראים טוב ונדרש להבליטם, תמונה בה יש רעש נלווה ונחוץ לנקותו, ועוד. כנגד כל משימה כזו ניתן להציע אלגוריתמים שונים ומגוונים. לכן, תחת הכותרת "שיפור תמונה" נכנס מגוון רחב מאוד של אלגוריתמים אשר אמנם משפרים את התמונה אך במובנים שונים, ולעיתים קרובות אין כל קשר ביניהם.

מבחינתנו היה נפלא לו יכולנו להציע אלגוריתם כללי המקבל תמונה ומוציא גרסה משופרת שלה, ללא מעורבותו של המשתמש. המכשלה העיקרית לקיומו של תהליך כזה היא חוסר בתיאור מתמטי של מהי תמונה טובה. כיוון שאין מדד כמותי כזה, שיפור תמונות הוא עניין של מומחיות וטעם אישי - צריך לזהות את הבעיה בתמונה ואז לתפור אלגוריתם ספציפי של פתרונה. לכן, התחום הקרוי שיפור תמונות הוא תחום עם הרבה היורסטיקות, ובו פעולות תתבססנה על ניסוי וטעייה. במסגרת פרק זה ננסה להציג את היסודות וכלי העבודה הבסיסיים לשיפור תמונות.

כנקודה אחרונה במסגרת זו, חשוב להבהיר את ההבדל בין שיפור תמונה ושחזור תמונה. קיים תחום נפרד הקרוי "שחזור תמונה", ובו לוקחים תמונה שעברה קלקול ידוע ומשחזרים את המקור. אנו נדון באלגוריתמים כאלו במסגרת הפרק הבא. במסגרת פרק זה איננו מניחים כי הקלקול ידוע, ולכן דרך הפתרון משתנה באופן מהותי.

5.1 פעולות נקודה והיסטוגרמת תמונה

אחת הפעולות הפשוטות ביותר שניתן לבצע על תמונה היא פעולה חסרת זיכרון. כפי שכבר ראינו בפרק 2, פעולה זו בצורתה הכללית ביותר תהיה $g[m,n] = T\{f[m,n]\}$, כאשר המוצא בנקודה הדיסקרטית $[m,n]$ תלוי רק בערך הכניסה באותה נקודה. מערכת כזו ניתן לתאר ע"י פונקציה מעבר פשוטה הקובעת כיצד כל ערך כניסה עובר לערך יציאה, כמתואר בציור 5.1. אם נניח למשל כי תמונת הכניסה והמוצא מיוצגות ע"י 8 סיביות לפיקסל, הרי שלפנינו מערכת המעבירה כל מספר בתחום $[0,255]$ לערך מוצא בתחום זה, ולכן המערכת ניתנת לתיאור ע"י טבלת התאמה פשוטה הקרויה Look Up Table (LUT).



השתכנענו כי אמנם מדובר במערכת פשוטה, והשאלה המתבקשת היא – מה ניתן להשיג באמצעות מערכת כזו? מסתבר כי לעיתים תכופות יכולה מערכת LUT לספק שיפור ממשי לתמונת הכניסה. לפני שנמחיש את הפעולות הללו נציג את מושג ההיסטוגרמה לתמונה, אשר יסייע בהבנתן של פעולות אלו.

היסטוגרמה לתמונה היא פונקציה השכיחות של רמות האפור בתמונה, בדומה לפונקציות שתוארו במסגרת פרק קודם בדיון על קוונטיזר מקס-לויז. בהינתן תמונה בה כל פיקסל מיוצג ע"י n סיביות, כל פיקסל יכול לקבל ערך רמת אפור שלם בתחום $[0, 2^n - 1]$. הערך $h(x)$ הוא כמות הפיקסלים

בתמונה שערכם הוא x . לכן, הפונקציה $h(x)$ היא פונקציה חד-מימדית המקבלת ערך חיובי כנגד כל x בתחום $[0, 2^n - 1]$. בדרך כלל מקובל לחלק את ערכי הפונקציה בכמות הפיקסלים הכוללת בתמונה, לקבלת פונקציה ששכום איבריה הוא 1. במקרה כזה, הפונקציה $h(x)$ נראית כפונקציה הסתברות לכל דבר המקיימת:

$$1. \forall x \in \mathbf{Z}, 0 \leq x < 2^n, h(x) \geq 0$$

$$2. \sum_{x=0}^{2^n-1} h(x) = 1$$

היסטוגרמה של תמונה מתארת תכונה מסוימת של התמונה – מידת אכלוס רמות האפור בתמונה. להיסטוגרמה אין מרכיב המתאר התנהגות מקומית, כלומר, מהיסטוגרמה לא ניתן לומר מהי רמת האפור במקום מסוים בתמונה, וברור כי מהיסטוגרמה לא ניתן לבנות מחדש את התמונה. בציר 5.2 מוצגות מספר תמונות וההיסטוגרמה שלהן. בכל המקרים מדובר בתמונות המיוצגות ב-8 סיביות לפיקסל. כפי שמודגם בציר זה (בשורה האמצעית), שינוי סדר הפיקסלים בתמונה אינו משפיע כלל על ההיסטוגרמה. ע"י שינוי סדר הפיקסלים בתמונה מתקבלת תמונה שונה לחלוטין. למרות זאת, ברור כי ההיסטוגרמה נותרת ללא שינוי כלל.

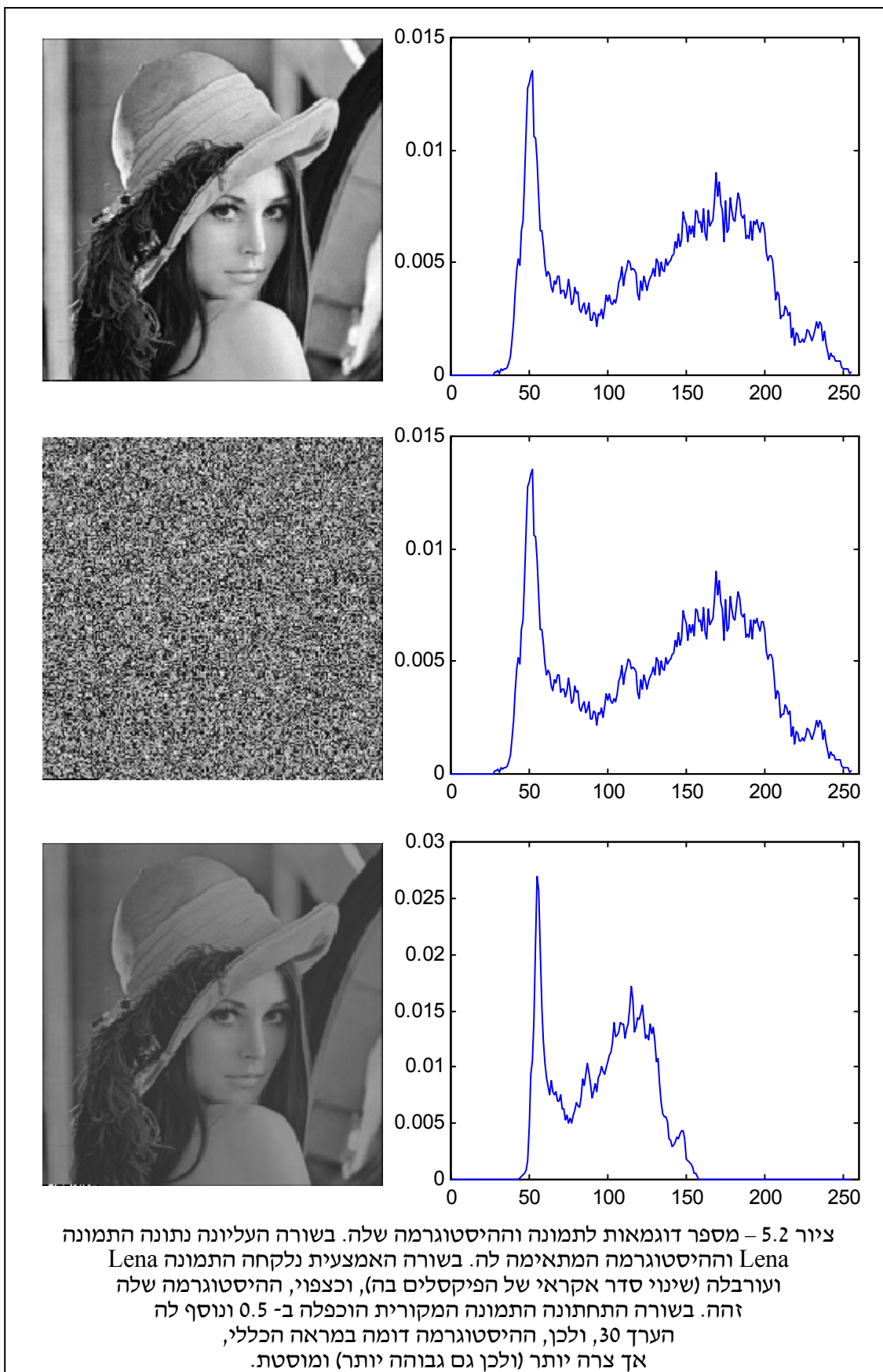
מדוע אנו מתעניינים בהיסטוגרמת תמונה? מסתבר כי עיון בהיסטוגרמת התמונה חושף מיד בעיות של ניצול תחום דינאמי. כך למשל, התמונה החשוכה מעט שמתוארת בתחתית ציור 5.2 נראית כך כיוון שערכי רמות האפור בה מאכלסים בקירוב את התחום $[50, 150]$ - זוהי תמונה שמבזבזת תחום דינאמי. ע"י פעולת LUT על תמונה זו מהצורה

$$\text{Out} = 2.55 \cdot [\text{In} - 50]$$

מתקבלת תמונה אשר מאכלסת את כל רמות האפור מאפס ועד 255. תוצאה זו ופעולת ה-LUT התואמת לה מתוארים בציר 5.3. פעולה זו תיקרא מתיחת קונטרסט (Contast Streching). נשים לב כי פעולה זו ממפה לערכים שליליים את רמות האפור הקטנות מ-50, ולערכים גבוהים מ-255 את ערכי הכניסה מעל 150. אבל בתמונה הנתונה אף פיקסל כמעט אינו בתחומי ערכים אלו, ולכן אין בעיה. על מנת להבטיח תוצאה נאותה, אנו מרתקים כל ערך שלילי לאפס, וכל ערך מעל 255 ל-255.

במקרה הכללי, בהינתן תמונה בה רמת האפור המינימלית תסומן f_{\min} ורמת האפור המקסימלית f_{\max} , פעולת המתיחה לשם ניצול כל הרמות בתחום $[0, 255]$ תהיה:

$$\text{Out} = 255 \cdot \frac{\text{In} - f_{\min}}{f_{\max} - f_{\min}}$$

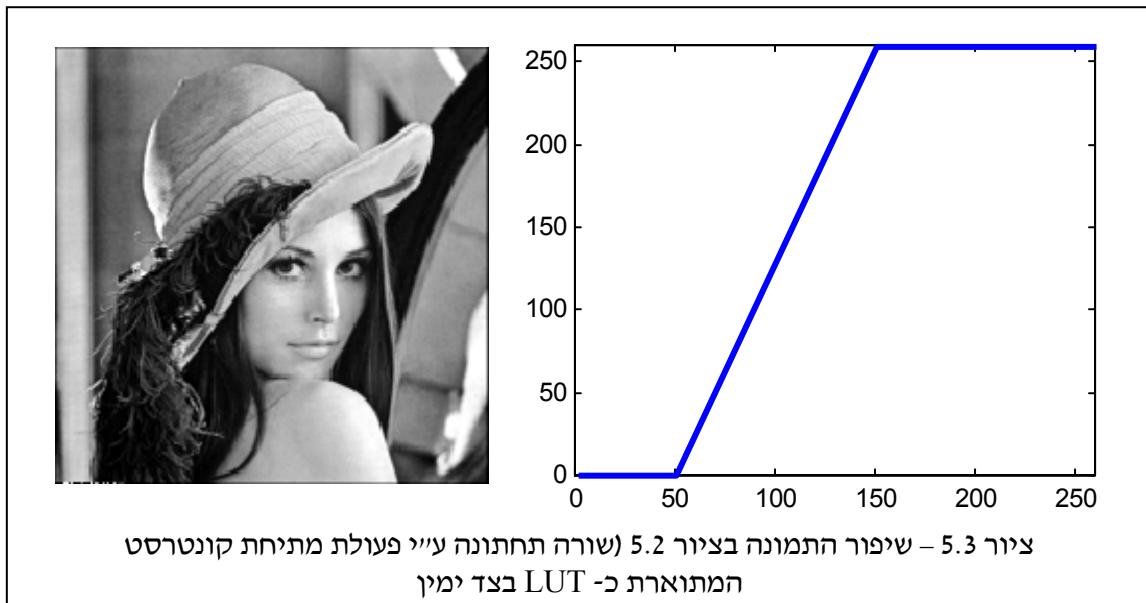


לעיתים שימוש בערכים הקיצוניים (f_{\max} ו- f_{\min}) אינה מוצלחת. לדוגמה, אם בתמונה כל רמות האפור למעט אחת מצויות בתחום $[0,100]$, וישנו פיקסל בודד בעל ערך 255, הרי שלפי המתואר לעיל, מתיחת קונטרסט לא תשנה את התמונה, בשל אותו פיקסל חריג. כדי להתגבר על בעיה זו, ניתן להגדיר את f_{\min} ו- f_{\max} באופן שונה מעט. נזכור כי פונקציית ההיסטוגרמה היא פונקציית צפיפות פילוג, ולכן נוכל להגדיר פונקציית הצטברות הפילוג $H(x)$ ע"י:

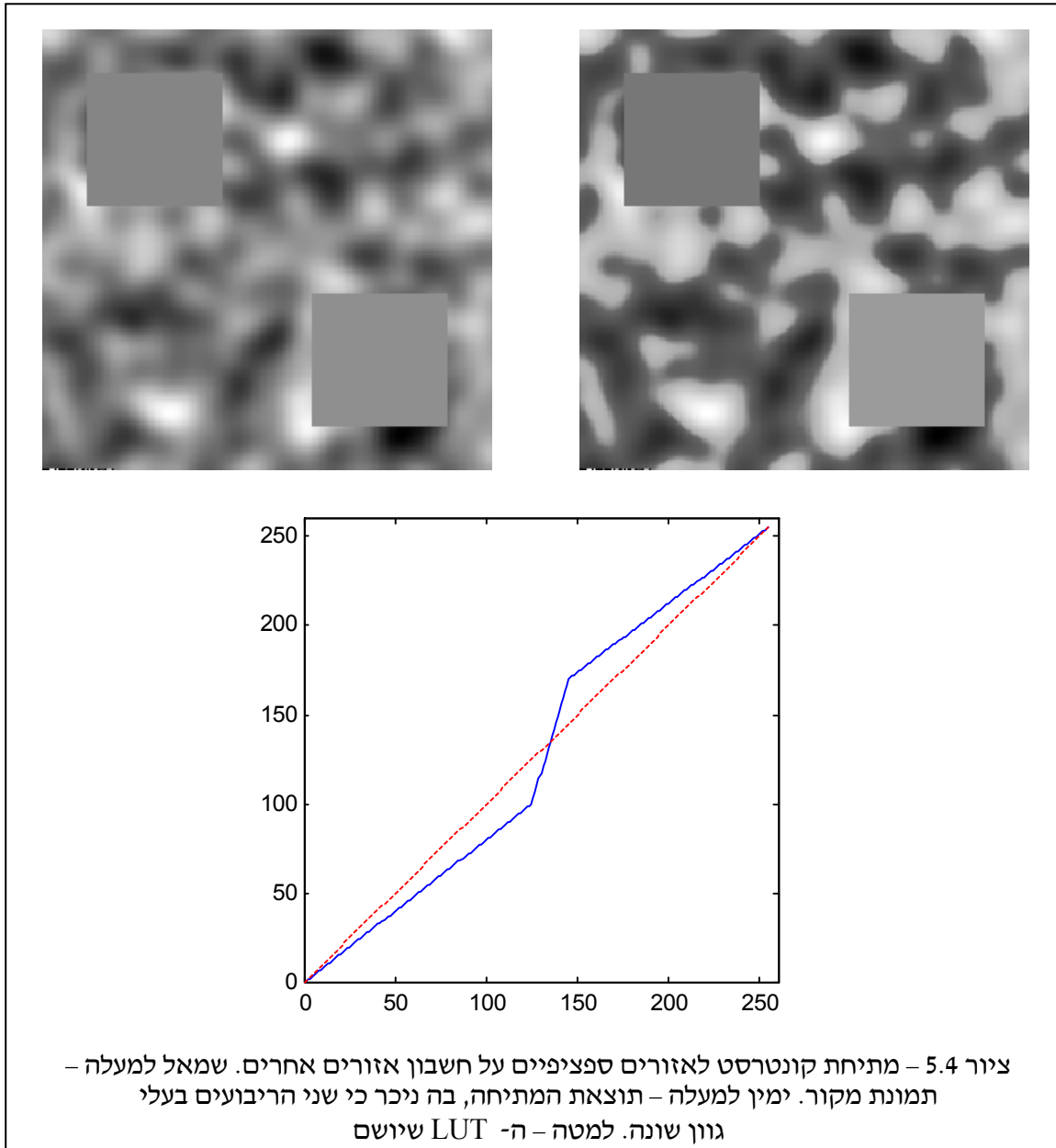
$$H(x) = \sum_{k=0}^x h(x) \Rightarrow H(0) = 0; H(2^n - 1) = 1$$

(n – כמות הסיביות לפיקסל). אזי, אם קודם הגדרנו את f_{\min} להיות הנקודה בה $H(x) = 0$, נגדיר אותה הפעם להיות הנקודה בה $H(x) = \varepsilon$. משמעות בחירה זו היא של- ε מהפיקסלים הכי נמוכים נתייחס כחריגים ונבחר את הגדול שביניהם בערך מינימלי. באופן דומה, בחירת f_{\max} כנקודה בה $H(x) = 1 - \varepsilon$ תוחלף בנקודה בה $H(x) = 1 - \varepsilon$. מתיחה על-פי שני ערכים אלו תגרום לכמות יחסית של 2ε מהפיקסלים בתמונה להיכנס לרוויה (מעל 255) ולקיטעון (מתחת 0), אך זו קבוצה מצומצמת של נקודות, ולעומתה, עיקר התמונה יעבור שיפור מוצלח.

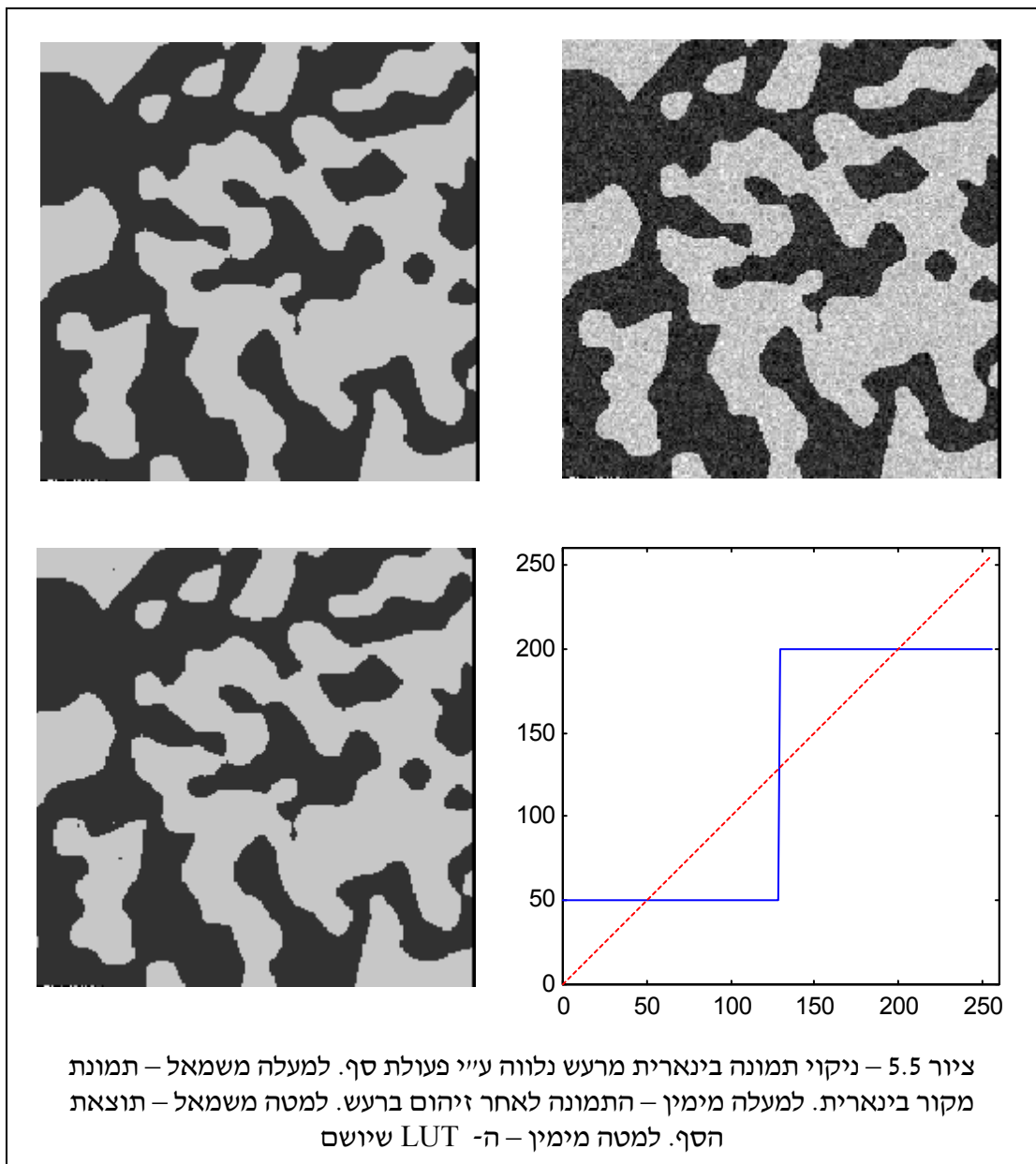
כדוגמה אחרת לפעולת מתיחת קונטרסט, בציור 5.4 מתוארת תמונה בה ישנם שני אובייקטים בעלי רמות אפור קרובות. נניח כי הם העיקר בתמונה ורצוננו להגדיל את יכולת האבחנה ביניהם. בציור 5.4 מוצע גם LUT להגשמת יעד זה. רמת האפור של המטרות הן 130 ו-140. פעולת ה-LUT מרחיקה שתי רמות אלה לערכים 100 ו-170 בהתאמה, תוך פגיעה ביתר התמונה (אשר כאמור, נחשב לפחות חשוב במשימה שלפנינו). גם זוהי פעולת מתיחת קונטרסט, אך רק לתחום מסוים של רמות האפור. בתוצאה ניכר כי אמנם שני הריבועים הינם בעלי גוון שונה, כנדרש.



פעולה אחרת הניתנת לתיאור כ- LUT היא פעולת סף. נניח כי נתונה תמונה אשר יודע כי במקורה היא תמונה בינארית (נניח ערכי רמת אפור 50 או 200 בלבד!). לתמונה זו צורף רעש בתהליך הרכשתה ורצוננו לנקותה. ע"י פעולת סף נוכל להעלים חלק גדול מהרעש. ציור 5.5 ממחיש תהליך זה ותוצאותיו כפי שניתן לראות אנו מצליחים לקבל תמונה כמעט זהה למקור לפני הרעש.



הראינו כיצד אנו משנים קונטרסט – כעת נעבור להראות שינוי בהירות. בהירות היא בעצם רמת האפור הממוצעת בתמונה. לכן, בחירה אינטואיטיבית לפעולת שינוי בהירות תהיה הוספת ערך קבוע לרמת האפור. בחירה כזו אינה מוצלחת בדרך כלל כיוון שהיא גורמת להוצאת הערכים מהתחום המותר. כתחליף, מקובל להשתמש בתיקון גאמה (Gamma Correction), אשר נראית כך:

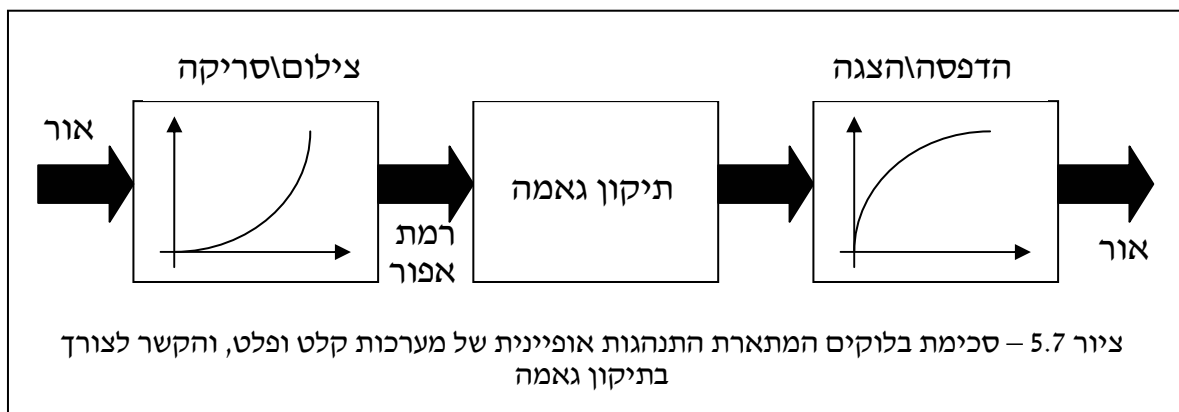
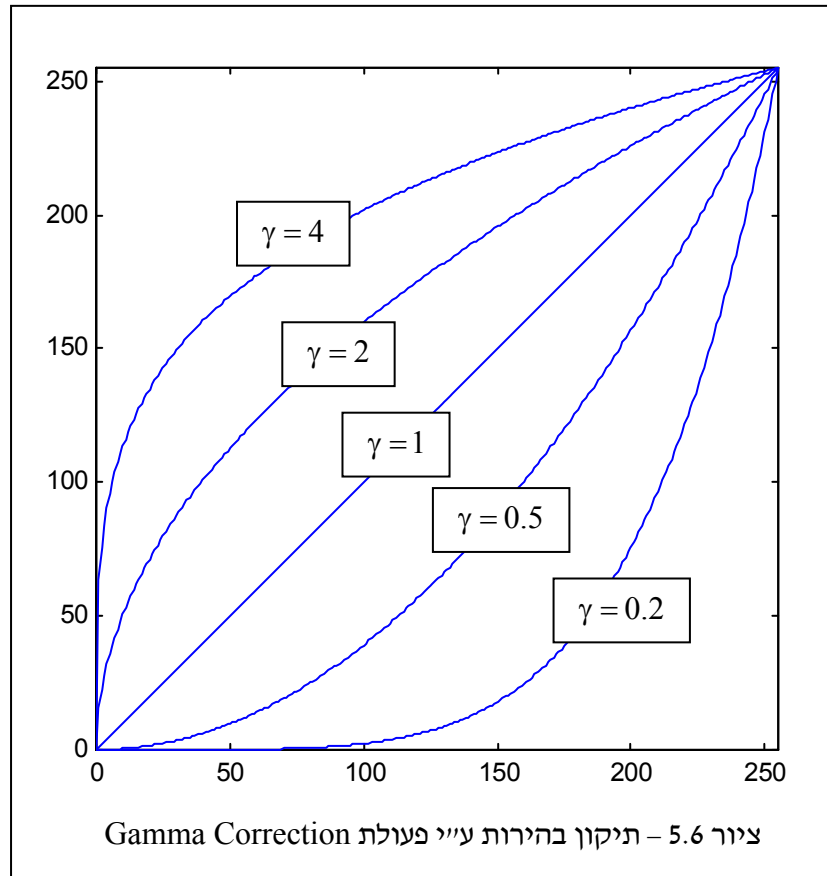


$$\text{Out} = \text{In}^{1/\gamma}$$

עבור ערכי גאמה גדולים מ-1 נקבל הבהרה של התמונה, ועבור ערכי גאמה קטנים מ-1 התמונה תוכהה. ציור 5.6 מראה את צורת ה-LUT למגוון ערכי גאמה.

לפעולת תיקון הגאמה שורשים מוצקים בתהליכי הצילום וההדפסה. מסתבר כי בצילום תמונה (לסרט צילום או לגלאי CCD) רגישות החישה לאור אינה ליניארית וניתנת לתיאור כללי כפעולת חזקה על הכניסה, עם מעריך חיובי וגדול מ-1. לעומת זאת, תהליך ההדפסה (או ההצגה למסך) מאופיינים אף הם בעקומה לא ליניארית של הצגת התמונה, כשעבורם מקובל להניח כי הפעולה היא פעולת חזקה עם מעריך חיובי קטן מ-1. ברור כי עבור ערכים תואמים (יחס של היפוך בין השניים) יתקבל ביטול הדדי ואין

צורך בתיקון נוסף. לעומת זאת, כאשר התקני הקלט ופלט אינם מתואמים, נדרשת פעולת תיקון ביניהם, וזוהי בדיוק פעולת הגאמה. ציור 5.7 מראה סכימת בלוקים של המערכת אשר תוארה.



5.2 פעולות עיצוב היסטוגרמה

עד כה עסקנו בפעולות LUT פשוטות למדי, שנובעות מתוך שיקולים בסיסיים הניכרים מההיסטוגרמה. במסגרת סעיף זה נציג דרכים מורכבות יותר לתכנון פעולת ה-LUT. היעד הראשון אותו נרצה להגשים הוא הבא – בהינתן תמונה כלשהי, נשאל מהי פעולת ה-LUT אשר תמיר את ההיסטוגרמה שלה להיות היסטוגרמה אחידה. מדוע אחידה? ובכן, אינטואיטיבית נראה כי תמונה שמנצלת כל רמת אפור באופן אחיד תהיה תמונה מאוזנת ונעימה לצפייה.

משימה שניה בה נעסוק דומה בעיקרה - בהינתן תמונה כלשהי, נשאל מהי פעולת ה-LUT אשר תמיר את ההיסטוגרמה שלה להיות היסטוגרמה אחרת נתונה? מדוע בעיה זו חשובה? ביישומים מסוימים כגון מציאת תנועה בין שתי תמונות, פסיפס של תמונות ועוד, משימה זו חשובה. הרעיון אם כך מכליל את הבעיה הקודמת, כך שהפעם אנו נדרשים לעצב את ההיסטוגרמה לפי צורה כלשהי.

האם פעולת LUT יכולה להגשים את היעדים הללו? מסתבר כי בכפוף לכמה דרישות סבירות, פעולות זו אפשריות. לשם בניית אלגוריתם למימוש משימה זו, נניח כי פונקציית ההיסטוגרמה הינה פונקציה רציפה מעל רצף הערכים בתחום $[0, 1]$. לכן, כל הניתוח שנתאר יתמקד בטיפול בפונקציות מעל הרצף. לשם פתרון הבעיות שהוצגו, נזדקק לתוצאה מתורת ההסתברות על הדרך בה משתנה חוק פילוג ע"י מערכת חסרת זיכרון.

נציג את הבעיה המתמטית באופן ממוקד: נתון משתנה אקראי x (זוהי רמת האפור שלנו) עם פונקציית צפיפות הסתברות $P_f(x)$ (וזוהי כמובן היסטוגרמת תמונת הכניסה f). המשתנה האקראי x עובר דרך מערכת לא ליניארית חסרת זיכרון כלשהי T והפיכה (כלומר חד-חד ערכית, ולמעשה, מונוטונית עולה) ומתקבל המוצא $y = T\{x\}$ (זוהי פעולת ה-LUT). השאלה היא מהי פונקציית צפיפות ההסתברות של המשתנה האקראי y , אותה נסמן ב- $P_g(y)$ (תמונת המוצא היא g)? הפונקציה $P_g(y)$ היא הנגזרת של פונקציית הצטברות ההסתברות, כמתואר במשוואה הבאה:

$$P_g(\alpha) = \frac{d}{d\alpha} \int_{-\infty}^{\alpha} P_f(\xi) d\xi$$

כיוון ש- T הפיכה, נוכל לומר כי $\alpha = T\{T^{-1}\{\alpha\}\}$ ולכן:

$$P_g(\alpha) = \frac{d}{d\alpha} \int_{-\infty}^{\alpha} P_f(\xi) d\xi = \frac{d}{d\alpha} \int_{-\infty}^{T^{-1}\{\alpha\}} P_f(\xi) d\xi = \frac{d}{d\alpha} \left[\int_{-\infty}^{T\{\beta\}} P_f(\xi) d\xi \right]_{\beta=T^{-1}\{\alpha\}}$$

המעבר האחרון הינו פשוט פעולת הצבה, ואין בו כל צעד אלגברי מהותי. אבל קיים הקשר הבא:

$$\int_{-\infty}^{T\{\beta\}} P_g(\xi) d\xi = \int_{-\infty}^{\beta} P_f(\xi) d\xi$$

כיוון שאגף ימין קובע מהו "מספרם" של הפיקסלים בעלי רמת אפור קטנה מ- β בתמונת הכניסה f , ואילו אגף שמאל פירושו "מספרם" של הפיקסלים בעלי רמת אפור קטנה מ- $T\{\beta\}$ בתמונת המוצא g . כיוון שהפעולה T מונוטונית עולה, הרי ששני ערכים אלו זהים לחלוטין. הצבת קשר זה נותנת:

$$P_g(\alpha) = \frac{d}{d\alpha} \left[\int_{-\infty}^{\beta} P_f(\xi) d\xi \right]_{\beta=T^{-1}\{\alpha\}} = \frac{d\beta}{d\alpha} \frac{d}{d\beta} \left[\int_{-\infty}^{\beta} P_f(\xi) d\xi \right]_{\beta=T^{-1}\{\alpha\}}$$

כשעשינו שימוש בגזירה בשרשרת. הנגזרת על-פי β פשוטה כיוון שמתקיים:

$$P_f(\beta) = \frac{d}{d\beta} \int_{-\infty}^{\beta} P_f(\xi) d\xi$$

ולכן, הקשר הנ"ל הופך ל:

$$P_g(\alpha) = \frac{d\beta}{d\alpha} \frac{d}{d\beta} \left[\int_{-\infty}^{\beta} P_f(\xi) d\xi \right]_{\beta=T^{-1}\{\alpha\}} = \frac{dT^{-1}\{\alpha\}}{d\alpha} \cdot P_f(T^{-1}\{\alpha\})$$

גזירה של היפוכה של פונקציה נתונה ע"י הנוסחה הבאה:

$$\frac{d}{d\alpha} T^{-1}\{T\{\alpha\}\} = \frac{d\alpha}{d\alpha} = 1 \quad \stackrel{\uparrow}{=} \quad \frac{d}{d\beta} T^{-1}\{\beta\} \Big|_{\beta=T\{\alpha\}} \cdot \frac{dT\{\alpha\}}{d\alpha}$$

Chain Rule

$$\Rightarrow \frac{d}{d\beta} T^{-1}\{\beta\} = \left[\frac{1}{\frac{dT\{x\}}{dx}} \right]_{x=T^{-1}\{\beta\}} \Rightarrow \frac{dT^{-1}\{\alpha\}}{d\alpha} = \frac{1}{\left[\frac{dT\{x\}}{dx} \right]_{x=T^{-1}\{\alpha\}}}$$

ולכן, התוצאה הסופית היא הביטוי הבא:

$$P_g(\alpha) = \frac{P_f(T^{-1}\{\alpha\})}{\left[\frac{dT\{x\}}{dx} \right]_{x=T^{-1}\{\alpha\}}}$$

דוגמה 5.1: נמחיש את הביטוי שהתקבל דרך מספר דוגמאות. עבור הפונקציה $T\{x\} = x + 1$, הנוסחה הנ"ל נותנת

$$P_g(\alpha) = \frac{P_f(T^{-1}\{\alpha\})}{\left[\frac{dT\{x\}}{dx}\right]_{x=T^{-1}\{\alpha\}}} = \frac{P_f(\alpha-1)}{1} = P_f(\alpha-1)$$

וזו תוצאה צפויה ואינטואיטיבית, לפיה ההסתברות לקבל ערך כלשהו במוצא זהה לחלוטין לקבלת אותו ערך פחות 1 בכניסה. באופן דומה, עבור הקשר $T\{x\} = 2x + 1$ מתקבל:

$$P_g(\alpha) = \frac{P_f(T^{-1}\{\alpha\})}{\left[\frac{dT\{x\}}{dx}\right]_{x=T^{-1}\{\alpha\}}} = \frac{P_f(0.5[\alpha-1])}{2} = \frac{1}{2} P_f\left(\frac{\alpha-1}{2}\right)$$

עבור הקשר $T\{x\} = \exp(3x)$ מתקבל:

$$P_g(\alpha) = \frac{P_f(T^{-1}\{\alpha\})}{\left[\frac{dT\{x\}}{dx}\right]_{x=T^{-1}\{\alpha\}}} = \frac{P_f\left(\frac{1}{3}\log\{\alpha\}\right)}{\left[3\exp\{x\}\right]_{x=\frac{1}{3}\log\{\alpha\}}} = \frac{P_f\left(\frac{1}{3}\log\{\alpha\}\right)}{3 \cdot \exp\left\{\frac{1}{3}\log\{\alpha\}\right\}}$$

נחזור כעת ליעד הראשון אותו רצינו להגשים: רצוננו למצוא את T אשר ייתן כי היסטוגרמת המוצא קבועה. עבור ערכי מוצא בתחום $[0,1]$ מתקבל כי נדרש $P_g(\alpha) = 1$. לכן:

$$1 = \frac{P_f(T^{-1}\{\alpha\})}{\left[\frac{dT\{x\}}{dx}\right]_{x=T^{-1}\{\alpha\}}} \Rightarrow P_f(T^{-1}\{\alpha\}) = \left[\frac{dT\{x\}}{dx}\right]_{x=T^{-1}\{\alpha\}}$$

כיוון ששני אגפי המשוואה הם פונקציות של $T^{-1}\{\alpha\}$, ומכיוון שפונקציה זו חד-חד-ערכית, נוכל להשמיטה ולקבל את הקשר:

$$P_f(\alpha) = \frac{dT\{\alpha\}}{d\alpha} \Rightarrow T\{\alpha\} = \int_{-\infty}^{\alpha} P_f(x) dx$$

דוגמה 5.2: ננסה להמחיש כיצד פילוג לא אחיד מומר לאחיד באמצעות הקשר שהתקבל. נניח כי

$$P_f(x) = \begin{cases} 2x & 0 \leq x \leq 1 \\ 0 & \text{Otherwise} \end{cases}$$

נשים לב לכך שהפונקציה בעלת אינטגרל כולל השווה 1 כנדרש. לפי האמור לעיל, עלינו להפעיל LUT מהצורה:

$$T\{\alpha\} = \int_{-\infty}^{\alpha} P_f(x) dx = \begin{cases} 0 & \alpha < 0 \\ \alpha^2 & 0 \leq \alpha \leq 1 \\ 1 & 1 < \alpha \end{cases}$$

על מנת להשתכנע ש-LUT זה "יספק את הסחורה", נמחיש זאת דרך סימולציית Matlab. ניצור תמונה בעלת היסטוגרמה השווה בקירוב לזו הנתונה כאן (אלכסונית), נסתכל על היסטוגרמת התמונה, ועל היסטוגרמת התמונה בריבוע. לשם ניטרול תופעת הקוונטיזציה בבדיקה זו, בחרנו 500 ערכים במרווחים של 0.002 באינטרוול [0,1]. תוצאות בדיקה זו מופיעות בציור 5.8, וכפי שניתן לראות, מתקבל שיווי היסטוגרמה. הסיבה לחריגות מהתיאוריה היא הדיסקרטיזציה של התהליך, עליה עוד נדבר בהמשך.

דוגמה 5.3: בדומה לדוגמה קודמת, נטפל הפעם במקרה המנוון בו הכניסה בעלת היסטוגרמה אחידה. לכן:

$$P_f(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & \text{Otherwise} \end{cases}$$

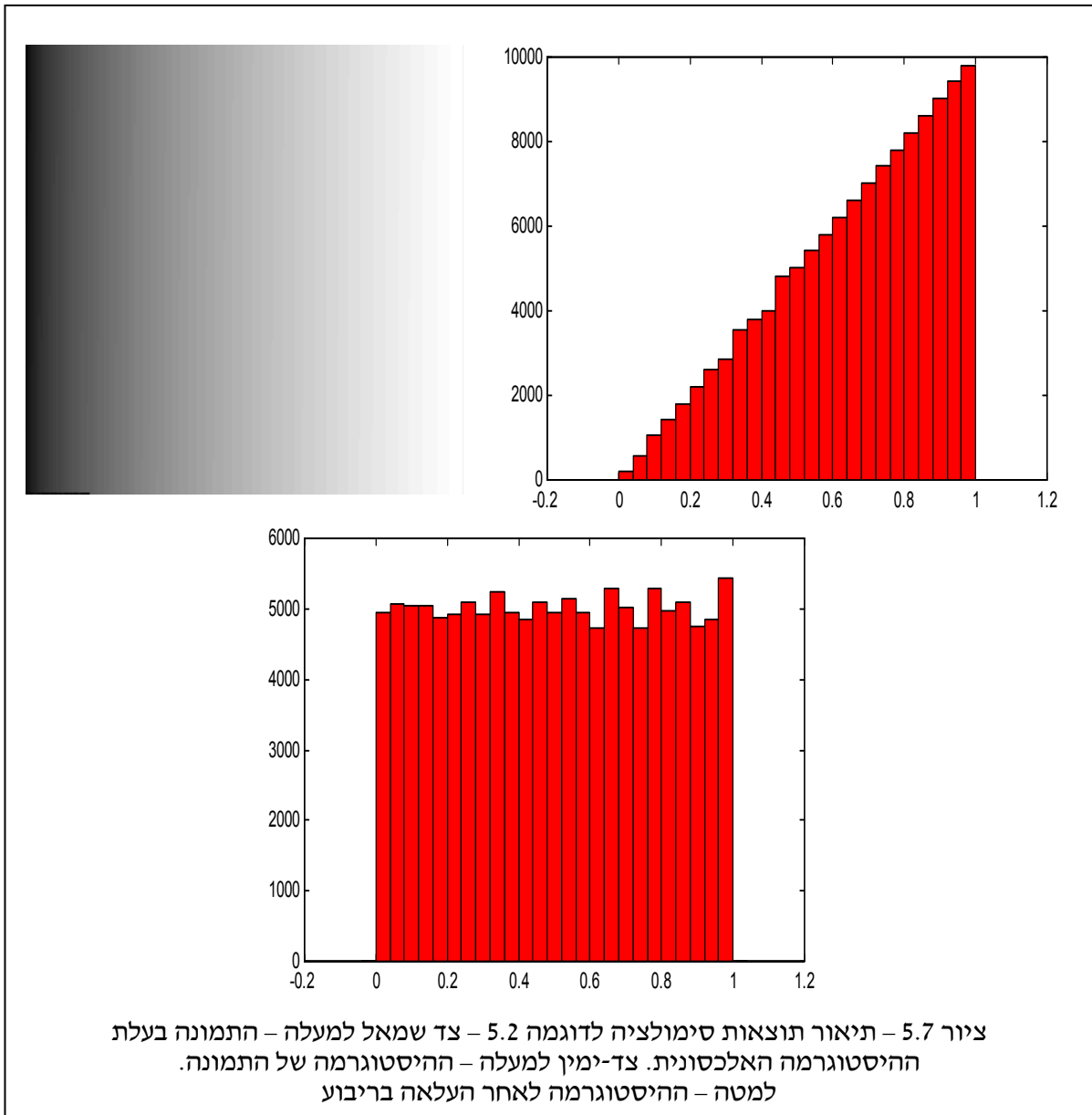
ה-LUT הנדרש הוא:

$$T\{\alpha\} = \int_{-\infty}^{\alpha} P_f(x) dx = \begin{cases} 0 & \alpha < 0 \\ \alpha & 0 \leq \alpha \leq 1 \\ 1 & 1 < \alpha \end{cases}$$

ולענייננו, זוהי בדיוק מערכת יחידה אשר אינה משנה את התמונה. תוצאה זו רצויה כיוון שהמוצא מובטח אף הוא להיות בעל היסטוגרמה אחידה.

נחזור כעת אל התיאוריה, ונטפל בבעיית התאמת היסטוגרמות בין תמונות. הפעם, בהינתן תמונה בעלת היסטוגרמת כניסה $P_f(\alpha)$ רצוננו לעצב את היסטוגרמת המוצא לפי בחירה קבועה מראש $P_g(\alpha)$. שימוש בקשרים שהושגו קודם מניב:

$$P_g(\alpha) = \frac{P_f(T^{-1}\{\alpha\})}{\left[\frac{dT\{x\}}{dx} \right]_{x=T^{-1}\{\alpha\}}} \Rightarrow P_g(T\{\alpha\}) = \frac{P_f(\alpha)}{\frac{dT\{\alpha\}}{d\alpha}} \Rightarrow \frac{dT\{\alpha\}}{d\alpha} P_g(T\{\alpha\}) = P_f(\alpha)$$



כששוב עשינו שימוש בעובדה שההתאמה T הפיכה. אינטגרציה על שני האגפים תיתן :

$$\int_{-\infty}^{\alpha} \frac{dT\{x\}}{dx} P_g(T\{x\}) dx = \int_{-\infty}^{\alpha} P_g(T\{x\}) dT\{x\} \stackrel{y=T\{x\}}{=} \int_{-\infty}^{T\{\alpha\}} P_g(y) dy = \int_{-\infty}^{\alpha} P_f(x) dx$$

את הקשר הני"ל ראינו כבר בעבר. נשתמש בסימונים הבאים לפונקציות הצטברות ההסתברות :

$$\int_{-\infty}^{\alpha} P_g(y) dy = F_g(\alpha) \quad \int_{-\infty}^{\alpha} P_f(x) dx = F_f(\alpha)$$

ונקבל:

$$\int_{-\infty}^{T\{\alpha\}} P_g(y) dy = F_g(T\{\alpha\}) = \int_{-\infty}^{\alpha} P_f(x) dx = F_f(\alpha)$$

$$\Rightarrow T\{\alpha\} = F_g^{-1}(F_f(\alpha))$$

עבור המקרה המנוון בו היסטוגרמת המוצא אחידה, מתקבל כי $F_g(x) = x$, ולכן המיפוי הינו פשוט פונקצית הצטברות ההסתברות, כפי שקיבלנו קודם. נמחיש את התוצאה שהתקבלה למקרה מורכב מעט יותר.

דוגמה 5.4: עבור שתי פונקציות ההסתברות הבאות:

$$P_g(x) = \begin{cases} 2x & 0 \leq x \leq 1 \\ 0 & \text{Otherwise} \end{cases} \Rightarrow F_g(x) = \begin{cases} x^2 & 0 \leq x \leq 1 \\ 0 & \text{Otherwise} \end{cases}$$

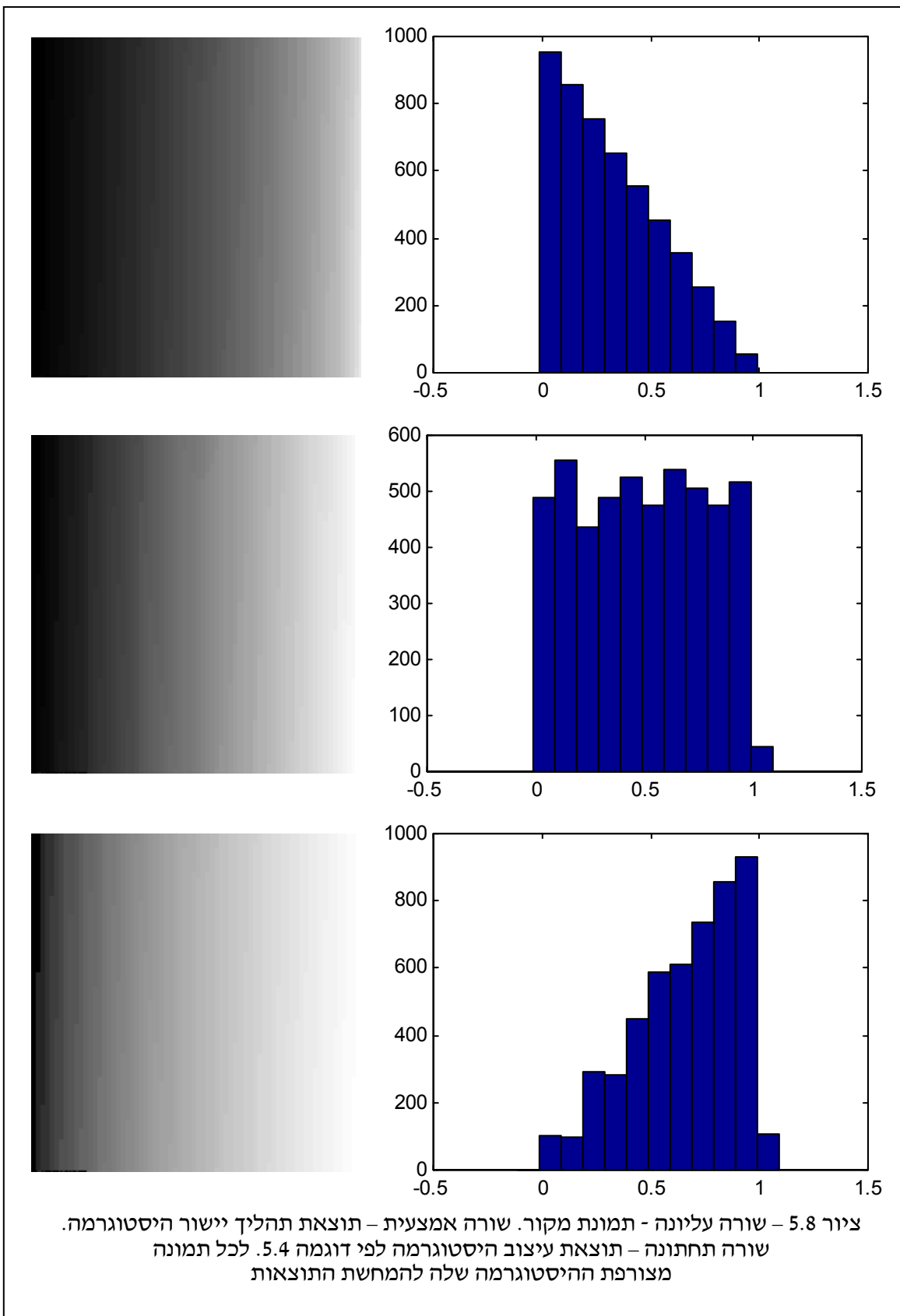
$$P_f(x) = \begin{cases} 2(1-x) & 0 \leq x \leq 1 \\ 0 & \text{Otherwise} \end{cases} \Rightarrow F_f(x) = \begin{cases} 2x - x^2 & 0 \leq x \leq 1 \\ 0 & \text{Otherwise} \end{cases}$$

מתקבל:

$$F_g(x) = \begin{cases} x^2 & 0 \leq x \leq 1 \\ 0 & \text{Otherwise} \end{cases} \Rightarrow F_g^{-1}(x) = \begin{cases} \sqrt{x} & 0 \leq x \leq 1 \\ 0 & \text{Otherwise} \end{cases}$$

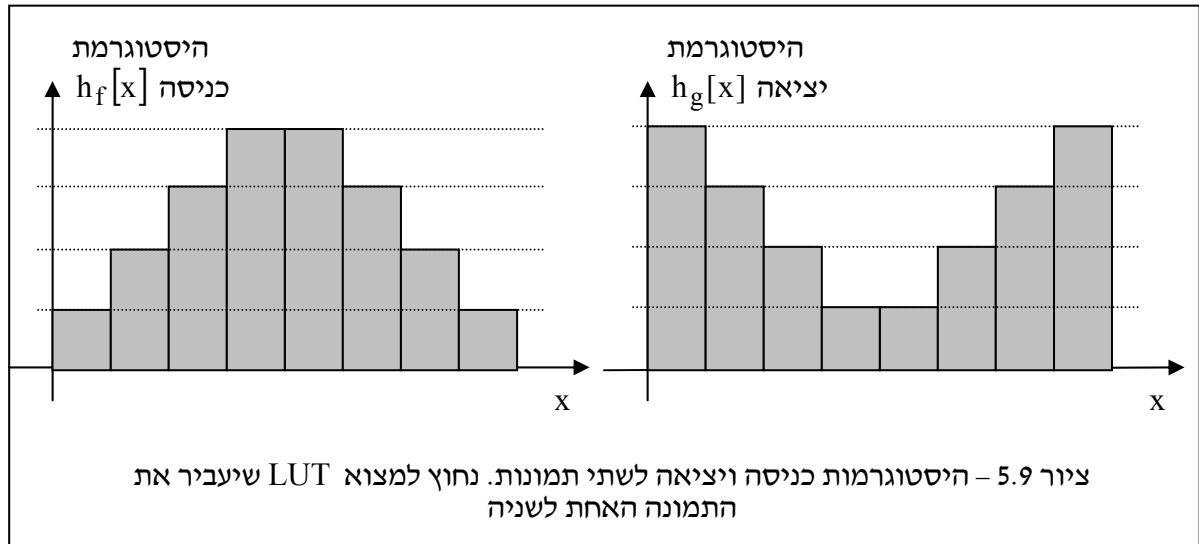
$$\Rightarrow F_g^{-1}(F_f(x)) = \begin{cases} \sqrt{2x - x^2} & 0 \leq x \leq 1 \\ 0 & \text{Otherwise} \end{cases}$$

ציור 5.8 מראה את תמונת המקור בעלת ההיסטוגרמה האלכסונית, בה הכהים דומיננטיים (שורה עליונה), תוצאת יישור ההיסטוגרמה ע"י ה-LUT $Out_1 = T_1(In) = In \cdot (2 - In)$ (שורה שניה), ותוצאת ה-LUT לעיצוב ההיסטוגרמה כנדרש בדוגמה - $Out_2 = T_2(In) = \sqrt{In \cdot (2 - In)}$. ההיסטוגרמות נראות כנדרש, למעט הפרעות הנובעות הדיסקרטיזציה של ערכי רמות האפור.



נדון כעת בעיצוב היסטוגרמות לתמונות אחרי קוונטיזציה. כפי שראינו דרך הדוגמאות, דיסקרטיזציה (גם במובן של קיומם של ערכי רמות אפור דיסקרטיות, וגם במובן של כמות פיקסלים סופית) גורמת לבעיות ולחריגות. נמחיש את דרך העבודה, ואת הבעיות המתעוררות דרך דוגמה, וכך נפתח אלטרנטיבה דיסקרטית לגישה האנליטית שפותחה.

דוגמה 5.5: נתונות שתי תמונות בנות 20 פיקסלים כל אחת, עם 8 רמות אפור אפשריות (כלומר – כל פיקסל מיוצג ב- 3 סיביות). ההיסטוגרמות לשתי תמונות אלה מתוארות בציור 5.9. כל ריבוע מייצג פיקסל בודד, וניכר שישנם בדיוק 20 ריבועים כאלה.

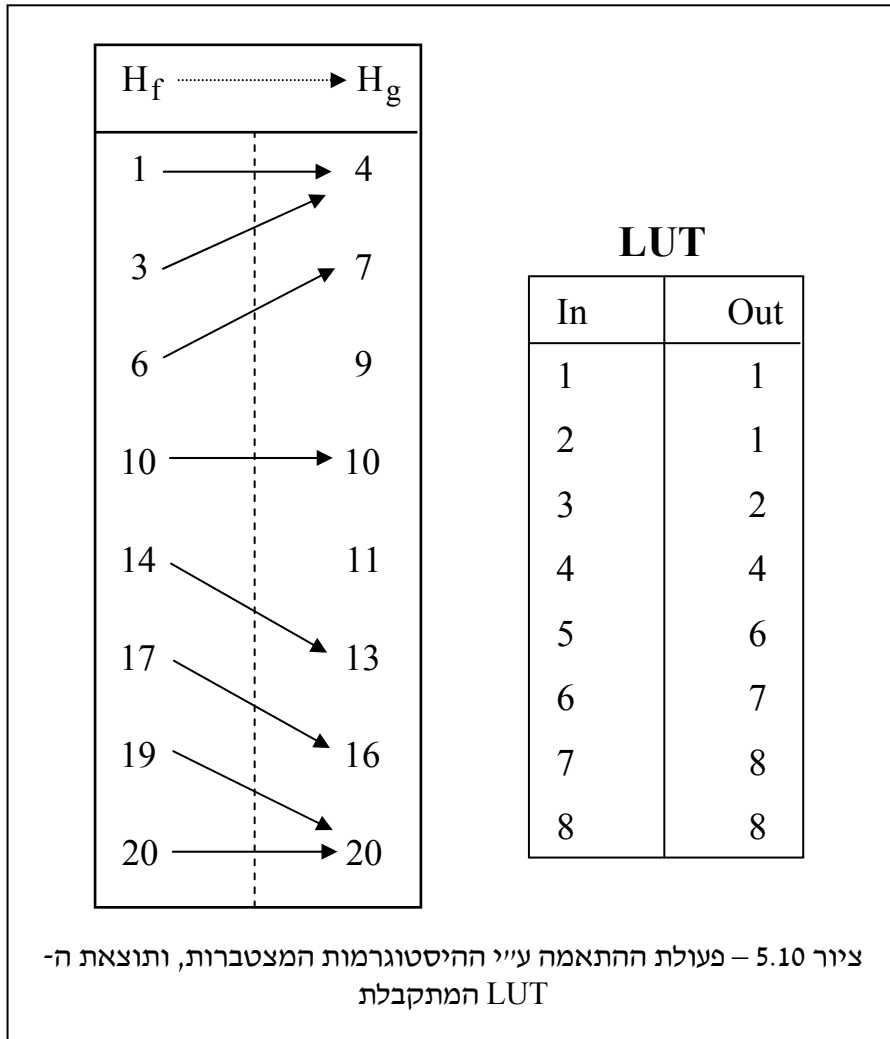


פעולת ה-LUT במקרה זה היא טבלת התאמה פשוטה בה יש 8 שורות – לכל ערך כניסה אפשרי (ויש 8 כאלו), עלינו להציע מה יהיה המוצא. בניית ה-LUT תיעשה באופן סידרתי מהערך הנמוך לגבוה. נוח יותר ונכון יותר לעבוד עם ההיסטוגרמות המצטברות:

$$H_f[x] = \{1 \ 3 \ 6 \ 10 \ 14 \ 17 \ 19 \ 20\}$$

$$H_g[x] = \{4 \ 7 \ 9 \ 10 \ 11 \ 13 \ 16 \ 20\}$$

הרעיון המרכזי הוא למצוא LUT בעל שגיאה מצטברת מינימלית. ערכי כניסה 1 ו-2 (שני הנמוכים) ימופו לרמת מוצא 1. כך נקבל כי היסטוגרמת המוצא המצטברת בעלת גובה 3 בעמודה ראשונה – סטייה של 1 מהדרוש. הוספת רמת כניסה 3 למוצא 1 תיתן גובה של 6 לעמודת המוצא הראשונה, ולכן לחריגה של 2 – לכן עדיף למפות כניסה 2 למוצא 2. כך נמשך התהליך כשאנו בודקים בכל שלב מהי ההחלטה שתיתן שגיאה מינימלית. למעשה, גישה פשוטה לבניית ה-LUT היא כתיבת שתי ההיסטוגרמות המצטברות זו מול זו, ומיפוי של כל ערך בהיסטוגרמת הכניסה (המצטברת) לערך הקרוב ביותר בהיסטוגרמת המוצא המצטברת. ציור 5.10 מראה תהליך זה. בהינתן המיפוי, ה-LUT פשוט קושר את האינדקסים, כך שלמשל, החץ הראשון קושר את האיבר הראשון בכניסה לאיבר הראשון בכניסה, ולכן ב-LUT, 1 יעבור ל-1.

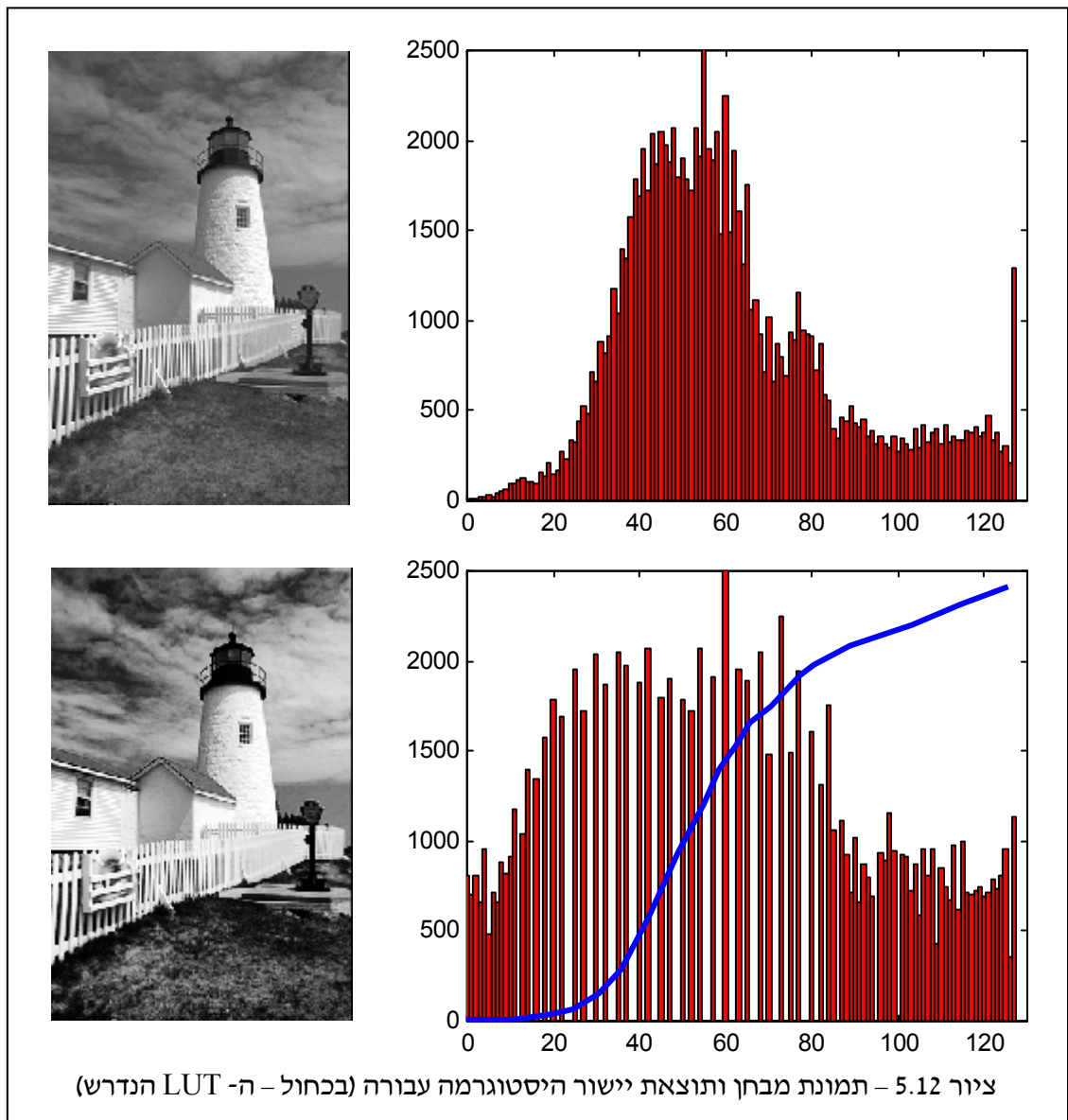
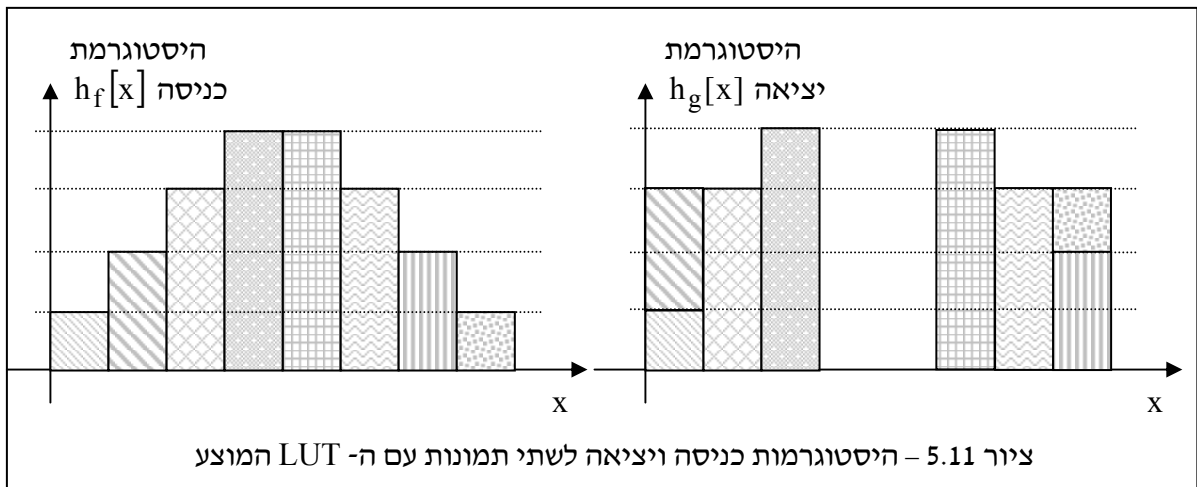


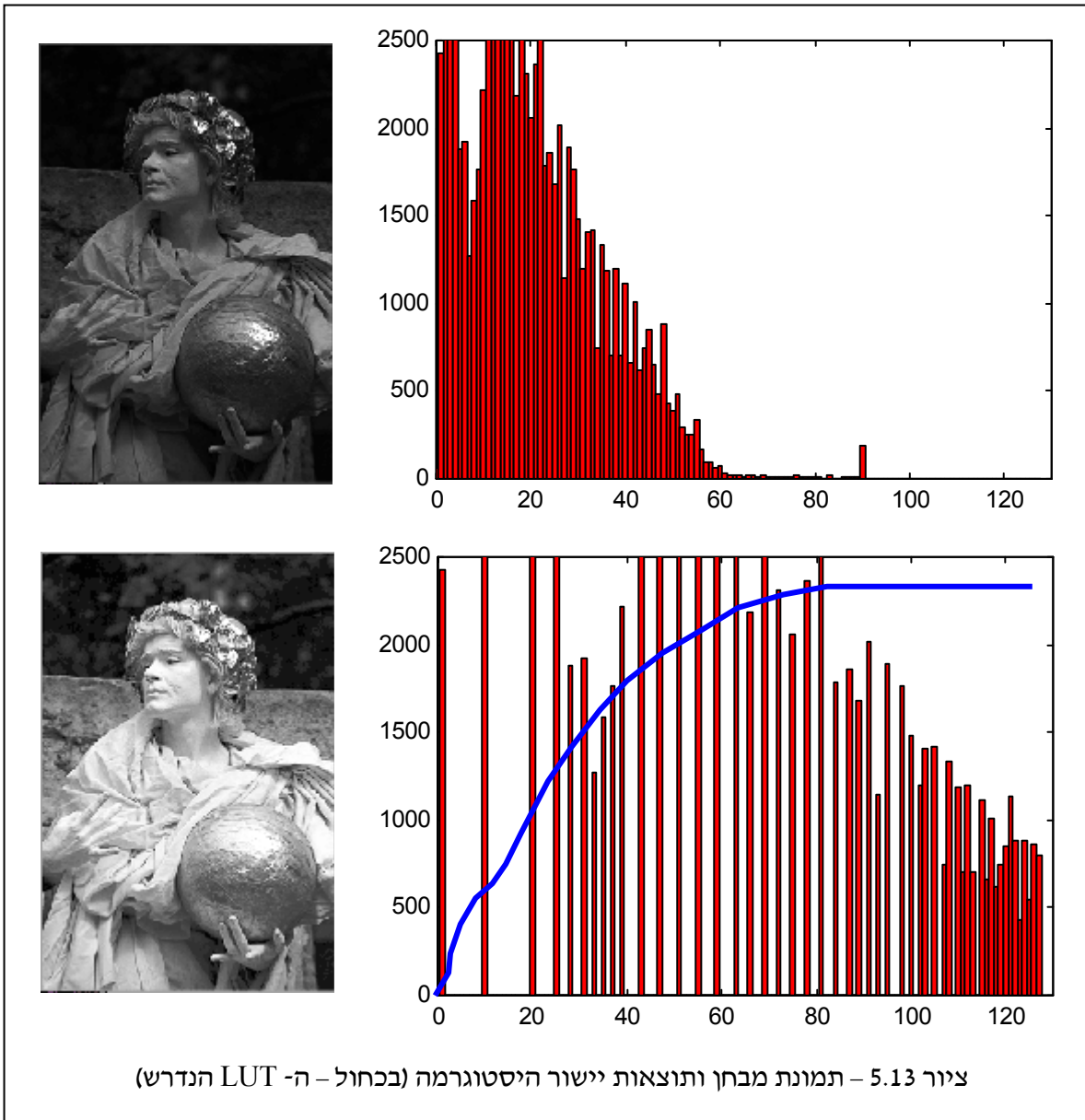
ציור 5.11 מראה את תוצאת ההיסטוגרמה ע"י ה-LUT המוצע. כפי שרואים, התוצאה אינה משיגה את היסטוגרמת היעד הנדרשת, בשל מגבלות הדיסקרטיזציה. באופן כללי, האלגוריתם שתואר לעיתים קרובות נותן אפקטים של False Contouring (היווצרם של עקומים מלאכותיים) בתמונה. זה קורה כאשר שתי רמות אפור סמוכות יכולות כתוצאה מהתהליך להתרחק זו מזו ולתת שפה בתמונה.

קיימות מספר אלטרנטיבות לאלגוריתם הנ"ל, הבאות לשפר את האלגוריתם הנ"ל, ולמנוע אפקטים לא רצויים הנובעים מהיות התמונה דיסקרטית. פתרון אפשרי אחד הוא תוספת רעש לתמונה. הוספת רעש יכולה להיעשות לפני או אחרי הפעולה T, ובאופן דומה לדיון בקוונטיזציה, גם כאן מושג אפקט של ניטרול העקומים המלאכותיים. אפשרות שניה היא להכניס מרכיב הסתברותי בפעולת ה-LUT. למשל, אם נתייחס לדוגמה 5.5, לאחר מיפויים של 1 ו-2 למוצא 1, חסר פיקסל אחד בעמודת המוצא (שגובהה 4 פיקסלים). רמת הכניסה 3, לעומת זאת, מאוכלסת ב-3 רמות אפור. לכן, אם בהסתברות 0.3333 נעביר את הכניסה 3 ל-1, ובהסתברות 0.6667 נעביר למוצא 2, נקבל היסטוגרמת מוצא כנדרש (בממוצע). ישם עוד פתרונות מגוונים, אך אנו נסתפק באלה.

פרק 5 – עיבודים בסיסיים לתמונות

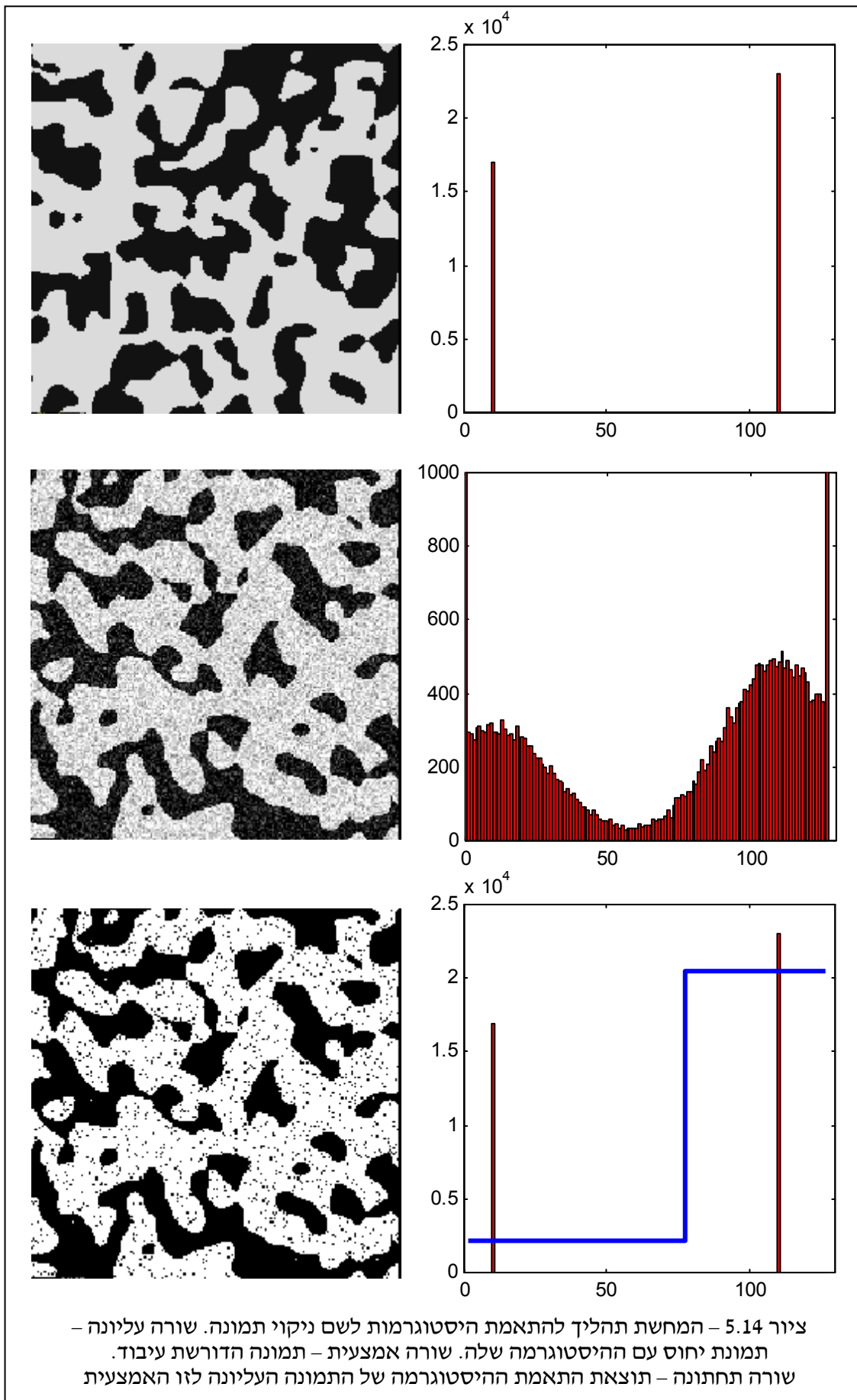
כצעד אחרון בפרק זה, נראה מספר דוגמאות להמחשת תוצאות תהליך יישור/התאמת ההיסטוגרמה. ציורים 5.12 ו-5.13 מספקים תמונת מקור ותוצאת יישור היסטוגרמה. הגרף הכחול מתאר את ה-LUT הנדרש (בכל המקרים, פיקסל מיוצג ע"י 7 סיביות).





ציור 5.14 מראה תהליך של התאמת היסטוגרמות. נתונות שתי תמונות בעלות אופי דומה (שורות עליונה ואמצעית בציור), כשאחת מהן מזוהמת, ונדרש לנקותה. לוקחים את התמונה המקולקלת ומבצעים עליה פעולת LUT לשם קבלת היסטוגרמה זהה לזו של התמונה הנקיה. התוצאה מוצגת בשורה התחתונה, ורואים כי אמנם ההיסטוגרמה במוצא התהליך זהה להיסטוגרמת הייחוס. התמונה נקיה יותר, ובעצם, הפעולה שנמצאה להיות היא השוואה לסף כלשהו, כמתואר בגרף הכחול בציור.

לסיכום דיון זה חשוב לומר כי יישור התאמת היסטוגרמה אינו בהכרח משפר את התמונה, וקורה לא אחת שהתמונה אף ניזוקה. תמונות בהן לא צפוי פיזור אחיד של רמות אפור הן תמונות בהן תהליך יישור היסטוגרמה צפוי להזיק. זה נכון למשל לתמונות בהן צילום של גרפיקה/טקסט הנראות כמעט בינאריות, תמונות בהן גוון אחד דומיננטי מאוד, תמונות בהן מטרות קטנות יחסית אך חשובות (בשל גודלן, אין להן השפעה על היישור, אך הן מושפעות ממנו בפועל).



ציור 5.14 – המחשת תהליך להתאמת היסטוגרמות לשם ניקוי תמונה. שורה עליונה – תמונת יחוס עם ההיסטוגרמה שלה. שורה אמצעית – תמונה הדורשת עיבוד. שורה תחתונה – תוצאת התאמת ההיסטוגרמה של התמונה העליונה לזו האמצעית

3. פעולות מרחביות בסיסיות

עד כה עסקנו בפעולות LUT אשר הינן פעולות חסרות זיכרון מרחבי - ערך מוצא המערכת תלוי בערך הסקלארי של ערך הכניסה באותו מקום. יתרונה של פעולת LUT בפשטותה, אך ברור כי מגוון התוצאות הניתנות להשגה באמצעותה מוגבל למדי. פעולה כללית מעט יותר אשר עדיין תהיה פשוטה יחסית היא פעולה מרחבית הפועלת על סביבת שכנים מצומצמת. כלומר, מוצא המערכת הפעם יהיה תלוי בפיסקל הכניסה ומספר פיקסלים שכנים לו. פעולה כזו יכולה להיות ליניארית או לא ליניארית, קבועה או משתנה במקום, ופעולת LUT הופכת להיות מקרה פרטי של מערכת מעין זו. במסגרת סעיף זה נתמקד בפעולות קבועות במקום (ליניאריות ושאינן ליניאריות).

באופן כללי, אם אנו דנים במערכת ליניארית על סביבה סמוכה לפיקסל, אנו בעצם דנים במסנן FIR דו-מימדי. ואמנם, ברוב מוחלט של המקרים, סינון ליניארי של תמונה נעשה ע"י מסנני FIR. מסנן כזה יאופיין ע"י תגובתו להלם אשר ניתנת לתיאור כסדרת המשקלות המכפילות כל פיקסל בסביבה, המתוארות במטריצה דו-מימדית הקרויה גרעין (Kernel). נמחיש זאת דרך דוגמה.

דוגמה 5.6: אם פעולת הסינון על הפיקסל $[m,n]$ תהיה

$$\text{Out}[m,n] = \text{In}[m,n]\alpha_1 + \text{In}[m-1,n]\alpha_2 + \text{In}[m,n-1]\alpha_3 + \text{In}[m+1,n]\alpha_4 + \text{In}[m,n+1]\alpha_5$$

אזי הגרעין של מסנן ה-FIR נתון ע"י

$$K = \begin{bmatrix} 0 & \alpha_2 & 0 \\ \alpha_3 & \alpha_1 & \alpha_5 \\ 0 & \alpha_4 & 0 \end{bmatrix}$$

כאשר הראשית במישור האינדקסים מצויה למעלה בצד שמאל, והגרעין מופעל על התמונה ע"י הנחתו כמסכה במקום $[m,n]$ וחישוב סכום המכפלות בין המשקלות וערכי הפיקסלים של התמונה המטופלת. איבר האמצע בגרעין מתייחס למיקום $[m,n]$.

השימוש במסנן לשיפור תמונות רווח לשלושה יישומים עיקריים - ניקוי תמונה מרעש, חידוד תמונה, ומציאת שפות בתמונה. באשר לדיכוי רעש בתמונה: רעש יכול להצטרף לתמונה כתוצאה מלכלוך על תמונה שאמורה להיסרק, הפרעות שידור לתמונה מועברת בערוץ תקשורת, ועוד. בדרך כלל מקובל להניח מודל אדיטיבי (חיבורי) לפיו התמונה הרועשת $f[m,n]$ הינה סכום של התמונה המקורית $s[m,n]$ ותמונת רעש $d[m,n]$, כלומר $d[m,n] + f[m,n] = s[m,n]$. ישנן גם מערכות צילום בהן הרעש הינו דווקא מולטיפליקטיבי (מכפלתי), בו הרעש כופל את התמונה האידיאלית, אך אנו לא נעסוק בהן במסגרת זו.

בדרך כלל אות הרעש יונח להיות אקראי בעל ממוצע אפס, הומוגני ולבן. משמעות הנחה זו היא שערכי הפיקסלים בתמונת הרעש d מפולגים לפי אותו חוק פילוג בכל מקום בתמונה, לחוק פילוג זה ממוצע אפס ועוצמה (ווריאנס) קבועה, ואין קורלציה בין פיקסלים סמוכים. במלים אחרות - מתוך ידיעת ערכו של $d[m,n]$ לא נוכל להקיש דבר על ערכי שכניו. בדרך כלל אף נהוג להניח כי הרעש הוא בעל חוק פילוג גאوسی.

הערה: המושג רעש מתאים יותר לאות שמע, וזהו באמת מקורו. הפרעה באות דיבור למשל תהווה "רעש". בתמונות, לפי הגיון זה, הפרעה הייתה צריכה להיקרא "לכלוך". בפועל, המושג רעש תפס גם בתמונות, והכוונה לאות אקראי לא רצוי.

הבעיה אם כך היא לנסות ולבנות את התמונה f מתוך ידיעת התמונה s , ואת זאת לעשות ע"י סינון ליניארי. בניגוד לרעש, ידיעת ערכו של הפיקסל $f[m,n]$ מלמדת הרבה על ערכי שכניו, כיוון שכפי שכבר ראינו בפרק ראשון, ערכי פיקסלים סמוכים הם לרוב קרובים (ואולי אפילו זהים). זוהי בדיוק האבחנה שבזכותה אנו יכולים לנקות רעש מתמונה.

נתחיל את הניתוח מתוך הנחה שערכי השכנים צריכים להיות זהים, ונניח שהמסנן פועל בגרעין בגודל של 3 על 3 פיקסלים. כשאנו מסתכלים בחלון של 3 על 3 סביב הנקודה $[m,n]$, אנו מחפשים מהו הערך האחד $f[m,n]$ שיחליף את $s[m,n]$ ואשר יהיה קרוב לכל תשעת הערכים בחלון. למעשה, אנו מחפשים את המינימום לביטוי הבא:

$$\varepsilon\{f[m,n]\} = \sum_{k=-1}^1 \sum_{j=-1}^1 (s[m-k, n-j] - f[m,n])^2 \rightarrow \text{Minimum}$$

והתוצאה נתונה ע"י

$$\frac{\partial \varepsilon\{f[m,n]\}}{\partial f[m,n]} = -2 \sum_{k=-1}^1 \sum_{j=-1}^1 (s[m-k, n-j] - f[m,n]) = 0$$

$$\Rightarrow f[m,n] = \frac{1}{9} \sum_{k=-1}^1 \sum_{j=-1}^1 s[m-k, n-j]$$

כלומר, יש להפעיל על התמונה מסנן ליניארי בעל הגרעין

$$K = \frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

כשמסנן זה פועל על $s[m,n]$ (אשר כזכור הוא $s[m,n]=f[m,n]+d[m,n]$), הוא נותן סכום של שני ממוצעים - מיצוע מקומי של התמונה f ומיצוע מקומי של התמונה d . אם אמנם ב- f הערכים בסביבת 3×3 קרובים מאוד זה לזה, הרי שממוצע מקומי של f כמוהו כ- f עצמו. לעומת זאת, מיצוע של 9 משתנים אקראיים שווי פילוג הוא משתנה אקראי חדש בו השונות קטנה פי $\sqrt{9} = 3$, וכך יצא שעוצמת הרעש הוקטנה.

מדוע אם כך להסתפק בחלון של 3×3 - אם נעבוד על חלון של 5×5 על 5 נקבל דיכוי עוצמת הרעש בפקטור 5 . אבל - עם הגדלת החלון, הנחתנו על ערכי רמות אפור קבועים ב- f נשברת והולכת. וכך נקבל כי אמנם הרעש ייקטן יפה, אך עמו נקבל גם אפקט של טשטוש התמונה. ציור 5.15 מראה תוצאת הסינון עם חלונות בגודל שונה, וכפי שרואים, הרעש אמנם פוחת, אך עמו גם פירטי התמונה. האזורים בהם הכי גדול הנזק אם אזורי המעברים (שפות בתמונה), בהם הנחת רמת אפור קבועה היא הנחה שגויה.



כדי לראות אם שיפרנו, נוכל לחשב את ההפרש בין תמונת המקור $f[m,n]$ לכל אחת מהתמונות הנ"ל (עם רעש, סינון ע"י 3×3 וסינון ע"י 5×5), ולסכם את ריבועי הפרשים אלו. רצוי גם לחלק במספר

הפיקסלים בתמונה ולהוציא שורש מהתוצאה, על מנת לקבל מספר ביחידות של רמות אפור. הנוסחה אם כך היא:

$$\text{Error} = \sqrt{\frac{1}{MN} \sum_{m=1}^M \sum_{n=1}^N (f[m,n] - \hat{f}[m,n])^2}$$

כאשר $\hat{f}[m,n]$ היא התוצאה שלנו המנסה לקרב את $f[m,n]$. קריטריון זה קרוי שגיאה ריבועית ממוצעת (Mean Squared Error - MSE), והוא רווח בעיבוד תמונות כמדד לאיכות תהליכי עיבוד. ברור כי שימוש בו אפשרי רק כאשר זמינה לנו תמונת המקור (בטרם הצטרפות הרעש). ברור כי ככל שערכו של סכום זה קטן יותר השיפור טוב יותר. התוצאה למקרה הנ"ל נותנת כי השגיאה המיוחסת לתמונה הרועשת היא 15.01 - זוהי כמעט בדיוק השונות בה השתמשנו. לתמונה לאחר סינון 3 על 3 התקבלה שגיאה של 9.8, ולסינון 5 על 5 שגיאה של 13.38. בשני המקרים חל שיפור, אך סינון 3 על 3 טוב יותר. נשים לב כי אם התמונה היתה קבועה לגמרי היינו צפויים לקבל שגיאות של 5 ו-3 בקירוב טוב (לפי יחסי דיכוי רעש בלבד 3 ו-5). בשל היות התמונה לא קבועה, השגיאה גדולה יותר, וזו הולכת ומחריפה עם הגדלת החלון.

שאלה מעניינת היא השאלה הבאה: עברנו מחלון של 1×1 (זהו המקרה בו בעצם לא הופעל סינון), לחלון בגודל 3×3 , וממנו לחלון בגודל 5×5 , כשבכל המקרים קיבלנו מסנן אחיד (ערכים זהים בגרעין). האם אנו יכולים להציע דרך אותה גישה מגוון רחב יותר של פילטרים? התשובה לכך נעוצה בביטוי שאותו הבאנו למינימום. אם נכניס לשגיאות הריבועיות משקלות לפי החשיבות לכל פיקסל

$$\varepsilon\{f[m,n]\} = \sum_{k=-L}^L \sum_{j=-L}^L c[k,j] (s[m-k,n-j] - f[m,n])^2 \rightarrow \text{Minimum}$$

נקבל:

$$\frac{\partial \varepsilon\{f[m,n]\}}{\partial f[m,n]} = -2 \sum_{k=-L}^L \sum_{j=-L}^L c[k,j] \cdot (s[m-k,n-j] - f[m,n]) = 0$$

$$\Rightarrow f[m,n] = \frac{\sum_{k=-L}^L \sum_{j=-L}^L c[k,j] \cdot s[m-k,n-j]}{\sum_{k=-L}^L \sum_{j=-L}^L c[k,j]}$$

והפעם הגרעין יהיה:

$$K = \frac{1}{\sum_{k=-L}^L \sum_{j=-L}^L c[k,j]} \begin{bmatrix} \ddots & & & & \\ & c[-1,-1] & c[-1,0] & c[-1,1] & \\ \dots & c[0,-1] & c[0,0] & c[0,1] & \dots \\ & c[1,-1] & c[1,0] & c[1,1] & \\ \ddots & & & & \ddots \end{bmatrix}$$

כיצד נקבע את המשקלות? גישה אפשרית אחת היא כערכים הדועכים לפי המרחק האויקלידי מהמרכז. קצב הדעיכה יקבע בפועל את התמך האפקטיבי של המסנן שיווצר (רוחבו). למשל, לחלון 3 על 3 ניתן לקבוע את המשקלות כ:

$$c[-1,-1]=c[-1,1]=c[1,-1]=c[1,1]=0.25 \text{ ו- } c[0,1]=c[1,0]=c[-1,0]=c[0,-1]=0.5, c[0,0]=1$$

והגרעין יהיה:

$$K = \frac{1}{4} \begin{bmatrix} 0.25 & 0.5 & 0.25 \\ 0.5 & 1 & 0.5 \\ 0.25 & 0.5 & 0.25 \end{bmatrix}$$

את התהליך הנ"ל יכולנו לתאר מתוך שיקולים ספקטריים, כלומר תכנון המסנן כמסנן LPF דו-מימדי במישור התדר. נראה את עיקרי הדברים כאן: לומר שהתמונה $f[m,n]$ קבועה בקירוב במקום שקול לאמירה שתמונה זו בעלת התמרת תדר הדועכת מהר לאפס כשמתרחקים מהראשית (כלומר יש ספקטרום עם תדרים נמוכים בלבד). לעומת זאת, הרעש הוא בעל ספקטרום אחיד בכל התדרים (מכאן למעשה בא המושג "רעש לבן" - לבן הוא צבע העושה שימוש בכל אורכי הגל באופן אחיד). לכן, סינון הרעש תוך כדי אי-פגיעה בתמונת המקור מחייב מסנן LPF שאינו מזיק כמעט בכלל לתמונה $f[m,n]$. לכן, המסנן לא יכול להיות LPF צר כרצוננו, אלא תלוי בהתמרת התדר של התמונה $f[m,n]$.

בקריטריון שהוצע לעיל לבניית המסנן העלינו כל שגיאה בריבוע. ניתן כתחליף להגדיר את השגיאה ע"י ערך מוחלט כ:

$$\varepsilon\{f[m,n]\} = \sum_{k=-L}^L \sum_{j=-L}^L |s[m-k, n-j] - f[m,n]| \rightarrow \text{Minimum}$$

השינוי הוא במהותו שינוי המטריקה המודדת את המרחק בין שתי תמונות, והמטריקה החדשה אינה פחות נכונה מקודמתה. הבאת ביטוי זה למינימום מניבה את המסנן הידוע בשם סינון חציון (Median).

$$\frac{\partial \varepsilon\{f[m,n]\}}{\partial f[m,n]} = -2 \sum_{k=-L}^L \sum_{j=-L}^L \text{sign}(s[m-k, n-j] - f[m,n]) = 0$$

נניח כי אוסף ה- $(2L+1)^2$ הפיקסלים בחלון סודרו בסדר עולה לפי ערכיהם (ללא כל חשיבות למקומם). אזי, הערך שיביא את הביטוי הנ"ל למינימום הוא הערך האמצעי - $(2L^2 + 2L + 1)$. למשל, עבור חלון 3 על 3 (בו $L=1$), האיבר האמצעי הוא החמישי מתוך 9. בחלון 5 על 5 יהיה זה האיבר ה-13. מדוע האיבר האמצעי דווקא? כיוון שאז מחצית מהאיברים בסכימה הנ"ל בעלי סימן שלילי, מחצית בעלי סימן חיובי, ואיבר האמצע נותן ערך אפס. סכימת הסימנים תיתן התאפסות הנגזרת כנדרש.

המסנן שהתקבל - סינון חציון - אינו ליניארי. מסנן זה שייך למשפחה של מסנני סדר (Order Filters), בהם מוצא המסנן תלוי בסידור (Sorting) של ערכי הכניסה. ניתן גם להציע בחירה של המקסימלימינימלי בחלון, וישנם יישומים העושים שימוש במסננים מעין אלו.

אז איזה סינון יותר טוב - ליניארי או חציון? התשובה תלויה בטיבו של הרעש האדיטיבי (מהו חוק הפילוג שלו), מהי הגישה לבחון את התוצאות (האם MSE או מדד אחר) ועוד. ככלל, מקובל כי ברעשי שיאים סינון חציון יותר טוב. רעש שיאים, הקרוי גם רעש מלח ופלפל (Salt & pepper) (נשמע יותר אמין משום מה), הוא רעש בו פיקסלים במיקום אקראי משנים את ערכם לשחור ולבן מוחלט באופן אקראי. בדוגמה הבאה ננסה להמחיש מדוע סינון מדיין טוב כל כך בטיפול ברעשי שיאים.

דוגמה 5.7: נניח כי אנו משווים סינון ליניארי אחיד וסינון חציון בחלון של 3 על 3. בחלון מסויים התקבלו הערכים הבאים - {50 55 60 60 65 65 70 75 75} (תשעה ערכים ומיקומם לא חשוב). ערך הממוצע יהיה ≈ 64 והחציון יהיה 65 - אין הבדל גדול. נניח כעת כי הערך 75 השתנה כתוצאה מרעש וערכו גדל ל-255. הממוצע החדש יהיה 84, בעוד שהחציון לא ישתנה כלל! כל שינוי באחד המספרים יגרום לשינוי מיידי (אך מוקטן) בתוצאת הממוצע. לעומת זאת, בחישוב החציון שינוי של אחד מארבעת הערכים הנמוכים מטה או שינוי של אחד מארבעת הגבוהים מעלה לא תשנה דבר. כמקרה שני, נניח כי חוזרים לערכים המקוריים והערך 60 עובר להיות 255. במקרה כזה, החציון יהיה עדיין 65, ואילו הממוצע האלגברי ייתן 85 - שוב שינוי משמעותי.

בציור 5.16 מתוארת תמונה אשר אליה צורף רעש Salt & pepper בה 20% מהפיקסלים "נגועים" ברעש שיאים לשני הכיוונים (שחור ולבן בהסתברות שווה). כפי שניתן לראות, מסנן חציון מצליח להתגבר על מרבית הרעש תוך שמירת מרבית הפרטים בתמונה, בעוד שהמסנן הליניארי מניב תוצאה ירודה למדי באיכותה.

חשוב לציין כי כאשר מניחים ידע כלשהו על סטטיסטיקת הרעש האדיטיבי (כגון האמירה - רעש לבן גאوسی וכו'), דיכוי הרעש יכול להיעשות מתוך שיקולים תיאורטיים מוצקים יותר, בהשוואה לגישה שתוארה כאן. אנו נראה זאת במסגרת פרק הבא, בו נדבר על שחזור תמונות.

משימה שניה שמסננים (ליניאריים ושיאנים ליניאריים) מגשימים היא פעולת חידוד תמונה. המידע העיקרי בתמונה הם שפות העצמים המופיעים בה, ולכן הדגשת שפות היא משימה חשובה. בהסתכלות בתחום התדר, שפות הם מידע בעל תדרים גבוהים (בגלל המעברים החדים), בעוד שפנים העצמים מיוחס לתדרים נמוכים. לכן, חידוד שפות פירושו מסנן אשר יעבוד על התדרים הגבוהים וידגיש אותם. זוהי בדיוק פעולה הפוכה לפעולת ניקוי הרעש ע"י LPF כפי שתוארה קודם. לכן, חידוד תמונה בא בחשבון רק עבור תמונות נקיות יחסית מרעש. העיקרון באלגוריתמי חידוד תמונה הוא להעביר את התמונה דרך HPF (מסנן מעביר גבוהים) ולחבר את מוצא מסנן זה לאחר מישקול נאות לתמונה המקורית. השאלה המתבקשת היא - כיצד מממשים HPF? ובכן, HPL הוא עצם 1-LPF, וליישם LPF אנו כבר יודעים. לדוגמה, עבור ה-LPF הפשוט בו מוצע גרעין של 3 על 3 אחיד, ה-HPL יהיה



$$K_{HPF} = 1 - K_{LPF} = 1 - \frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = \frac{1}{9} \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

מסנן HPF יהיה תמיד בעל סכום איברים השווה 0 (למה?). לכן, המסנן הכולל שיבצע את פעולת החידוד במקרה זה יהיה:

$$K_{Sharp} = 1 + K_{HPF} = 1 + \alpha \frac{1}{9} \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix} = \frac{1}{9} \begin{bmatrix} -\alpha & -\alpha & -\alpha \\ -\alpha & 9 + 8\alpha & -\alpha \\ -\alpha & -\alpha & -\alpha \end{bmatrix}$$

עבור $\alpha = 0$ נקבל את תמונת המקור בדיוק, וכש- α חיובי וגדל, נקבל אפקט חידוד חזק יותר ויותר, ומנקודה מסוימת ואילך, תתקבל תוצאה כה חדה שתיראה לא טבעית ולא יפה. ציור 5.17 מראה תוצאת חידוד תמונה למגוון ערכי α עם המסנן הני"ל. כתמונת מקור בחרנו לקחת את Lena לאחר טשטוש קל.

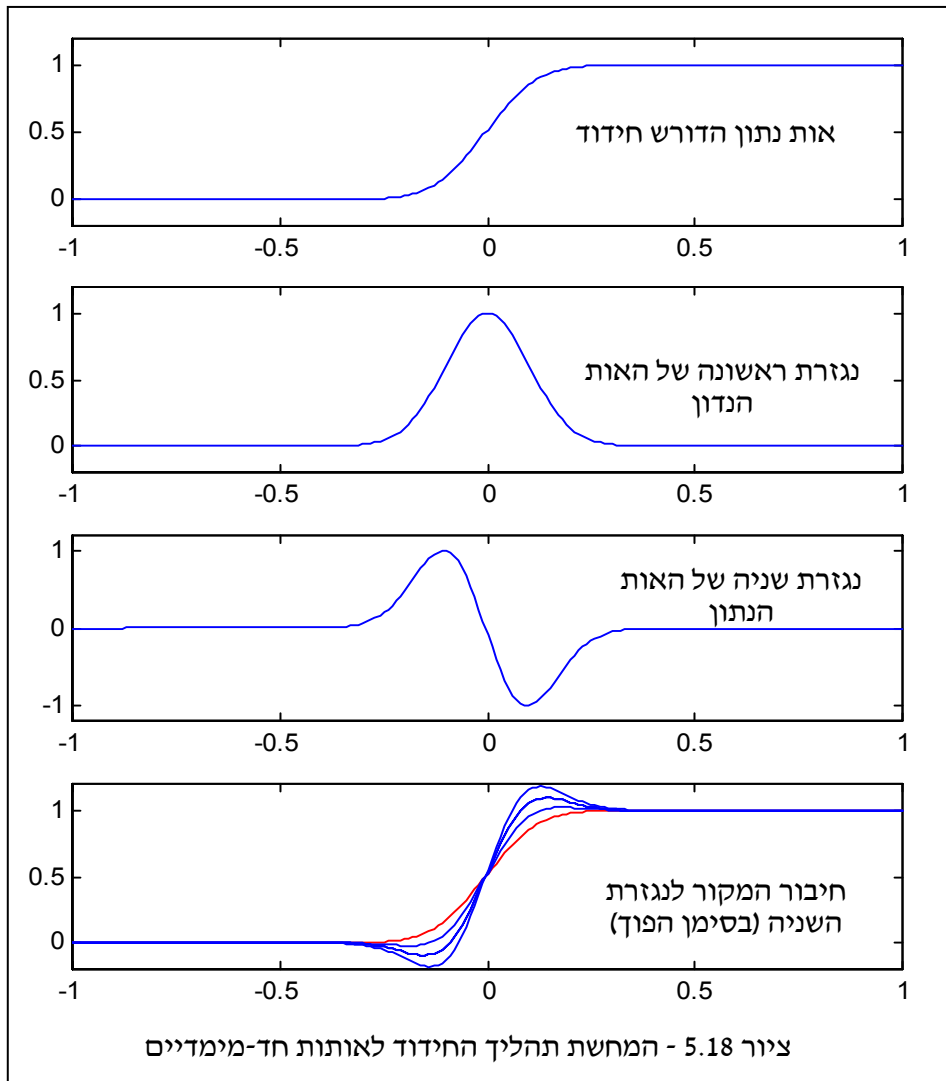


מסנן שרווח בו השימוש כמסנן HPF לטובת חידוד שפות הוא אופרטור הלפלסיאן הנתון ע"י

$$K_{\text{Laplacian}} = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix}$$

כפי שנראה בהמשך, לאופרטור זה קשר הדוק לאופרטור הנגזרת השניה הרציף.

מה קורה בתהליך החידוד? ציור 5.18 ממחיש את התהליך עבור אותות חד-מימדיים. אנו רואים שפה מטושטשת, ורואים כי נגזרת שניה המוחסרת מאות המקור נותנת אפקט של מעבר חד יותר, כנדרש. מדוע סימן שלילי לנגזרת? מדוע נגזרת שניה? לעת עתה שאלות אלו פתוחות כיוון שלא ענינו על שאלה בסיסית יותר - מה הקשר בין נגזרות (ראשונה ושניה) ומסנן ה-HPF שרצינו מלכתחילה? בכל אופן, אינטואיטיבית נוכל לומר את הדבר הבא - הפעלת מסנן כדוגמת הלפלסיאן נותנת ערך קרוב לאפס באזורים בהם רמת האפור מהווה מישור מוטה. במקומות אלו לא יורגש כלל החידוד. במקומות בהם תוצאת הלפלסיאן גבוהה משמעותית מאפס, פירוש הדבר הוא שהפיקסל הנוכחי גבוה יותר מממוצע ארבעת שכניו. כלומר - ישנו שם פרט. על מנת לחדדו, יש להגדיל עוד את ההפרש בין הפיקסל לבין שכניו, וזוהי בדיוק הפעולה המתבצעת. תהליך דומה קורה לערכים שליליים.



נעבור כעת לדון במציאת שפות בתמונה, וקירוב נגזרות מעל שריג דיסקרטי דו-מימדי. כזכור מפרק 1, אנו יכולים להתייחס לתמונה כמשטח (פונקציה) מעל המישור, כשרמת האפור $f(x,y)$ היא גובהו של המשטח בנקודה (x,y) . לפונקציה זו אנו יכולים לחשב בנקודה (x,y) את הגרדיאנט בכל נקודה. הגרדיאנט הוא וקטור הבנוי מהנגזרת האופקית והנגזרת האנכית, אשר ברצף מוגדר לפי:

$$\nabla f(x,y) = \frac{\partial f(x,y)}{\partial x} \hat{x} + \frac{\partial f(x,y)}{\partial y} \hat{y}$$

לוקטור הגרדיאנט חשיבות רבה. חישוב ערכו המוחלט יכול לומר לנו האם הנקודה בה אנו עומדים היא שפה או לא, שהרי נורמה גדולה לגרדיאנט פירושה שיפוע חזק. נורמה זו נתונה ע"י:

$$\|\nabla f(x,y)\| = \sqrt{\left(\frac{\partial f(x,y)}{\partial x}\right)^2 + \left(\frac{\partial f(x,y)}{\partial y}\right)^2}$$

אופרטור הלפלסיאן מתקבל ע"י :

$$\Delta f(x, y) = \nabla^2 f(x, y) = \frac{\partial^2 f(x, y)}{\partial x^2} + \frac{\partial^2 f(x, y)}{\partial y^2}$$

כך שאם נחזור למה שהצגנו קודם, חידוד הפונקציה, לו היה נעשה ברצף, היה מתבצע ע"י

$$g(x, y) = f(x, y) - \alpha \Delta f(x, y)$$

כל האמור לעיל טוב ויפה אך אינו מועיל בבואנו לטפל בתמונות דיסקרטיות. לכן, עלינו למצוא דרך לקרב נגזרות באות דגום. הרעיון הבסיסי פשוט עד טריויאלי. כיוון שנגזרת ברצף מייצגת הפרשים בין ערכים סמוכים במרחק אינפיטיסימלי וחלוקה במרחק זה, הצעה מתבקשת היא לקבוע את הנגזרת האופקית, למשל, כהפרש בין דגימות סמוכות באותה שורה, תוך הנחה שהמרחק ביניהם הוא 1. הבעיה היא שכשאנו עומדים על הדגימה $[m, n]$, ישנן שתי אפשרויות :

$$\frac{\partial f}{\partial x} \rightarrow f[m, n] - f[m-1, n] \text{ or } f[m+1, n] - f[m, n]$$

האפשרות הראשונה קרויה גזירה אחורית, והשנייה גזירה קדמית. באופן דומה לגזירה אנכית :

$$\frac{\partial f}{\partial y} \rightarrow f[m, n] - f[m, n-1] \text{ or } f[m, n+1] - f[m, n]$$

כתחליף להגדרות הנ"ל ניתן להציע גזירה מרכזית לפי :

$$\frac{\partial f}{\partial x} \rightarrow \frac{1}{2} [f[m+1, n] - f[m-1, n]]$$

$$\frac{\partial f}{\partial y} \rightarrow \frac{1}{2} [f[m, n+1] - f[m, n-1]]$$

בהמשך ישיר לקו הגיון זה, גזירה שניה היא הפרש על הנגזרת עצמה. לכן, גזירה שניה אופקית תהיה :

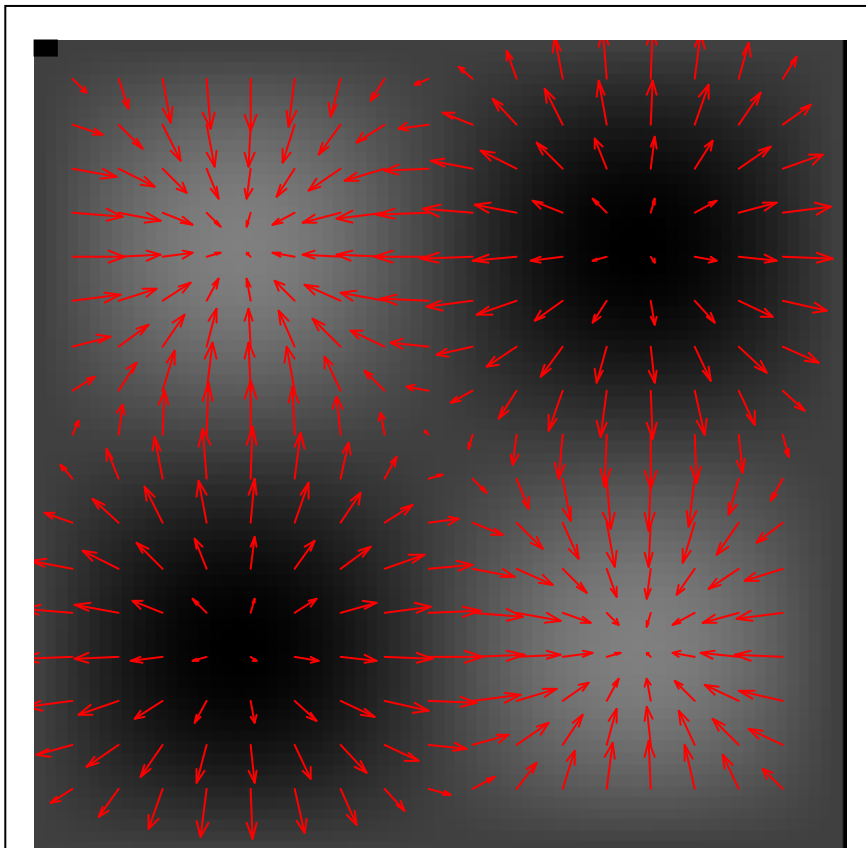
$$\frac{\partial^2 f}{\partial x^2} \rightarrow (f[m+1, n] - f[m, n]) - (f[m, n] - f[m-1, n]) = f[m+1, n] + f[m-1, n] - 2f[m, n]$$

כשהפעם עשינו שימוש בגזירה הקדמית והאחורית במשולב על מנת לקבל נגזרת שניה מאוזנת (כלומר כזו שממוקדת במקום הנכון). לפי האמור לעיל, אופרטור הלפלסיאן הדיסקרטי מתקבל ע"י :

$$\begin{aligned} \Delta f[m,n] &= (f[m+1,n] - f[m,n]) - (f[m,n] - f[m-1,n]) + \\ &\quad + (f[m,n+1] - f[m,n]) - (f[m,n] - f[m,n-1]) = \\ &= f[m+1,n] + f[m-1,n] + f[m,n+1] + f[m,n-1] - 4f[m,n] \end{aligned}$$

וכבר ראינו מסנן זה קודם לכן, כשדנו בחידוד.

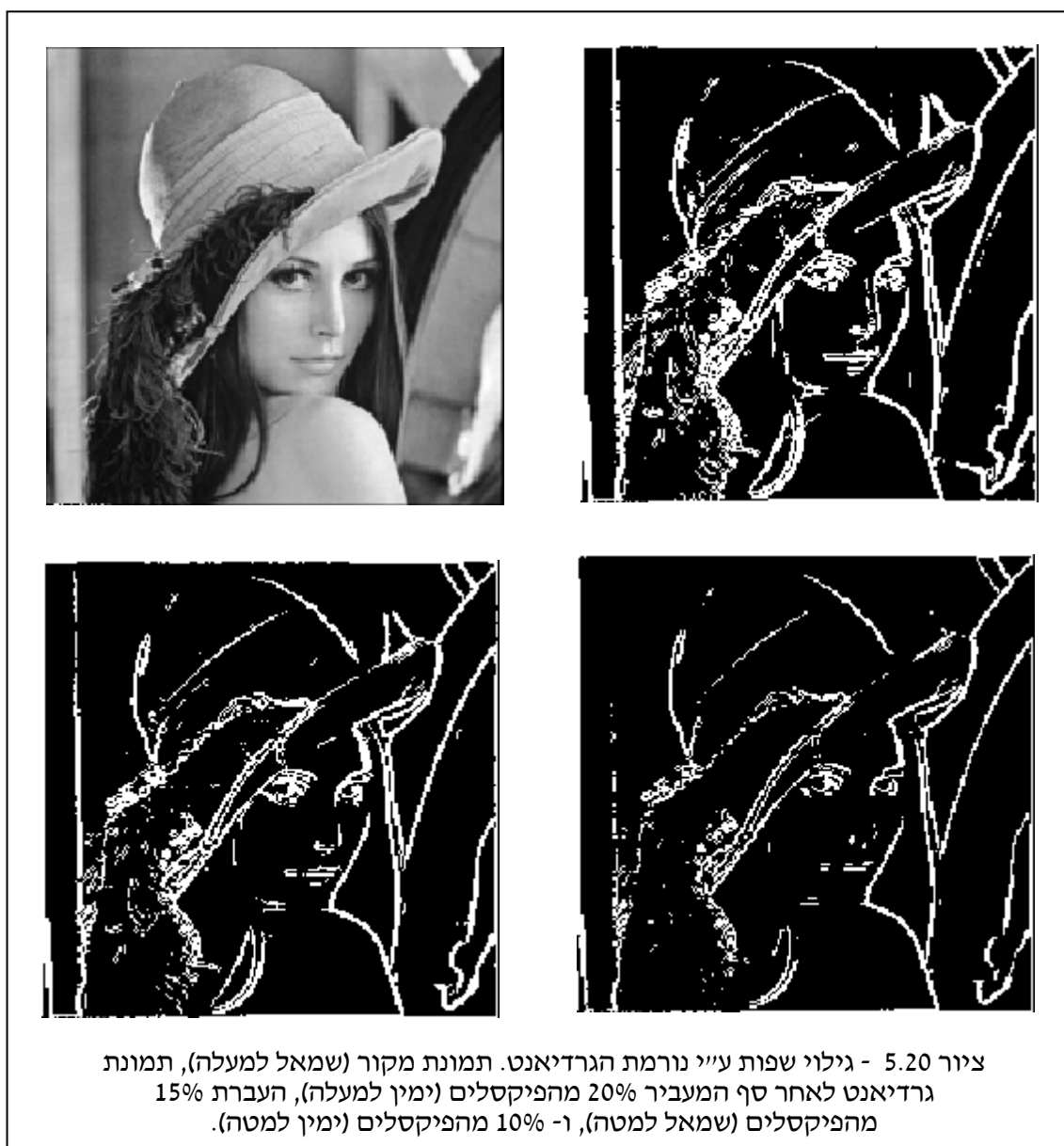
בציור 5.19 מוצגת תמונה ועליה חיצו הגרדיאנט, אשר חושבו ע"י נגזרת מרכזית. כצפוי, חיצו הגרדיאנט מצביעים לכיוון העלייה המרבית במשטח התמונה.



ציור 5.19 - תמונה ועליה מפת וקטורי הגרדיאנט שחושבו נומרית לפי נגזרת מרכזית

ציור 5.20 מראה תמונה ומציאת שפותיה ע"י חישוב הגרדיאנט והשוואת הנורמה שלו לסף. הסף נבחר כך שאחוז שנקבע מראש של ערכים יעבור אותו. הציור מראה העברה של 10%, 15% ו- 20% מכלל הפיקסלים, וניכר כי השפות בתמונה מתגלות יפה, ובאופן רובסטי לערכו של הסף.

כנקודה אחרונה בסעיף זה נזכיר כי חישוב נגזרות רגיש מאוד לרעש אדיטיבי. אין זה מפתיע כיוון שרעש מתבטא חזק יותר בתדרים הגבוהים, בהם עיקר האינפורמציה היא הרעש ולא תמונת המקור. פעולת גזירה היא פעולת חיזוק התדרים הגבוהים, ולכן אם רעש קיים, הוא יוגבר ויזיק. על מנת להתגבר על בעיה זו מקובל לבצע סינון מקדים על התמונה לשם דיכוי הרעש בה, בטרם חישוב הנגזרות. אם



הסינון המקדים הוא ליניארי (כפי שמקובל) והגזירה ליניארית (וכל הנגזרות שהראינו הינן ליניאריות!), הרי שהמסנן הכולל ליניארי וניתן לביצוע כמסנן אחד. למשל, אם הסינון המקדים הוא 3 על 3 על 3 אחיד, ורצוננו לבצע גזירה אופקית ממורכזת, המסנן הכולל יהיה:

$$\frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \otimes [1 \ 0 \ -1] = \frac{1}{9} \begin{bmatrix} 1 & 1 & 0 & -1 & -1 \\ 1 & 1 & 0 & -1 & -1 \\ 1 & 1 & 0 & -1 & -1 \end{bmatrix}$$

ואנו רואים כאן הפרש אופקי בין ממוצעי אזורים במקום בין פיקסלים בודדים - תוצאה די צפויה.

5.4 פעולות מרחביות מסתגלות

במסגרת הסעיף הקודם הכרנו פעולות סינון ליניאריות ושאינן ליניאריות, אך משותפת לכולן הייתה התכונה שהם קבועים במקום. באופן עקרוני ברור כי עם פעולות שמשתנות במקום ניתן להשיג תוצאות טובות יותר, אך הקביעה כיצד המסנן יהיה תלוי במקום היא שאלה קשה, וללא מענה מוצלח עליה, לא השגנו דבר. במסגרת סעיף זה נראה מספר מצומצם של מסננים תלויי מקום ונציג תוצאות אופייניות להן.

האלגוריתם הראשון אותו נציג קרוי מסנן Wallis. הבעיה בה מנסה מסנן זה לטפל היא תיקון תחום דינאמי וחיידוד התמונה. כבר ראינו שכאשר בתמונה מנוצלים מספר מצומצם של ערכי רמות אפור, פעולת מתיחת קונטרסט או יישור היסטוגרמה מצליחים להשיג שיפור. מה קורה אם בתמונה מנוצלות הרמות כולן, אך באופן שאינו מוצלח, כך שלמשל באזור מסוים ישנם רק גוונים כהים ולא רואים פרטים, ובאזור אחר רק גוונים בהירים ושוב לא רואים פרטים? במצב כזה כל הגוונים מצויים, אך בפרטים קשה להבחין. הפתרון לבעיה זו הינו מתיחה לוקלית של רמות האפור - פעולה לוקלית במקום פעולה גלובלית.

מתיחת היסטוגרמה לוקלית יכולה להיעשות ע"י חישוב בכל פיקסל של היסטוגרמת הסביבה, ותיקון ערך הפיקסל על-פי היסטוגרמה זו, כך שאם רמת האפור של הפיקסל המרכזי מצויה בצד התחתון של ההיסטוגרמה הלוקלית, יש להקטין את רמת האפור עוד, ובכך להשיג חידוד מקומי, ובאופן דומה, אם היא בצד העליון יש להעלותה. הבעיה עם גישה פשטנית זו היא מורכבותה החישובית, שכן בניית היסטוגרמה סביב כל פיקסל היא פעולה כבדה.

כתחליף, מסנן Wallis מציע להניח כי היסטוגרמות לוקליות הן בקירוב גאוסיות, ולכן, במקום לבנותן במלואן, נדרש רק חישוב של שני פרמטרים - ממוצע ושונות. נתחיל את תיאור האלגוריתם בתיאור דרך החישוב של הממוצע והשונות באופן מקומי בצורה נוחה. אם נניח כי התמונה הנתונה היא $x[m,n]$, חישוב הממוצע בכל פיקסל נעשה ע"י קונבולוציה עם מסנן בצורת גאוסין רחב אשר רוחבו σ קובע עד כמה הפעולה לוקלית (ערך מומלץ הוא $\sigma = 10-3$ פיקסלים). מסנן זה נבנה כמטריצה של 15 על 15 (בחירה שרירותית) ע"י המשוואה:

$$G_{\sigma}[m,n] = \frac{\exp\left\{\frac{(m-8)^2 + (n-8)^2}{2\sigma^2}\right\}}{\sum_{k=1}^{15} \sum_{j=1}^{15} \exp\left\{\frac{(k-8)^2 + (j-8)^2}{2\sigma^2}\right\}} \quad 1 \leq n, m \leq 15$$

לכן, תמונת הממוצע תהיה:

$$M_{In}[m,n] = x[m,n] \otimes G_{\sigma}[n,m]$$

קירוב השונות בכל פיקסל נעשית ע"י החסרת תמונת הממוצע מהתמונה המקורית, העלאת התוצאה בריבוע, ושוב קונבולוציה עם אותו מסנן גאוסני, כלומר:

$$S_{In}[m, n] = (x[m, n] - M_{In}[m, n])^2 \otimes G_{\sigma}[n, m]$$

ידוע כי בהינתן משתנה אקראי x בעל ממוצע M ושונות S , הערך $(x-M)/S$ הוא משתנה אקראי חדש בעל ממוצע אפס ושונות 1. באופן דומה, אם נתון משתנה אקראי x בעל ממוצע אפס ושונות 1, ורצוננו להמיר להיות בעל ממוצע M ושונות S , עלינו לבצע את הפעולה $Sx+M$. לכן, אם רצוננו להגיע בכל פיקסל לממוצע יעד M_0 (למשל - 128), ולשונות יעד S_0 (למשל 40), אזי עלינו לבצע את הפעולה הבאה:

$$x_{new}[m, n] = \frac{x[m, n] - M_{In}[m, n]}{S_{In}[m, n]} S_0 + M_0$$

כפי שראויים, תחילה נורמל המשתנה האקראי להיות בעל ממוצע 0 ושונות 1, ומשם הומר לערכי היעד. הפעולה הנ"ל היא פעולה מקומית, אך תלויה (בשל השימוש ב-תמונות M ו- S) בסביבת הפיקסל. מסנן זה אינו ליניארי ואינו חסר זכרון!

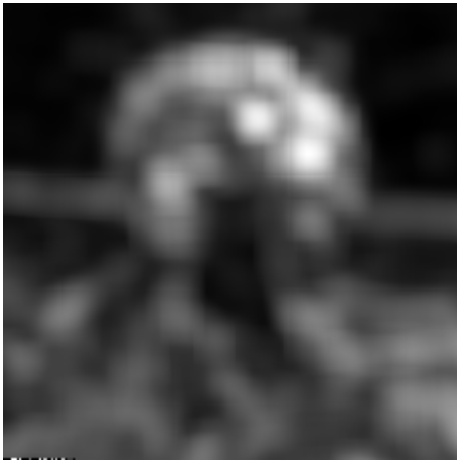
למעשה שיטה זו מניבה תוצאות קיצוניות מידי, כיוון שכפיית ממוצע ושונות יעד (S_0, M_0) זהים בכל מקום היא דרישה חזקה. ריסון לאלגוריתם זה מוצע ע"י קביעת תמונות לממוצע ולשונות היעד, כממוצע משוקלל של הערכים המקוריים וערכי היעד הקבועים. לכן:

$$\begin{aligned} M_{Dest}[m, n] &= (1 - \alpha)M_{In}[m, n] + \alpha M_0 \\ S_{Dest}[m, n] &= (1 - \alpha)S_{In}[m, n] + \alpha S_0 \\ \Rightarrow x_{new}[m, n] &= \frac{x[m, n] - M_{In}[m, n]}{S_{In}[m, n]} S_{Dest}[m, n] + M_{Dest}[m, n] \end{aligned}$$

בגישה זו, שימוש ב- $\alpha = 0$ לא משנה את התמונה, ושימוש ב- $\alpha = 1$ נותן את האלגוריתם הקיצוני. עבור ערכי ביניים נקבל אפקט נאה של חידוד ומתיחת קונטרסט לוקליים. ציור 5.21 מראה תמונת מקור ותוצאות העיבוד עם מסנן Wallis למגוון ערכי α .

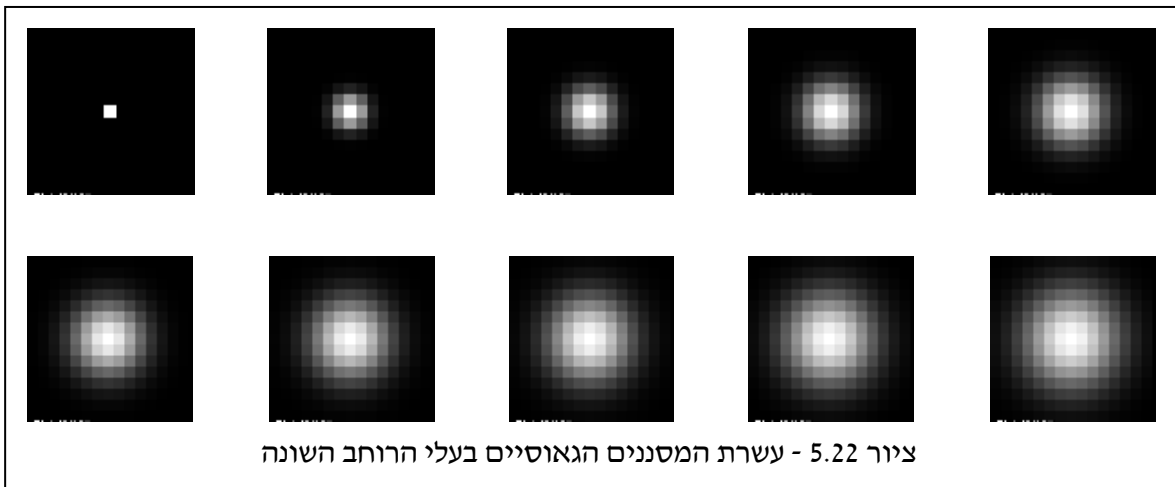
האלגוריתם הבא בו נדון נועד לנקות תמונה מרעש אדיטיבי. כבר ראינו כי החלקה (LPF) מצליחה יפה בניקוי רעש כאשר מדובר באזורים חלקים. לעומת זאת, באזורי שפות, החלקה גורמת לתופעה הצורמת לצופה האנושי. הרעיון אם כך הוא להפעיל מסנן גאוסי בעל שונות תלוית מקום - באזור חלק המסנן יהיה רחב, ובאזור שפה או מרקם המסנן יהיה צר. כך נקבל כי רעש יידחה יפה במקומות החלקים, וכמעט לא יידחה כלל באזורי שפה.

למעשה, במערכת הראיה האנושית קיימת תופעה הקרויה "תופעת המיסוך", אותה אנו מנצלים כאן. מסתבר כי קיומו של הרעש צורם לנו הרבה יותר באזורים חלקים, ושם אנו עושים עבודה טובה של דיכוי. לעומת זאת, באזורי שפות איננו כה רגישים לקיומו של רעש, ולכן ניתן להשאירו.



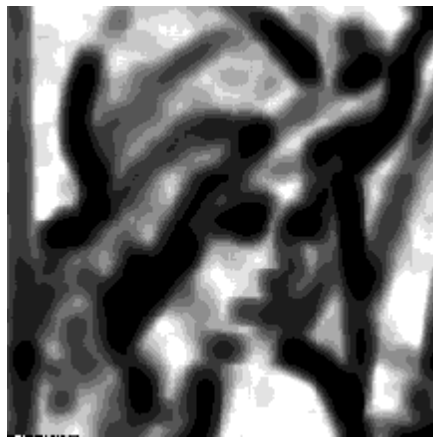
ציור 5.21 - תוצאות סינון Wallis - שורה עליונה (שמאל) תמונת מקור, (ימין) תמונת הממוצע המקומי, שורה אמצעית (שמאל) תמונת השונות המקומית, (ימין) תוצאת הסינון עם $\alpha = 0.1$, שורה תחתונה (שמאל) תוצאת הסינון עם $\alpha = 0.4$, (ימין) תוצאת הסינון עם $\alpha = 1$.

כיצד נקבע אם אזור הוא כולל שפה? אפשרות אחת היא לבצע גזירה כפי שתואר בסעיף קודם, ולהתייחס לערכים המוחלטים של הגרדיאנט. ברור כי גישה זו חייבת לבוא עם החלקה מקדימה של התמונה לשם מניעת השפעה רעה של הרעש המצוי בתמונה. גישה אחרת בא אנו נעשה שימוש היא להשתמש בתמונת שונויות מקומיות, בדיוק כמו זו שנבנתה עבור מסנן ה-Wallis. נדרש לבנות סט של מסננים גאוסיים בעלי רוחב משתנה, ולבחור מסנן לפי ערך השונויות (ביחס הפוך לה - שונויות גבוהה - רוחב נמוך). ציור 5.22 מראה את המסננים הגאוסיים בהם בחרנו לעשות שימוש, וציור 5.23 מראה תוצאות של ניקוי רעש ע"י מסנן אדפטיבי כזה. בציור זה מתוארת גם תמונת בחירת המסננים (לפי אינדקסים בין 1 ל-10, כמספר המסננים הקיימים).

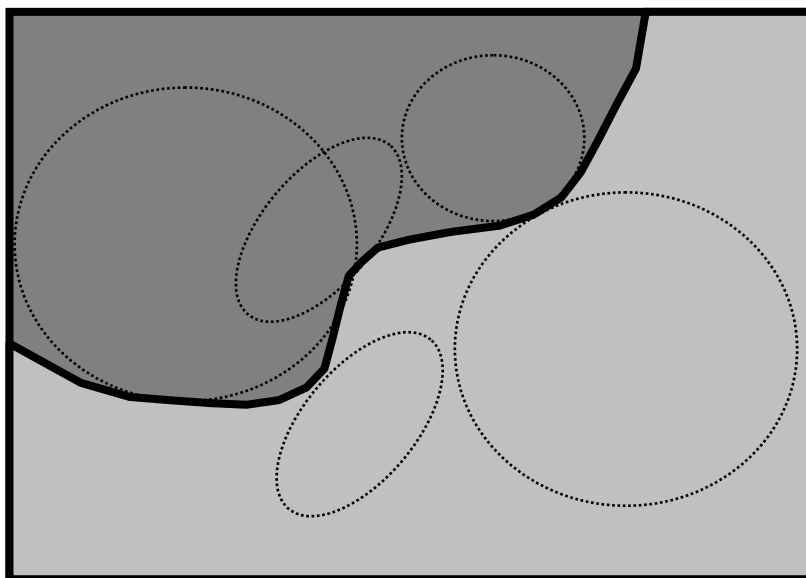


מסתבר כי האלגוריתם הנ"ל אינו הכי טוב שנוכל לעשות, כשאנו מפעילים מסנן תלוי פעילות מקומית. כשאנו באזור שקט אנו פועלים נכון כשאנו מטשטשים ע"י מסנן רחב תמך, וקשה יהיה להתעלות על כך. באזורי שפה, לעומת זאת, אנו בוחרים לא לטשטש, וכתוצאה מכך נותר הרעש כבמקורו. כתחליף לכך, היה ראוי להפעיל באזור כזה מסנן בעל צורת שאינה עיגול אשר היה מכסה רק פיקסלים מהצד הנכון של השפה. ציור 5.24 ממחיש רעיון זה באופן גרפי. תמך המסנן נבחר להיות עיגול אם אין שפה באזור התמך. באשר מתקרבים לשפה, מתעוות המסנן ומנה להשיק לקו השפה, ולא לחצותו. לכן, ממוצע משוקלל המבוצע עם מסנן כזה מערב רק אוכלוסייה מצד אחד של השפה.

לרעיונות אלו של סינון כיווני עומק תאורתי שלא יוצג כאן. מסתבר כי ניתן לבסס תכנון סינון שכזה ע"י כלים מחשבון וריאציות וגיאומטריה דיפרנציאלית, ולהראות כי התוצאה הינה נקודת מינימום של פונקציה על מידת חלקות מסוימת של התמונה. אנו נכיר גישה זו באופן מפורט יותר במסגרת הפרק הבא, הדן בשחזור תמונה.

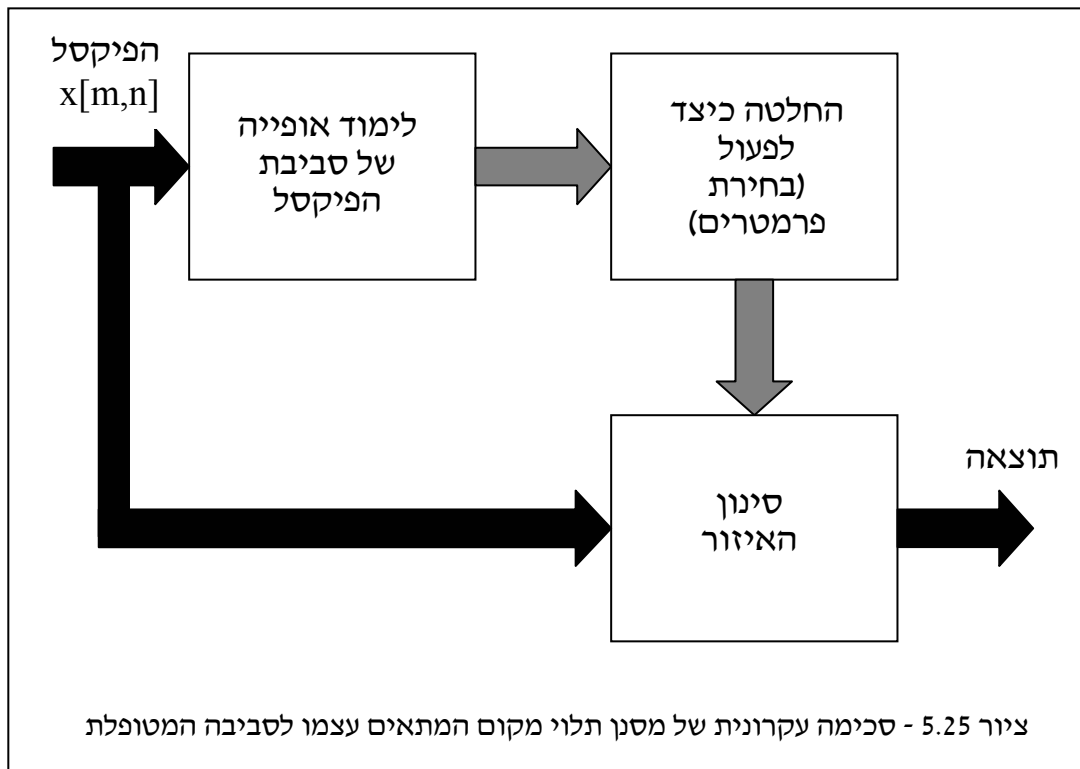


ציור 5.23 - תוצאות סינון אדפטיבי עם גאוסיאן ברוחב משתנה. למעלה (שמאל) תמונת מקור, (ימין) תוצאת אלגוריתם הסינון, למטה (שמאל) תוצאת סינון קבוע במקום ברוחב ממוצע זהה, (ימין) תמונת האינדקסים של המסננים לפי המקום



ציור 5.24 - שימוש במסנן בעל חתך לא איזוטרופי לשם השגת דחיית רעש גם ליד שפות

ניתן להמשיך ולתאר עוד מגוון של מסננים אשר מתאימים עצמם לסביבת הפיקסל לשם קביעת אסטרטגית הטיפול בו, אך אנו נעצור את הדיון כאן. למרבית המסננים מהסוג שתואר כאן מבנה עקרוני כמתואר בציור 5.25, בו נעשית בדיקה של אופי המידע בסביבת הפיקסל, עפ"י מידע זה נלקחת החלטה איזו פעולה לבצע, ולבסוף מבוצעת פעולה תלויית תוכן. חשוב להבהיר שאפילו אם המסנן שמושם בסוף הוא ליניארי (כלומר, צירוף משוקלל של רמות אפור), אין המסנן הכולל ליניארי, בשל שני הבלוקים האחרים בסכימה.



פרק 6 – שחזור תמונה

113	<u>פרק 6 – שחזור תמונה</u>
113	6.1 בעיית השחזור הקלאסית
116	6.2 שיטות שחזור בסיסיות
122	6.3 משערכים סטטיסטיים – ML ו-MAP
125	6.4 שחזור מבוסס משערך ML ו-MAP
132	6.5 שיפור פונקציות ההסתברות לתמונות
140	6.6 בעיות היפוך אחרות והקשר לבעיית השחזור הקלאסית

במסגרת פרק זה נדון בשחזור תמונות מקלוקל מסוים - טשטוש ורעש. הבעיה חשובה ומתעוררת במערכות רבות. אנו נניח כי הן האופרטור המטשטש והן אופי הרעש האדיטיבי ידועים לנו, ונבסס שיטות לבנייתה של התמונה המקורית אשר קדמה לקלוקלים. זוהי בעיה קלאסית בעיבוד אותות בכלל, ובעיבוד תמונות בפרט, אשר זכתה לטיפול עשיר. למרות זאת, כשמסתכלים על התוצאות המתקבלות, נדמה כי עדיין לא נאמרה המילה בתחום.

תיאור נאות של תחום זה מחייב היכרות עם כלים מתורת השערוך (פרק בסטטיסטיקה והסתברות). אנו נציג את הכלים הנחוצים, כך שפרק זה אינו מחייב קריאה נוספת לשלם הבנתו.

1. בעיית השחזור הקלאסית

נתחיל בתיאור הבעיה, ונמחיש אותה דרך דוגמה. בידינו מצלמה דיגיטלית ורצוננו לצלם יעד כלשהו. לו היתה המצלמה שלנו אידיאלית (נטולת הפרעות), היינו מצלמים את התמונה $X[m,n]$ בגודל $L \times L$. לצערנו, בשל בחירתנו לעשות שימוש במצלמה מאיכות ירודה, אנו מקבלים תמונה אחרת $Y[m,n]$ בגודל $L \times L$ (ניתן להכניס לתוך הסיפור גם מצב של יצירת תמונה קטנה יותר). מהם הקלוקלים שהכניסה המצלמה? ובכן, אנו נניח כי המצלמה עושה שימוש באופטיקה מאיכות ירודה (כלומר עדשות ירודות באיכותן) אשר גורמת לטשטוש H . אנו נניח כי טשטוש זה הוא ליניארי וקבוע במקום. לכן, גרעין הטשטוש $H[m,n]$ מספר את סיפורה המלא של פעולה זו.

ישנם מקרים בהם מקור הטשטוש אינה האופטיקה אלא תנועה של המצלמה. כאשר המצלמה זהה בזמן החשיפה, תתקבל מריחה כיוונית הקרויה Motion Blur. מתוך התמונה המקולקלת ניתן לרוב לחשב את גרעין הטשטוש אשר הינו ליניארי וקבוע במקום.

נחזור לסיפור על המצלמה: בנוסף, מטריצת הגלאים בה המצלמה עושה שימוש סובלת מרעשים, ולכן התמונה לאחר הטשטוש מקבלת מרכיב אדיטיבי $N[m,n]$ של רעש אקראי בעל ממוצע אפס. אנו נניח לכל אורכו של פרק זה כי רעש זה הינו רעש לבן והומוגני בעל פילוג גאوسي. פירוש הדבר שבכל פיקסל מצטרף מספר אקראי גאوسي בעל שונות σ_N קבוע במקום, וממוצע אפס.

לכן, הקשר בין התמונה הרצויה $x[m,n]$ לתמונה המתקבלת $Y[m,n]$ נתון ע"י הביטוי:

$$Y[m,n] = X[m,n] \otimes H[m,n] + N[m,n]$$

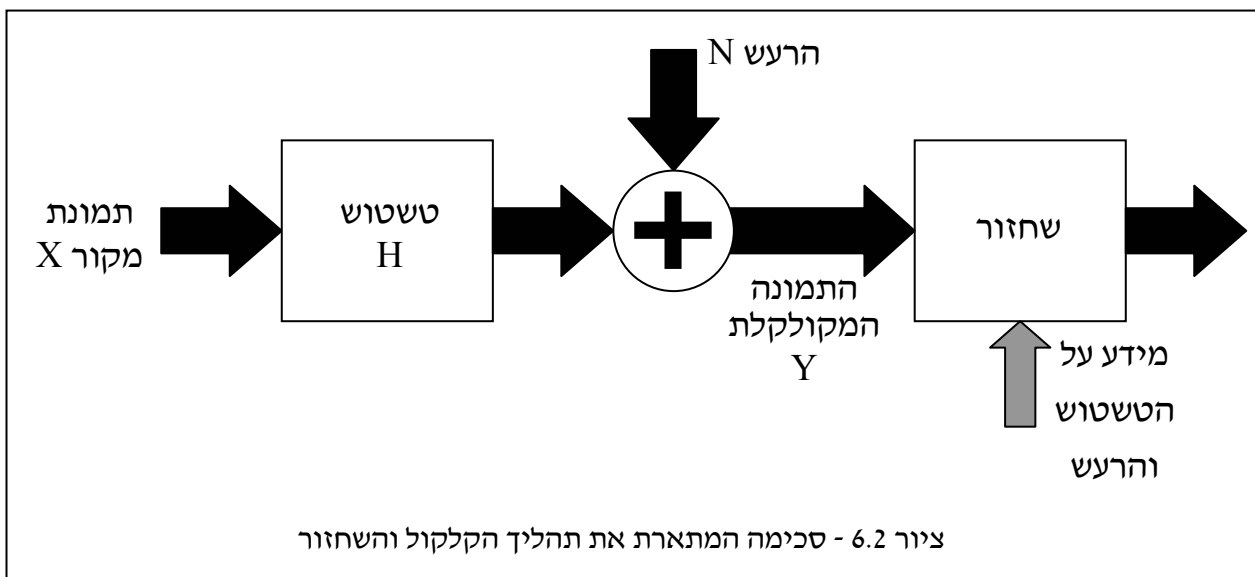
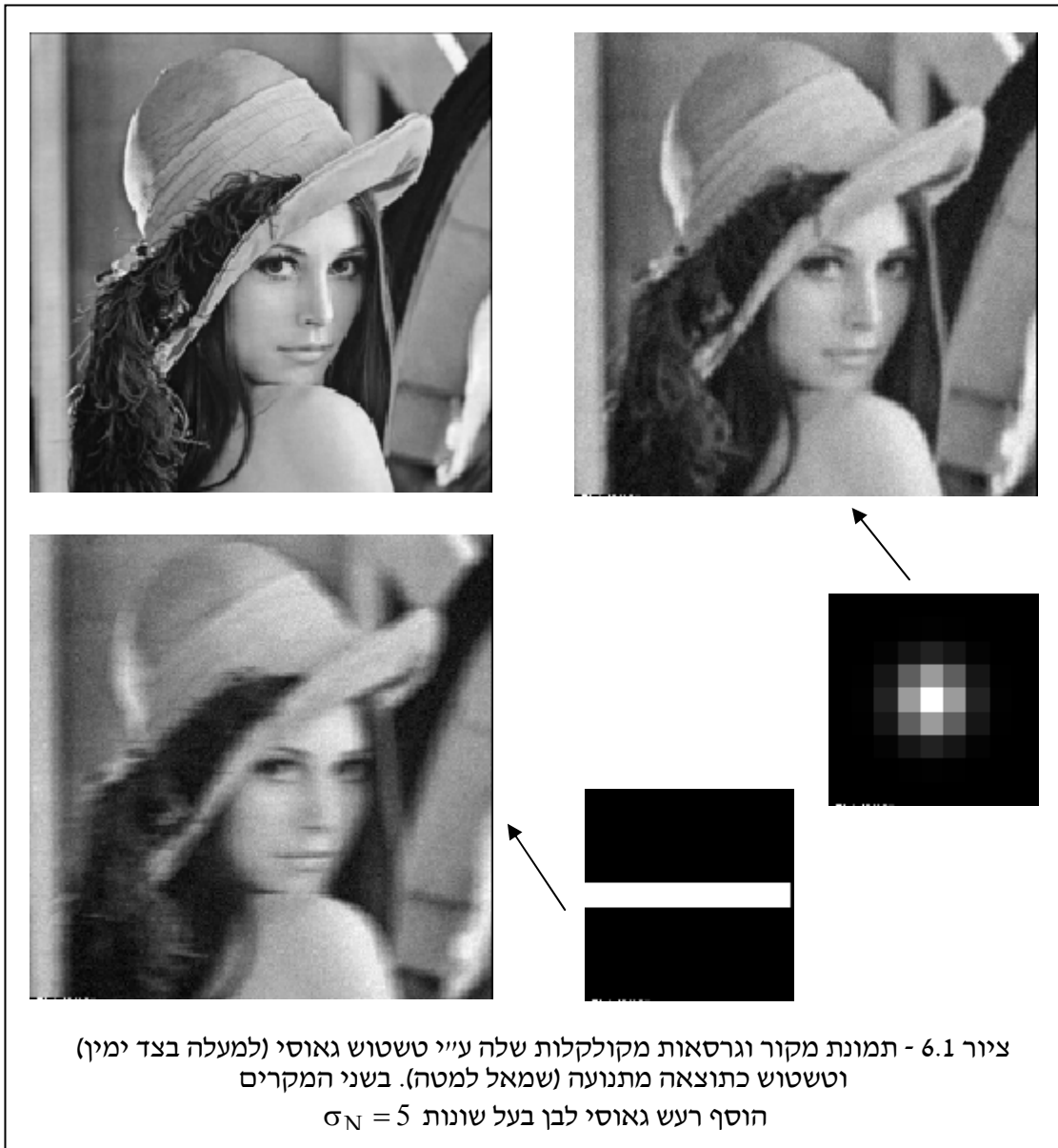
לשם נוחות מקובל לסדר את התמונות בסידור לקסיקוגרפי לפי עמודות (בחירה שרירותית), ואז הקשר הנ"ל הוא:

$$\underline{Y} = \underline{H}\underline{X} + \underline{N}$$

המטריצה \underline{H} היא אופרטור הטשטוש בגודל $[L^2 \times L^2]$. בהמשך טבעי לדיון מפרק 2, נוכל לומר כי כיוון ש- \underline{H} מייצגת אופרטור קבוע במקום, הרי שמטריצה זו בעלת מבנה בלוק טופליץ. אם מטריצה זו גדולה מאוד ביחס לגודל הגרעין המטשטש, נוכל אף להניח בקירוב טוב כי יש לה מבנה בלוק סיבובי. אנו נזדקק להנחות אלו בהמשך.

ציור 6.1 מראה תמונת מקור X וגרסה מקולקלת שלה ע"י טשטוש אופייני לאופטיקה (טשטוש גאوسي בעל חתך רדיאלי), וגרסה של מריחה הנובעת מתנועה. בשני המקרים הוסף רעש אדיטיבי. כאמור, המטרה היא לקחת את התמונה המקולקלת, את המידע על גורמי הקלקול, ולשחזר את תמונת המקור. ציור 6.2 מראה סכימת בלוקים של התהליך במלואו (קלקול ותיקון). המידע המוזן לתהליך השחזור הוא התמונה המקולקלת עצמה, גרעין הטשטוש, ואופיו הסטטיסטי של הרעש (במקרה שלנו רק הפרמטר σ_N).

נשים לב כי לפי המודל הנ"ל יכולנו בקלות לטפל בטשטוש ליניארי תלוי מקום. גם אז אופרטור הטשטוש ניתן לתיאור כמטריצה \underline{H} , אלא שאז אין לה מבנה בלוק טופליץ. הבעיה עם טשטוש תלוי מקום היא הצורך לדעת אותו, ולבנות ממנו את המטריצה \underline{H} . אם בשל הבנה עמוקה של התהליך הפיסיקלי של יצירת התמונה טשטוש זה ידוע, הרי שרוב הדברים שייאמרו בהמשך רלוונטיים גם לטשטוש תלוי מקום.



לעיתים קרובות מקובל להציג את בעיית השחזור בתחום התדר. נציג את המטריצה T_L אשר תהיה פעולת DFT חד-מימדית על וקטור באורך L . אזי המטריצה $T_{L^2} = T_L \otimes T_L$ (מכפלת קרונקר של המטריצה בעצמה) תהיה מטריצה בגודל $[L^2 \times L^2]$ (כמו H) המייצגת פעולת DFT דו-מימדי על תמונה בגודל $L \times L$, כשהיא נתונה בסידור לקסיקוגרפי. לכן, המכפלה $T_{L^2} \underline{X}$ היא בעצם התמרת פוריה דיסקרטית על התמונה $X[m,n]$, אשר תהיה גם היא וקטור באורך L^2 (עם מספרים קומפלקסיים!). ניקח את המשוואה הוקטורית ונכניס לתוכה התמרת פוריה (נשתמש בסימון מקוצר T) באופן הבא:

$$\underline{Y} = \underline{H}\underline{X} + \underline{N} \quad \Rightarrow \quad \underline{TY} = \underline{TH}\underline{X} + \underline{TN} = \underline{THT}^H \underline{TX} + \underline{TN}$$

$$\text{For: } \underline{TY} = \underline{Y}_T, \quad \underline{TX} = \underline{X}_T, \quad \underline{TN} = \underline{N}_T, \quad \underline{THT}^H = \underline{H}_T$$

$$\Rightarrow \underline{Y}_T = \underline{TX} = \underline{H}_T \underline{X}_T + \underline{N}_T$$

כאשר הגדלים עם סימון T פירושים התמרת תדר. עבור H בלוק סיבובית, כבר ראינו בפרק 2 כי התמרת תדר תביא ללכסון, ולכן H_T מטריצה אלכסונית (זו הסיבה שאנו אוהבים כל כך פעולות קבועות במקום וציקליות). איברי האלכסון הראשי הם הערכים העצמיים של המטריצה המקורית H , ואלה הם גם הערכים שהיו מתקבלים בהתמרת DFT דו-מימדית ישירה על גרעין הטשטוש שבנה את H (לאחר שרופד באפסים לגודל המלא של התמונה).

2. שיטות שחזור בסיסיות

אם נתעלם לרגע מהרעש, כיוון שהקלקול הוא טשטוש H , מדוע לא ליישם אופרטור הפוך לו,

H^{-1} , לשם השחזור? אם קיים ההיפוך, הרי שאם ל- H מבנה בלוק סיבובי, אזי גם להיפוך מבנה בלוק סיבובי. זאת קל לראות כיוון שמתקיים:

$$\underline{H}_T = \underline{THT}^H \Rightarrow \underline{H}_T^{-1} = \left[\underline{THT}^H \right]^{-1} = \underline{TH}^{-1} \underline{T}^H$$

כלומר, רואים כי התמרת DFT על האופרטור ההפוך גם היא מטריצה אלכסונית, ולכן האופרטור H^{-1} גם הוא בלוק סיבובי. פירוש הדבר הוא שהמסנן ההפוך אף הוא קבוע במקום, ולכן ניתן להפעלה כקונבולוציה פשוטה.

אבל, מתעוררות שתי בעיות: ראשית, כלל לא ברור כי ל- H קיים היפוך. למעשה, במרבית

המקרים ניתן לקבוע בודאות כי לא קיים היפוך! קל יותר לראות זאת דווקא בייצוג בתדר. אם H לא

הפיכה (כלומר סינגולרית), אזי גם H_T סינגולרית. לכן, באיברי אלכסונה הראשי יש אפסים. מטריצה כזו לא ניתנת להיפוך! מבחינה אינטואיטיבית ניתן להסביר זאת באופן הבא: נניח כי קיים אפס כזה בשורה K כלשהי. אפס זה כפל את האיבר ה- K של התמרת התדר של תמונת הכניסה X . לכן, ערך זה אבד ולא ניתן יהיה לשחזרו.

תיקון לבעיה זו יהיה (ושוב, אנו "חושבים" בתדר) לקחת כל ערך אלכסון ראשי שהוא אפס, ולהמירו בערך אחר קרוב לאפס - נניח $1e-10$. גישה דומה אך מעט שונה היא להשתמש במסנן המשחזר הבא (הנתון בתדר):

$$G_T[k] = \frac{H_T^*[k]}{|H_T[k]|^2 + \text{Const}}$$

כאשר הקבוע קטן משמעותית מ-1 (סדר גודל מומלץ - 0.001). לכן, בתדרים נמוכים בהם $H_T^*[k]$ בערכו המוחלט קרוב ל-1, הקבוע זניח ואנו מקבלים התנהגות של מסנן הפוך. עבור תדרים גבוהים בהם $H_T^*[k]$ מתקרב לאפס, הקבוע משמש כהגנה מחלוקה באפס.

עד כה בחרנו להתעלם מקיומו של הרעש. מסתבר כי אפילו אם הרעש חלש, השפעתו בתהליך של סינון הפוך היא דרמטית. נשתמש שוב בתיאור אינטואיטיבי של הבעיה ע"י ייצוג הבעיה בתדר: אם נניח כי בשורה ה- k של H_T יש ערך נמוך יחסית (נניח $1e-3$). נניח כמו כן כי בנקודת תדר זו, עוצמת האות המקורי ועוצמת הרעש נמוכות מאוד אך הרעש חזק יותר מהאות. תופעה זו מתרחשת כמעט תמיד בתדרים גבוהים בהם כמעט אין נוכחות לאות המקור ואז יוצא שהרעש דומיננטי. נניח לכן - $X_T[k] = 1e-3$ ואילו הרעש חזק יותר - $N_T[k] = 1e-2$. הפעלת המסנן ההפוך פירושה הכפלת שני ערכים אלו בערך $1e3$. גרמנו באופן זה להכנסה חזקה מאוד של רעש לתמונת התוצאה.

המסנן המשחזר שהוצע קודם יתנהג יותר טוב. במקרה כזה, נקבל שהן הרעש והן האות מועברים כשהם מוכפלים בגודל:

$$\frac{1 \times 10^{-3}}{1 \times 10^{-6} + \text{Const}} \stackrel{\cong}{\uparrow} 1$$

Const=0.001

תוצאה זו לא נוראית כמו הכפלה ב-1000, אך אינה טובה לכשעצמה. בנקודת תדר זו היה ראוי שנכפול את הכניסה בערך נמוך מ-1 לשם דיכוי הרעש. ממילא האות טובל שם ברעש גדול ממנו, ולכן אין מה להפסיד. נציע אם כך את הגישה הבאה: בכל נקודת תדר נברר מהו היחס בין עוצמת האות לעוצמת הרעש, ולבצע מסנן הפוך רק בתדרים בהם יחס זה גבוה. במקומות בעלי יחס נמוך, בהם הרעש דומיננטי, כתחליף למסנן הפוך, יש להפעיל דיכוי (כלומר הכפלה בערך נמוך מאוד מ-1. דיכוי זה יקטין את האות שם, אך גם את הרעש שממילא דומיננטי יותר.

הבעיה עם גישה זו היא הצורך בשני גדלים שאת שניהם אין לנו - עוצמת האות המקורי ועוצמת הרעש בכל נקודת תדר. כתחליף, ניתן להניח כי עוצמת הרעש קבועה וזהה בכל נקודת תדר, ושווה σ_N . הנחה זו נכונה כאשר הרעש לבן. באשר לאות המקורי, הנחה סבירה היא שעוצמתו יחסית הפוך למרחק מהראשית בתדר. כלומר, עבור תדרים נמוכים עוצמתו צפויה להיות גבוהה, ועם העלייה בתדר עוצמתו יורדת. ביצוע התמרות פוריה למגוון גדול של תמונות מגלה כי כלל זה נכון ברוב מוחלט של המקרים. ההבדל בין התמונות מתבטא בדרך כלל כמעט רק בפאזה, ומזו אנו מתעלמים פה.

לסיכום, התהליך המוצע הוא ביצוע התמרת פוריה לאופרטור הטשטוש, ולקבל את \mathbf{H}_T . אם

נסמן את ערכי האלכסון הראשי של מטריצה זו ב- $\mathbf{H}_T[k]$, מסנן השחזור בתדר, \mathbf{G}_T יתקבל ע"י:

$$\mathbf{G}_T[k] = \frac{\mathbf{H}_T^*[k]}{|\mathbf{H}_T[k]|^2 + \alpha \sigma_N^2 (r[k])^2}$$

כאשר * מייצג צמוד קומפלקסי, ו- $r[k]$ הוא המרחק אל הראשית במישור התדר. כך למשל, עבור $k=1$ זהו איבר ה- DC ולכן $r[1]=0$. גישה זו דומה להצעתנו הקודמת, למעט העובדה שכעת ה-"קבוע" שלנו גדל עם התדר, ולכן הופך דומיננטי מהר יותר.

מדוע ביטוי זה? בשל היות המכנה גדול מאפס לכל k , ברור כי לא נקבל היפוך של 0, וכך נפתרה הבעיה הראשונה. שנית, עבור k נמוכים, כאשר האות חזק מאוד ביחס לרעש, הגודל $r[k]$ קטן מאוד, ונקבל כי המסנן הוא בקירוב טוב היפוך מדויק כיוון שהאיבר הימני במכנה יהיה זניח. כאשר הרעש דומיננטי מאוד, איבר זה (הימני במכנה) יהיה הדומיננטי במכנה ולכן המסנן המשחזר יהיה קרוב לאפס, כנדרש.

לאחר שהתקבל \mathbf{G}_T (מטריצה אלכסונית) ניתן לבצע התמרת תדר הפוכה לקבלת \mathbf{G} . הפעלת \mathbf{G}

על התמונה \underline{Y} תניב את השחזור. כתחליף, ניתן לבצע את השחזור בתדר ע"י הכפלת \mathbf{G}_T ב- \underline{Y}_T וביצוע התמרת תדר הפוכה על התוצאה - שתי גישות אלו זהות לחלוטין.

המסנן המוצע מוכר מאוד וקרוי מסנן Wiener. בנייתו באופן מדויק וקבלת הביטוי שהוצע נעשית מתוך שיקולים הסתברותיים, שלא הצגנו כאן. ציור 6.3 מראה תמונת מקור מטושטשת ורועשת, ושחזור שלה ע"י מסנן היפוך. בעוד שהמרחק (במובן MSE) בין תמונת המקור לתמונה המקולקלת הוא 103, הרי שהמרחק בין תמונת המקור לשחזור כאן הוא $7e6$! התוצאה ללכ כל ספק גרועה!

בציור 6.4 אנו מראים לאותה תמונה בדיוק את תוצאת מסנן Wiener, עבור שלושה ערכי α שונים. עבור $\alpha = 5e-6$ מתקבלת שגיאה של 87.1, עבור $\alpha = 1e-6$ השגיאה היא 78.8, ועבור $\alpha = 5e-7$ השגיאה עולה ל-94.8. אנו רואים כי חל שיפור בתהליך השחזור, הן באופן ויזואלי, והן באופן

כמותי. עם זאת, קשה להתנתק מתחושה של החמצה כשמסתכלים בתוצאות - השיזור אינו מתקרב כלל לאיכות של תמונת המקור, ונשאלת השאלה האם זה הכי טוב שניתן להשיג? למזלנו, התשובה על כך היא לא! ניתן להציע פתרונות איכותיים יותר, אך הדבר מחייב שיטות מורכבות יותר, אשר יידונו בסעיף הבא.



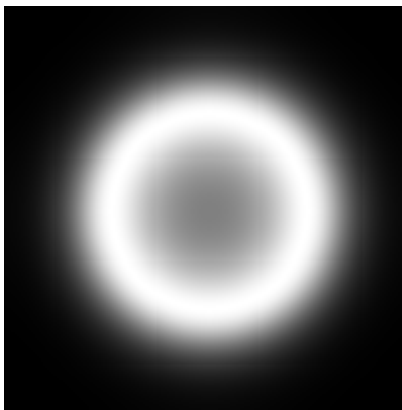

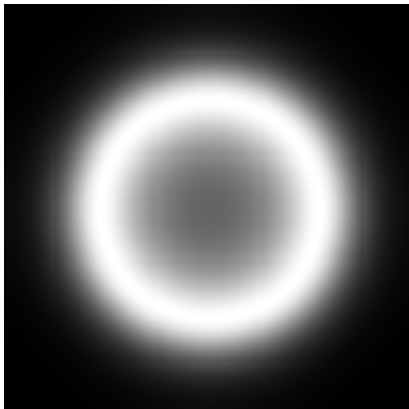



דוגמה 6.1: נניח כי אנו עוסקים באותות חד מימדיים, הטשטוש הוא הפונקציה הבאה (בתדר):

$$H_T[k] = \frac{10}{k^2 + 10} \quad k = 0, 1, \dots, 100.$$

המסנן המשחזר שהוצע ראשון בסעיף זה יהיה:

$$G_T[k] = \frac{\frac{10}{k^2 + 10}}{\left[\frac{10}{k^2 + 10} \right]^2 + 0.001} = \frac{10(k^2 + 10)}{100 + 0.001(k^2 + 10)^2}$$

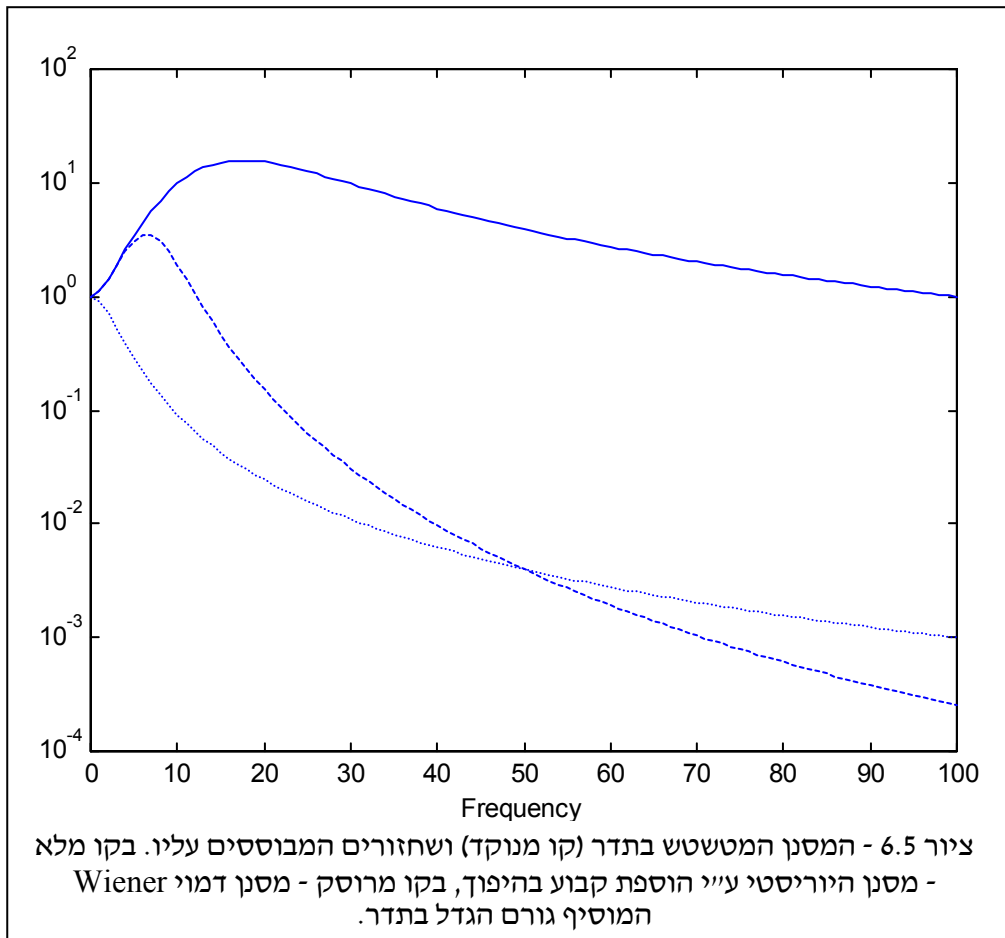
	
	
	

צור 6.4 - שחזור עיני מסנן Wiener. שמאל - המסננים (בערך מוחלט), ימין - תוצאות השחזורים. שלושת התוצאות מתייחסות לשלושה ערכים שונים של הפרמטר α

ואילו המסנן המשחזר השני יהיה (נבחר כאן קבוע באופן שרירותי):

$$G_T[k] = \frac{\frac{10}{k^2 + 10}}{\left[\frac{10}{k^2 + 10}\right]^2 + 0.001[0.4k^2]} = \frac{10(k^2 + 10)}{100 + 0.0004(k^2 + 10)^2 k^2}$$

ציור 6.5 מראה את שלושת מסננים אלו (המסנן המטשטש, ושני המשחזרים).



אנו רואים כי המסנן הבנוי על הוספת קבוע נותן אמנם דעיכה בהגבר אחרי תדר מסוים, אך בכל מקרה הוא גבוה. מסנן ה-Wiener מצליח לרדת חזק יותר ואף לתת ניחות חזק יותר משל המסנן המטשטש - תוצאה רצויה כיוון שאנו מניחים שבאזור זה חזק הרעש מהאות. בשני המקרים, ההתנהגות באיזור התדרים הנמוכים דומה - התנהגות של מסנן הפוך אשר אמורה לשחזר יפה מרכיבי תדר אלה.

6.3 משערכים סטטיסטיים – ML ו- MAP

בסעיף זה נציג באופן תיאורטי שני משערכים סטטיסטיים, בהם נעשה שימוש לטובת בניית אלגוריתמי שחזור תמונה טובים יותר. משערכים אלו הם משערכך הסבירות המירבית (ML - Maximum Likelihood), ומשערכך ההסתברות הסופית המקסימלית (Maximum A-posteriori) (MAP - Probability). בשני המקרים הסיפור מתחיל די דומה - נתון לנו סט נתונים שיסומן Y , ורצוננו לשערכך ערך לא ידוע X . ידוע לנו קשר בין X ל- Y באופן כלשהו (במונחים של פונקציות הסתברות), ורצוננו להציע דרך "חכמה" לחישוב של הנעלם X . נמחיש את הסיפור הנ"ל דרך דוגמה פשוטה.

דוגמה 6.1: נניח נתונה לנו סידרה בת 100 מספרים y_k אשר ידוע כי הוגרלו עפ"י חוק פילוג גאוסית חסר זיכרון עם ממוצע לא ידוע ושונות 1. הוקטור Y יהיה סט מספרים אלו. הנעלם יהיה M - הממוצע. ברמה אינטואיטיבית ידוע לנו כי שערך ממוצע במצב כזה יכול להיעשות פשוט ע"י ממוצע המדידות, דהיינו:

$$\hat{M} = \frac{1}{100} \sum_{k=1}^{100} y_k$$

משערכך הסבירות המרבית דורש מאתנו לבנות את פונקצית ההסתברות $P(Y/X)$, כלומר פונקציה שתקבע מהי הסבירות שהמדידות תהינה Y עבור ערך הנעלם X . פונקציה זו היא פונקציה של המשתנה X . אם הצלחנו לנסח פונקציה כזו, המשערכך מציע כי נמצא את ערכו של X המביא את הסתברות זו למקסימום:

$$\hat{X}_{ML} = \underset{X}{\text{ArgMax}} P(Y/X)$$

נחזור לדוגמה הקודמת ונראה כיצד פונקציה זו מתקבלת.

דוגמה 6.1 (המשך): חוק ההסתברות של הוקטור Y הוא פשוט מכפלה של 100 הסתברויות לכל אחד מאיבריו, כאשר כל איבר כזה הוא הסתברות גאוסית. אם נניח כי אנו יודעים שהממוצע הוא M נקבל:

$$\begin{aligned} P(Y/M) &= \prod_{k=1}^{100} P(y_k/M) = \prod_{k=1}^{100} C \cdot \exp\left\{-\frac{(y_k - M)^2}{2}\right\} = \\ &= C^{100} \cdot \exp\left\{-\frac{1}{2} \sum_{k=1}^{100} (y_k - M)^2\right\} = L(M) \end{aligned}$$

הפונקציה שהתקבלה נקראת פונקצית הסבירות L , והיא מעל המשתנה M - הממוצע הלא נודע.

כזכור, רצוננו למצוא את M שיביא פונקציה זו למקסימום, ולכן:

$$\text{ArgMax}_M \left\{ C^{100} \cdot \exp \left\{ -\frac{1}{2} \sum_{k=1}^{100} (y_k - M)^2 \right\} \right\} = \text{ArgMin}_M \left\{ \sum_{k=1}^{100} (y_k - M)^2 \right\} = \frac{1}{100} \sum_{k=1}^{100} y_k$$

ואנו רואים כי ה- ML נתן את התוצאה שהצענו קודם מתוך אינטואיציה.

דוגמה 6.2: נניח מקרה דומה, אלא שהפעם סט המדידות נובע מתוך התפלגות אקפוננציאלית מהצורה:

$$P(y_k / M) = C \cdot \exp\{-\lambda|y_k - M|\}$$

מהו משעריך ה- ML לממוצע M במקרה זה? הפתרון בדומה לקודם יהיה לכתוב את ההסתברות המותנית:

$$\begin{aligned} P(Y/M) &= \prod_{k=1}^{100} P(y_k / M) = \prod_{k=1}^{100} C \cdot \exp\{-\lambda|y_k - M|\} = \\ &= C^{100} \cdot \exp\left\{-\lambda \sum_{k=1}^{100} |y_k - M|\right\} = L(M) \end{aligned}$$

מינימיזציה הפעם תניב:

$$\text{ArgMax}_M \left\{ C^{100} \cdot \exp \left\{ -\lambda \sum_{k=1}^{100} |y_k - M| \right\} \right\} = \text{ArgMin}_M \left\{ \sum_{k=1}^{100} |y_k - M| \right\} = \text{MED}\{y_1, y_2, \dots, y_{100}\}$$

התוצאה הפעם פחות צפויה. אנו רואים כי חישובו של ממוצע הפילוג תלוי למעשה בחוק הפילוג עצמו. יכולנו לשערך את הממוצע ע"י מיצוע אלגברי, אך במקרה זה שערך זה צפוי להיות פחות טוב (לפחות במובן ML).

עד כה דיברנו על משעריך ה- ML. לפי גישה זו הנחנו כי הנעלם הוא פרמטר בחוק הפילוג ולא משתנה אקראי בעצמו. מה אם ידוע לנו כי הנעלם X הוא אכן משתנה אקראי, וידוע לנו חוק הפילוג שלו? למשל, עבור הדוגמה הקודמת, מה אם ידוע לנו כי הממוצע M הוא משתנה אקראי גאוסי בעצמו עם ממוצע 5 ושונות 2. כיצד נרתום מידע זה על מנת לשערך טוב יותר את הנעלם?

התשובה היא משעריך ה- MAP, המציע את השימוש בחוק Bayes. משעריך ה- MAP מציע להביא למקסימום את ההסתברות $P(X/Y)$, שהינה ההסתברות לקבלת הערך X כשנתונות המדידות Y (בדיוק הפוך ל- ML). נוסחת Bayes קובעת כי:

$$P(X/Y) = \frac{P(X, Y)}{P(Y)} = \frac{P(X)P(Y/X)}{P(Y)}$$

ולכן משערך ה-MAP ינסה להשיג מקסימום להסתברות זו ביחס לערכו של X . מקסימיזציה ביחס ל- X פירושה שהמכנה $P(Y)$ אינו משפיע כלל, ולכן:

$$\hat{X}_{MAP} = \underset{X}{\text{ArgMax}} P(X/Y) = \underset{X}{\text{ArgMax}} \{P(X)P(Y/X)\}$$

ישנו הרבה דמיון לשערוך ה-ML. ההבדל הוא בכך שפונקציית הסבירות $P(Y/X)$ מוכפלת בפונקציית ההסתברות של X הקרויה לעיתים Prior (כקיצור ל- Prior-Knowledge - מידע מקדים). מעניין לציין שכאשר אין לנו מידע מקדים על X , אנו נניח כי הפילוג שלו מתפזר אחיד על פני כל הערכים האפשריים, ואז נקבל כי משערך ה-MAP הופך להיות משערך ה-ML.

דוגמה 6.3: בהמשך לדוגמה 6.1, נניח כי ידוע כי הממוצע הלא נודע מפולג גאוסית לפי:

$$P(M) = C_0 \exp\left\{-\frac{(M-5)^2}{2\sigma_M^2}\right\}$$

במקרה זה, שערוך ה-ML לא משתנה. לעומת זאת שערוך ה-MAP יהיה:

$$\begin{aligned} \underset{M}{\text{ArgMax}} \left\{ C^{100} \cdot \exp\left\{-\frac{1}{2} \sum_{k=1}^{100} (y_k - M)^2\right\} \cdot C_0 \cdot \exp\left\{-\frac{(M-5)^2}{2\sigma_M^2}\right\} \right\} &= \\ = \underset{M}{\text{ArgMin}} \left\{ \frac{1}{2} \sum_{k=1}^{100} (y_k - M)^2 + \frac{(M-5)^2}{2\sigma_M^2} \right\} &= \frac{5/\sigma_M^2 + \sum_{k=1}^{100} y_k}{100 + 1/\sigma_M^2} \end{aligned}$$

עבור $\sigma_M \rightarrow 0$ נקבל כי התוצאה שואפת ל-5. זה צפוי כי כאשר $\sigma_M \rightarrow 0$ פירוש הדבר שבודאות מוחלטת אנו יודעים כי הממוצע שווה 5 (עוד בטרם התקבלו מדידות). עבור $\sigma_M \rightarrow \infty$ נקבל כי הממוצע האלגברי הוא השערוך, כמו שהציע ה-ML. גם זו תוצאה צפויה כיוון שעבור שונות גדולה מאוד, פירוש הדבר הוא שאין לנו בעצם מידע מקדים על הנעלם, ולכן יש לדלותו מהמדידות בלבד.

6.4 שחזור תמונות מבוסס ML ו- MAP

נחזור כעת לבעיית השחזור שלנו, כשאנו מנסים להשתמש במשעך ה- ML. וקטור המדידות הוא

\underline{Y} , ווקטור הנעלמים שלנו הוא \underline{X} . הקשר ביניהם נתון ע"י $\underline{Y} = \mathbf{H}\underline{X} + \underline{N}$, כאשר \mathbf{H} מטריצת טשטוש ו- \underline{N} וקטור רעש אקראי גאוסי בעל חוק פילוג:

$$P_N(\underline{N}) = C \cdot \exp\left\{-\frac{\underline{N}^T \underline{N}}{2\sigma_N^2}\right\}$$

עפ"י ביטוי זה, ההסתברות $P(\underline{Y}/\underline{X})$ אינה אלא:

$$P_N(\underline{Y}/\underline{X}) = C \cdot \exp\left\{-\frac{[\underline{Y} - \mathbf{H}\underline{X}]^T [\underline{Y} - \mathbf{H}\underline{X}]}{2\sigma_N^2}\right\}$$

זאת כיוון שאם מתייחסים ל- \underline{X} כוקטור פרמטרים נתון, אזי הגורם $\mathbf{H}\underline{X}$ מהווה תוספת דטרמיניסטית, והוקטור החדש $\underline{Y} = \mathbf{H}\underline{X} + \underline{N}$ הינו גאוסי גם כן, בעל אותה שונות, ועם ממוצע חדש שהינו הגודל שהוסף. משעך ה- ML יהיה הוקטור \underline{X} אשר יביא את ההסתברות הנ"ל למקסימום. לכן:

$$\begin{aligned} \hat{\underline{X}}_{ML} &= \underset{\underline{X}}{\text{ArgMax}}\{P_N(\underline{Y}/\underline{X})\} = \underset{\underline{X}}{\text{ArgMax}}\left\{\exp\left\{-\frac{[\underline{Y} - \mathbf{H}\underline{X}]^T [\underline{Y} - \mathbf{H}\underline{X}]}{2\sigma_N^2}\right\}\right\} = \\ &= \underset{\underline{X}}{\text{ArgMin}}\left\{\frac{[\underline{Y} - \mathbf{H}\underline{X}]^T [\underline{Y} - \mathbf{H}\underline{X}]}{2\sigma_N^2}\right\} = \underset{\underline{X}}{\text{ArgMin}}\{[\underline{Y} - \mathbf{H}\underline{X}]^T [\underline{Y} - \mathbf{H}\underline{X}]\} = \\ &= [\mathbf{H}^T \mathbf{H}]^{-1} \mathbf{H}^T \underline{Y} \end{aligned}$$

התוצאה שהתקבלה נכונה רק אם קיימת מטריצה הפוכה ל- $\mathbf{H}^T \mathbf{H}$, שפירושה שעל \mathbf{H} עצמה להיות לא סינגולרית. למעשה, משעך ה- ML מציע לנו להביא למינימום את הפונקציה $f(\underline{X}) = [\mathbf{H}\underline{X} - \underline{Y}]^T [\mathbf{H}\underline{X} - \underline{Y}]$ - הצעה בהחלט אינטואיטיבית שפירושה, מצא את התמונה שאם תטושטש ע"י \mathbf{H} היא תהיה קרובה ככל האפשר ל- \underline{Y} .

התוצאה שהתקבלה אינה אלא ביצוע של מסנן הפוך כיוון שאם \mathbf{H} ריבועית מתקבל:

$$\hat{\underline{X}}_{ML} = [\mathbf{H}^T \mathbf{H}]^{-1} \mathbf{H}^T \underline{Y} = \mathbf{H}^{-1} \mathbf{H}^{-T} \mathbf{H}^T \underline{Y} = \mathbf{H}^{-1} \underline{Y}$$

כאשר H אינה ריבועית (בניגוד למקרה שלנו), התוצאה הנ"ל הינה היפוך מוכלל (Pseudo-Inverse) המוכר לנו מאלגברה ליניארית. תוצאה מעניינת שהתקבלה כאן היא שלעוצמת הרעש אין משקל בקביעת המסנן המשחזר.

בכל מקרה, ההנחה שקיים היפוך ל- H היא הנחה מגבילה. על מנת להיפטר מהצורך בהנחה זו, מקובל לעשות שימוש ברגולריזציה. הרעיון די פשוט - המטריצה $H^T H$ חיובית חצי מוגדרת - עובדה שפירושה שכל ערכיה העצמיים ממשיים ואי-שליליים. לוי היו כל הערכים העצמיים מעל האפס ממש, הרי ש- $H^T H$ הייתה הפיכה. לכן, נוסף ל- $H^T H$ מטריצת יחידה (I) מוכפלת בקבוע חיובי קטן מאוד ε^2 . תוספת זו תזיז את כל הערכים העצמיים למעלה, ותביא למצב בו $H^T H + \varepsilon^2 I$ הפיכה. לכן, השחזור יהיה:

$$\hat{X}_{ML} = [H^T H + \varepsilon^2 I]^{-1} H^T Y$$

ההנחה היא שאם ε^2 קטן דיו לא להשפיע על ההתנהגות הכללית, ועם זאת, בעל ערך גבוה דיו למניעת בעיות יציבות נומריות, נקבל שחזור לא רע. בעקרון, יכולנו להוסיף למטריצה $H^T H$ כל מטריצה חיובית מוגדרת M ולהשיג את אותו אפקט, אך בעצם איננו יודעים איזו מטריצה לבחור. גילינו, אם כך, כי משיקולי ML יוצא שיש להפעיל מסנן הפוך לשם שחזור, וניתן לייצבו ע"י בחירה שרירותית של מטריצת רגולריזציה.

נשתמש כעת במשעך ה- MAP. לצורך השימוש בו עלינו לעשות שימוש בפונקציה הסתברות $P(\underline{X})$. פונקציה הסתברות זו פירושה שבהינתן תמונה כלשהי \underline{X} , נוכל לומר מהי מידת סבירותה. מה עושה תמונה לסבירה יותר? שאלה קשה אשר עד היום אין תשובה מוצלחת עליה. אבל, ישנן בהחלט תשובות חלקיות בהן נוכל לעשות שימוש.

הצעה מאוד פופולרית לאומדן הסתברות של תמונה היא ההצעה לעשות שימוש במידת חלקותה של התמונה. ניתן לטעון כי תמונה חלקה (כמשטח בתלת-מימד) יותר סבירה מתמונה שאינה חלקה. ברור כי באזורים חלקים בתמונה הנחה זו נכונה, ובאזורים שפות הנחה זו אינה נכונה. נתעלם בשלב זה מפגם זה (מאוחר יותר נראה דרכים להתגבר על סטייה זו). את חלקותה של התמונה נוכל למדוד ע"י הפעלת נגזרת כלשהי (גרדיאנט, לפלסיאן ...) וחשוב סכום ריבועי ערכי התוצאה. אם התמונה חלקה, הרי שתוצאת הנגזרת תהיה נמוכה בכל מקום ולכן הערך הכולל יהיה נמוך. אם בתמונה יש הרבה שפות נקבל כי הנגזרות גבוהות לכן הערך הכולל יהיה גבוה. אם נניח כי המטריצה D מייצגת פעולת נגזרת מסדר שנבחר, פונקציה הסתברות המוצעת היא:

$$P(\underline{X}) = C \cdot \exp \left\{ - \frac{[\underline{DX}]^T [\underline{DX}]}{2\sigma_D^2} \right\}$$

הביטוי \mathbf{DX} מקבל את התמונה X ומבצע עליה פעולת גזירה, והביטוי $[\mathbf{DX}]^T [\mathbf{DX}]$ לוקח את תוצאת הגזירה מעלה בריבוע כל איבר ומסכם. לכן, ערך זה יהיה נמוך לתמונות חלקות, וגבוה לתמונות שאינן חלקות. השימוש במבנה של פונקציית אקספוננט נובע מנוחות שתתבהר בהמשך. הפרמטר σ_D קובע את שונותה של הסתברות זו. בכל אופן, רואים שעבור תמונה חלה ההסתברות גבוהה, ועבור תמונה עם ריבוי שפות, ההסתברות נמוכה, כנדרש.

משערך ה- MAP מביא את מכפלת פונקציית הסבירות (הלוקחה מה- ML) ופונקציית ה- Prior הנ"ל למקסימום, ולכן:

$$\begin{aligned} \hat{X}_{MAP} &= \underset{X}{\text{ArgMax}} \{P_N(\underline{Y}/X)P_X(X)\} = \\ &= \underset{X}{\text{ArgMax}} \left\{ \exp \left\{ -\frac{[\underline{Y} - \mathbf{HX}]^T [\underline{Y} - \mathbf{HX}]}{2\sigma_N^2} \right\} \cdot \exp \left\{ -\frac{[\mathbf{DX}]^T [\mathbf{DX}]}{2\sigma_D^2} \right\} \right\} = \\ &= \underset{X}{\text{ArgMin}} \left\{ \frac{[\underline{Y} - \mathbf{HX}]^T [\underline{Y} - \mathbf{HX}]}{2\sigma_N^2} + \frac{[\mathbf{DX}]^T [\mathbf{DX}]}{2\sigma_D^2} \right\} = \\ &= \underset{X}{\text{ArgMin}} \left\{ [\underline{Y} - \mathbf{HX}]^T [\underline{Y} - \mathbf{HX}] + \frac{\sigma_N^2}{\sigma_D^2} [\mathbf{DX}]^T [\mathbf{DX}] \right\} = \\ &= \left[\mathbf{H}^T \mathbf{H} + \frac{\sigma_N^2}{\sigma_D^2} \mathbf{D}^T \mathbf{D} \right]^{-1} \mathbf{H}^T \underline{Y} \end{aligned}$$

אנו רואים כי הפעם, המסנן התקבל עם רגולריזציה שנבעה מתוך שיקולי ה- MAP. כלומר, אם רצונו בשחזור מוצלח, מוטב שנבחר לעשות שימוש מטריצת גזירה כלשהי. כמו כן, אנו רואים כי מנת שתי השונויות פועלת במשקל לגורם הרגולריזציה הנוסף.

בהשוואה ל- ML, משערך ה- MAP מציע מינימיזציה של הפונקציה

$$f(X) = [\underline{Y} - \mathbf{HX}]^T [\underline{Y} - \mathbf{HX}] + \frac{\sigma_N^2}{\sigma_D^2} [\mathbf{DX}]^T [\mathbf{DX}]$$

בפונקציה זו המרכיב הראשון זהה לזה של ה- ML, והמרכיב השני מושך את הפתרון לעבר תמונות חלקות יותר. אם תמונה מסוימת מצליחה להביא לערך מאוד נמוך את הגורם הראשון אין בכך די לשם קבלתה כפתרון - המחיר השני גם הוא משחק תפקיד בקביעת הפתרון. כאשר σ_D נמוך מאוד פירוש הדבר שאנו דורשים שהתוצאה תהיה חלקה מאוד, ואמנם בפונקציה f הגורם השני נהיה דומיננטי, ולכן הפתרון יהיה חלק. עבור $\sigma_D \rightarrow 0$ הבעיה דורשת חלקות מוחלטת, ולכן התמונה תהיה קבוע עבור גזירה

ראשונה ומישור מוטה עבור גזירה שניה. כאשר $\sigma_D \rightarrow \infty$ אנו מחלישים את דרישת החלקות ואז יתנקז הפתרון לתוצאת ה-ML.

כנקודה אחרונה בסעיף זה נתייחס לפתרון מעשי של בעיית השחזור ע"י הכלים שהוצגו לעיל. ניקח את תוצאת ה- MAP כיוון שהיא מכלילה את ה- ML, ונראה כיצד מעשית משיגים שחזור באמצעותה. נניח כי לשם רגולריזציה אנו משתמשים במסנן 3 על 3 לפלסיאן. אם נניח כי התמונה בגודל 200×200 , הרי שהמטריצות המעורבות (\mathbf{D}, \mathbf{H}) תהיינה בגודל של $40,000 \times 40,000$. מטריצות אלו דלילות (מרבית איבריהן אפסים), אך היפוך מטריצה בגודל כזה אינו בא בחשבון. שתי אלטרנטיבות קיימות הן פתרון במישור התדר, ושיטות איטרטיביות.

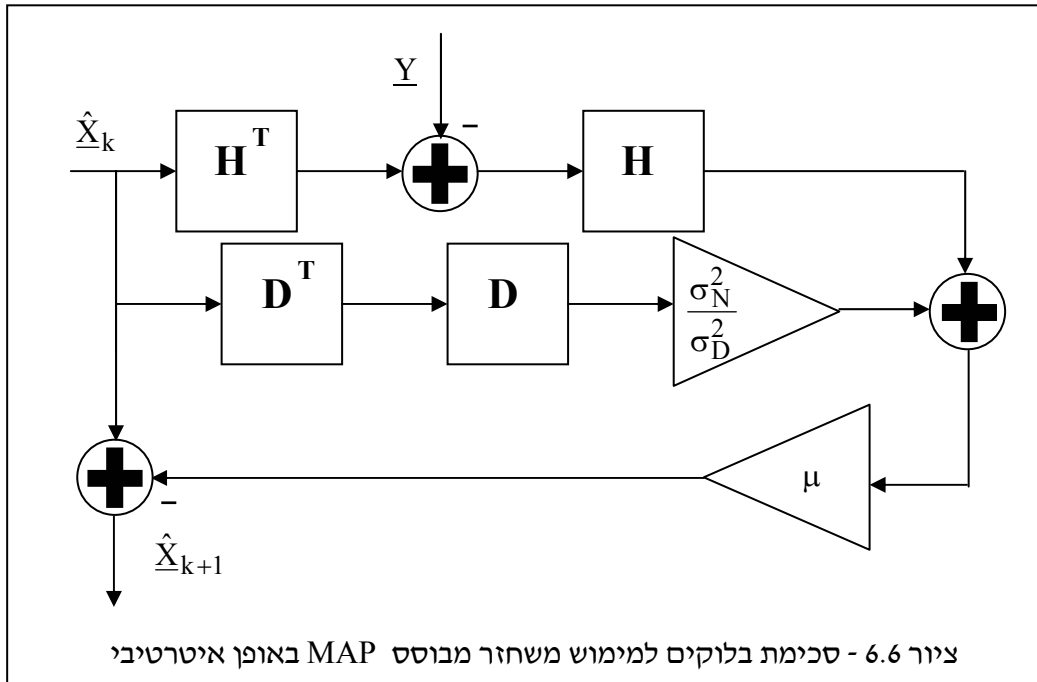
באשר לפתרון איטרטיבי, הרעיון הבסיסי ביותר הוא אלגוריתם השיפוע המירבי (Steepest Descent). הרעיון הוא להתחיל מניחוש מוצלח לפתרון (למשל $\hat{\mathbf{X}}_0 = \mathbf{Y}$) ולשפרו שוב ושוב ע"י ירידה בכיוון הפוך לכיוון הגרדיאנט של הפונקציה $f(\mathbf{X})$. הגרדיאנט מצביע על הכיוון בו חלה העלייה הגדולה ביותר, ולכן הליכה בכיוון הפוך פירושה ירידה מירבית. לכן, משוואת האיטרציה תהיה:

$$\hat{\mathbf{X}}_{k+1} = \hat{\mathbf{X}}_k - \mu \left. \frac{\partial f(\mathbf{X})}{\partial \mathbf{X}} \right|_{\hat{\mathbf{X}}_k} = \hat{\mathbf{X}}_k - \mu \left[\mathbf{H}^T (\mathbf{H} \hat{\mathbf{X}}_k - \mathbf{Y}) + \frac{\sigma_N^2}{\sigma_D^2} \mathbf{D}^T \mathbf{D} \hat{\mathbf{X}}_k \right]$$

הגודל μ הוא גודל הצעד לעדכון - ערך גדול מידי ייתן התבדרות, ערך קטן מידי ייתן התכנסות איטית מידי, ויש לבחור אותו בחוכמה (קיימות דרכים לקביעתו באופן אוטומטי, שלא יידונו כאן). התהליך המתואר מובטח להתכנס לפתרון במדויק של הבעיה בשל היותה של הפונקציה f קמורה, ובשל תכונות האלגוריתם המוצע.

כיוון ש- \mathbf{H} ו- \mathbf{D} מסננים ליניאריים וקבועים במקום, הרי שגם לאחר שחלוף פעולות אלו הינן ליניאריות וקבועות במקום. יתרה מזו, כאשר הטשטוש בעלת גרעין סימטרי, מתקבל כי $\mathbf{H}^T = \mathbf{H}$ (גם הלפלסיאן מקיימת תכונה זו). לכן, אין כלל צורך לעבור עם תמונות בסידור לקסיקוגרפי או מטריצות כה גדולות. ניתן פשוט להתייחס למשוואה הנ"ל כסדרת צעדי סינון שעל הפתרון הקודם לעבור לשם עדכונו. ציור 6.6 מראה את סכימת הבלוקים של תהליך העדכון.

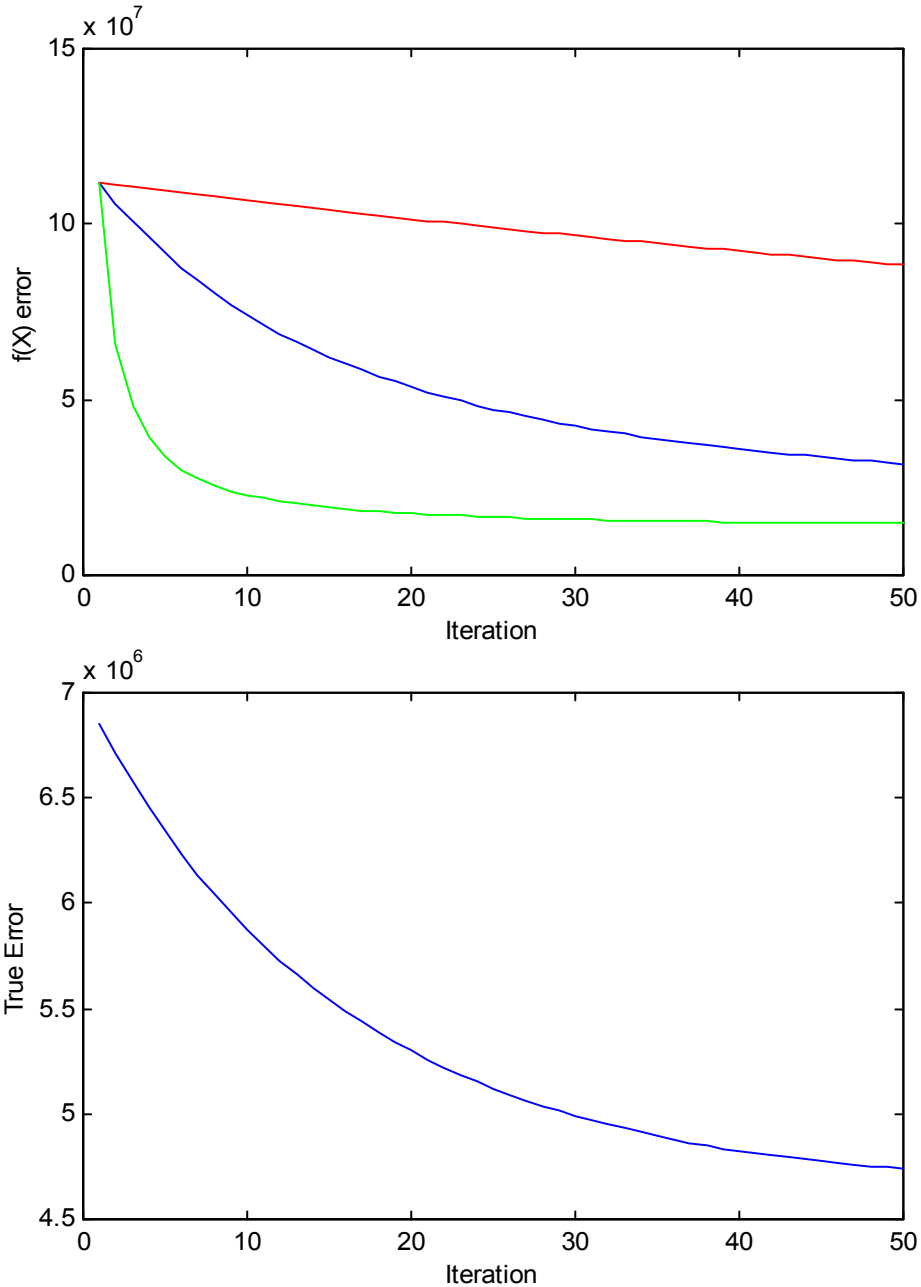
ציור 6.7 מראה גרף של ערך הפונקציה $f(\mathbf{X})$ שמובאת למינימום בתהליך האיטרטיבי כפונקציה של מספר האיטרציה. בהרצה זו נעשה שימוש בלפלסיאן לצורך רגולריזציה, ו- $\sigma_D^2 = 25$. רואים כי חלה ירידה עקבית עד להתכנסות. הגרפים השונים מתייחסים לערכי μ שונים. כפי שכבר נאמר, עבור ערך נמוך מידי ההתכנסות איטית, ועבור ערך גדול מידי חלה התבדרות.



הגרף השני בציור זה מראה את השגיאה $\|\underline{X} - \hat{\underline{X}}\|^2$. כפי שרואים, גם כאן מושגת ירידה עקבית, אם כי אין זו סתירה לתיאוריה לקבל גם עליה - נזכור כי רצוננו האמיתי הוא בהשגת מינימום לפונקציה זו, אך התמונה המקורית אינה קיימת בדיננו. לכן, הפונקציה אותה אנו מביאים למינימום $f(\underline{X})$ הינה חיקוי לשגיאה האמיתית.

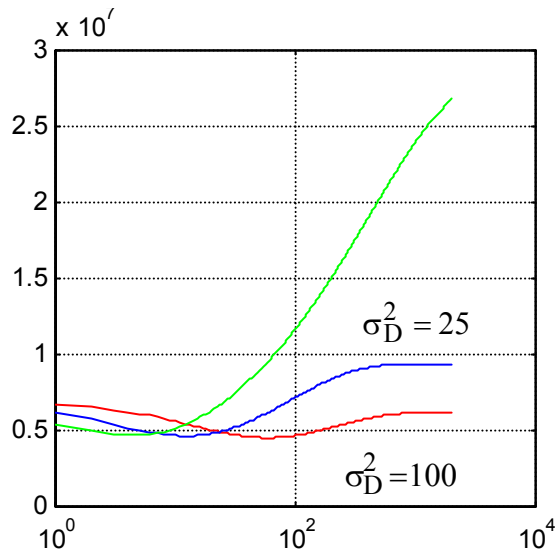
ציור 6.8 מראה תוצאות שחזור MAP ע"י הגישה האיטרטיבית למגוון ערכי σ_D . אנו רואים כי עבור בחירת ערך גבוה התוצאה רועשת מידי (תופעת ה- "צלצולים" - Ringing) בשל הגבר הרעש. עבור ערך נמוך התוצאה חלקה. בציור זה גם מובאת השגיאה האמיתית (מול התמונה האידיאלית) לשם השוואה. רואים כי לשלושת ערכי σ_D מתקבל אפקט דומה בו השגיאה יורדת באיטרציות הראשונות ואז עולה ומתייצבת על ערכה הסופי (כעבור 2000 איטרציות!).

נציג כעת את הפתרון המבוסס על התמרת תדר, כששוב אנו מתייחסים לפתרון ה-MAP. כזכור מפרקים קודמים ומתחילת פרק זה, המטריצה \mathbf{T} , המייצגת את התמרת הפורייה הדיסקרטית הדו-מימדית, מלכסנת מטריצות בלוק סיבוביות. במקרה שלנו, הן \mathbf{H} והן \mathbf{D} בעלות מבנה כזה בשל היותן ייצוג לפעולה ליניארית וקבועה במקום. לכן:



ציור 6.7 - תהליך שחזור איטרטיבי. הגרף העליון מראה את תהליך ההתכנסות ע"י הצגת ערך הפונקציה $f(\underline{x})$ כפונקציה של מספר האיטרציה. הגרף מתייחס לשלושה ערכים $\mu = 1e-2, 1e-3, 1e-4$. בגרף התחתון רואים עבור $u = 1e-3$ את השגיאה מול אות המקור.

$$\begin{aligned} & \left[\mathbf{H}^T \mathbf{H} + \frac{\sigma_N^2}{\sigma_D^2} \mathbf{D}^T \mathbf{D} \right] \hat{\underline{X}} = \mathbf{H}^T \underline{Y} \\ \Rightarrow & \left[\mathbf{T} \mathbf{H}^T \mathbf{T}^H \mathbf{T} \mathbf{H} \mathbf{T}^H + \frac{\sigma_N^2}{\sigma_D^2} \mathbf{T} \mathbf{D}^T \mathbf{T}^H \mathbf{T} \mathbf{D} \mathbf{T}^H \right] \mathbf{T} \hat{\underline{X}} = \mathbf{T} \mathbf{H}^T \mathbf{T}^H \mathbf{T} \underline{Y} \\ \Rightarrow & \left[\mathbf{H}_T^H \mathbf{H}_T + \frac{\sigma_N^2}{\sigma_D^2} \mathbf{D}_T^H \mathbf{D}_T \right] \hat{\underline{X}}_T = \mathbf{H}_T^H \underline{Y}_T \end{aligned}$$



ציור 6.8 - תמונת מקור (למעלה בצד שמאל), תמונה מקולקלת (למעלה בצד ימין), תוצאות שחזור MAP עם ערכים שונים של σ_D^2 (100 - שורה שניה בצד שמאל, 25 - שורה שניה בצד ימין, 1 - שורה תחתונה בצד שמאל). כן מצורף גרף המתאר את השגיאה האמיתית כפונקציה של מספר האיטרציה

הוקטור $\hat{\underline{X}}_T$ הוא התמונה הנדרשת במישור התדר. באופן דומה, \underline{Y}_T היא התמונה הנמדדת, גם בתדר. הקשר בין שתי תמונות אלו נתון ע"י המשוואה הנ"ל, כשהמטריצות הכופלות (בשני האגפים) אלכסוניות! לכן, נוכל לקבל קשר סקלרי פשוט בין האיבר ה- k של \underline{Y}_T לאיבר ה- k של $\hat{\underline{X}}_T$:

$$\left[\mathbf{H}_T^H \mathbf{H}_T + \frac{\sigma_N^2}{\sigma_D^2} \mathbf{D}_T^H \mathbf{D}_T \right] \hat{\underline{X}}_T = \mathbf{H}_T^H \underline{Y}_T \Rightarrow \hat{X}_T[k] = \frac{(\mathbf{H}_T[k])^*}{\left(|\mathbf{H}_T[k]|^2 + \frac{\sigma_N^2}{\sigma_D^2} |\mathbf{D}_T[k]|^2 \right)} Y_T[k]$$

קשר זה הוא בדיוק הקשר שהוצג כמסנן Wiener, כך שהפעם בנינו ממש את המסנן מתוך שיקולים תיאורטיים מבוססי שערך MAP. עם זאת, לגישה אשר נדונה בסעיף זה יש יתרון חשוב - הכללתה לשחזור ע"י סינון לא ליניארי תהיה קלה ופשוטה. נשים לב כי פעולת הלפלסיאן היא מסנן מעביר גבוהים (HPF), ולכן התמרת התדר של פעולת הלפלסיאן הינה גורם שעולה עם התדר. לכן, במכנה בביטוי הנ"ל הגורם של הרגולריזציה יהיה זניח בתדרים נמוכים (כנדרש) ומשמעותי בתדרים גבוהים.

לסיכום ביניים - גישת ה- MAP הביאה לנו יכולת לבנות את מסנן ה- Wiener מתוך שיקולים עמוקים יותר. אבל - לא קיבלנו בינתיים משהו חדש, והביצועים הצפויים (וזה ניכר מציור 6.8) אינם מלבבים. היתרון של שיטת ה- MAP הוא בכך שהיא ניתנת להכללה לסינון לא ליניארי, כפי שנראה בסעיף הבא.

6.5 שיפור פונקציית ההסתברות לתמונות

גילינו כי פעולת הרגולריזציה המבוססת על חלקות גורמת להחלקת יתר של התמונה. כאשר מקטינים את השפעת ה- Prior, אנו מתקרבים לביצועי מסנן מהפך, וזה כמובן גרוע. כך יוצא שעלינו למצוא ערך ביניים אשר מנסה ליהנות מכל העולמות, ובפועל סובל מכל המחלות, והתוצאות מאכזבות. הסיבה לאי-הצלחה כאן היא העובדה שהנחנו הנחה לא נכונה במודע - שכל סביבה בתמונה תעדיף להיות חלקה. ברור שבאזורים חלקים הדבר נכון, אך אזורי שפות הם העיקר בתמונה, ובהם פגענו. הפתרון לפיכך הוא לשנות את חוק ההסתברות שהנחנו על תמונות. נזכור כי הנחתנו המקורית היא:

$$P(\underline{X}) = C \cdot \exp \left\{ - \frac{[\underline{DX}]^T [\underline{DX}]}{2\sigma_D^2} \right\}$$

גישה אפשרית אחת לשפר את חוק ההסתברות זה היא להישאר במסגרת פילוג גאوسي, אך להרשות שונות תלויית מקום. אפשרות אחרת היא לשנות את חוק הפילוג מגאوسي למשהו אחר נכון יותר. אנו נכיר שתי גישות אלו, ונראה כי קיים דמיון ביניהן.

נתחיל באפשרות של הרחבת ההסתברות הגאוסית לשונות תלויית מקום. ההסתברות הגאוסית במקרה הכללי ביותר לוקטור \underline{X} היא:

$$P(\underline{X}) = C \cdot \exp\left\{-\frac{1}{2}[\underline{X} - \underline{M}]^T \mathbf{Q}^{-1}[\underline{X} - \underline{M}]\right\}$$

כאשר \underline{M} הוא הממוצע והמטריצה \mathbf{Q} היא מטריצת הקווריאנס. לפי הגירסה הקודמת בחרנו ממוצע אפס, ומטריצת שונות

$$\mathbf{Q}^{-1} = \mathbf{D}^T \mathbf{D}$$

באשר לממוצע, לא נציע שינוי. לעומת זאת, חסרון בולט במטריצת הקווריאנס היא היותה בעלת מבנה בלוק טופליץ (כיוון ש- \mathbf{D} בעלת מבנה כזה). במילים אחרות, מטריצת הקווריאנס מניחה כי בכל פיקסל יש אותה התנהגות (מקביל לתכונת ה- Linear Space Invariance). על מנת לגוון מעט את הבחירה, נציע הכנסתה של מטריצת משקל אלכסונית \mathbf{W} , ונקבל

$$\mathbf{Q}^{-1} = \mathbf{D}^T \mathbf{W} \mathbf{D}$$

כאשר מטריצה זו תכיל ערכים זהים על אלכסונה, נתלכד עם הבחירה הקודמת, וכאשר הערכים ישתנו לאורך האלכסון, נקבל התנהגות תלויית מקום. משעריך ה- MAP הפעם יתקבל כנקודת המינימום של הביטוי

$$f(\underline{X}) = [\underline{Y} - \mathbf{H}\underline{X}]^T [\underline{Y} - \mathbf{H}\underline{X}] + \frac{\sigma_N^2}{\sigma_D^2} [\mathbf{D}\underline{X}]^T \mathbf{W} [\mathbf{D}\underline{X}]$$

ומביטוי זה נוכל להסיק כיצד עלינו לפעול בבחירת ערכי \mathbf{W} . הוקטור $\mathbf{D}\underline{X}$ היא תמונת נגזרת (כלשהי) של תמונת המקור \underline{X} . באזורים שאמורים להיות חלקים נכפה חלוקת ע"י בחירת משקל 1. במקומות בהם צפויה שפה, נבחר ערך הקרוב לאפס. באופן זה נכפה חלקות באופן תלוי מקום. באשר לבנייתה של \mathbf{W} , נוכל לקחת את התמונה המקולקלת \underline{Y} , ולהשתמש בה כדי לקבוע בכל פיקסל האם ישנה שפה אם לאו.

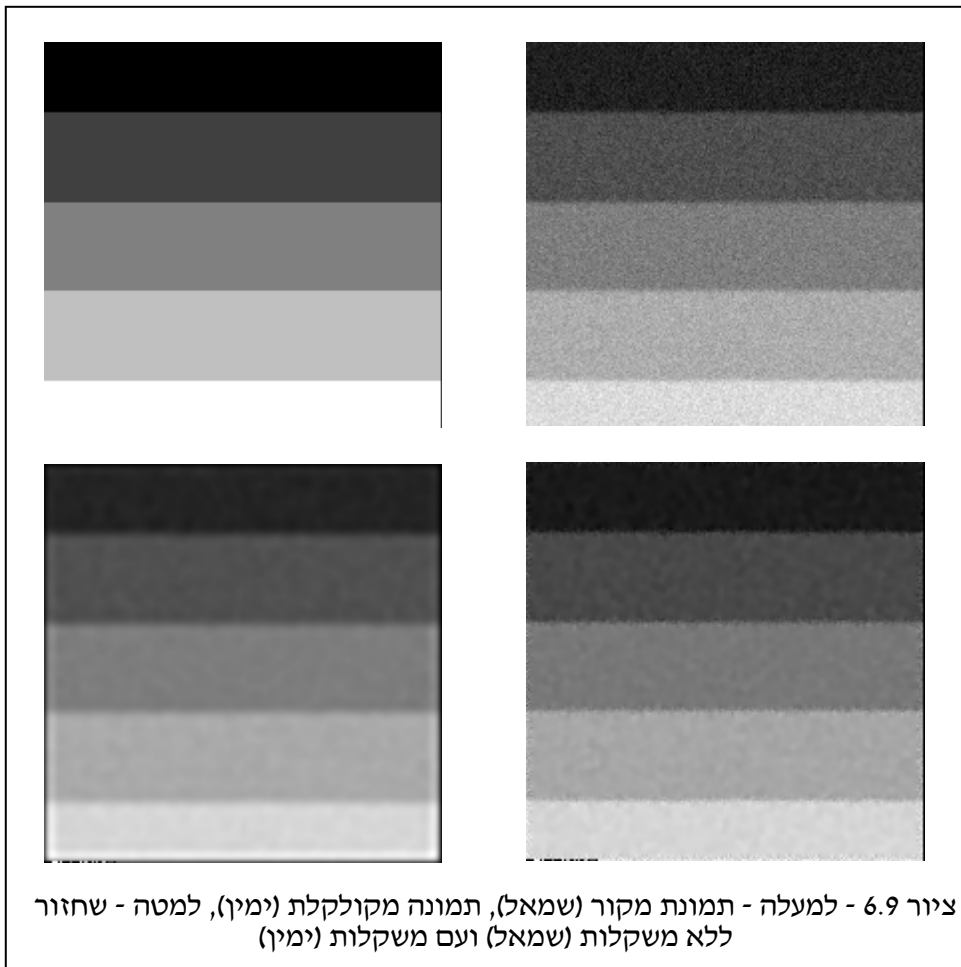
בניגוד למקרה הקודם, הפעם לא ניתן לעבוד בתדר כיוון שחלק מהפעולות אינן קבועות במקום. גישת האיטרציות פועלת באופן הבא:

$$\hat{\underline{X}}_{k+1} = \hat{\underline{X}}_k - \mu \frac{\partial f(\underline{X})}{\partial \underline{X}} \Big|_{\hat{\underline{X}}_k} = \hat{\underline{X}}_k - \mu \left[\mathbf{H}^T (\mathbf{H}\hat{\underline{X}}_k - \underline{Y}) + \frac{\sigma_N^2}{\sigma_D^2} \mathbf{D}^T \mathbf{W} \mathbf{D} \hat{\underline{X}}_k \right]$$

כך שהסכימה דומה מאוד לקודמת, עם שינוי אחד - לאחר הפעלת מסנן גוזר על התמונה, יש לכפול את התמונה בתמונת מסיכה שהינה בדיוק המשקלות שנבחרו.

לעיתים ניתן אף לבחור ערך משקל שלילי במקום בו צפויה שפה לשם השגת פעולה הפוכה של חידוד. בחירת הערך השלילי מסוכנת משום שאז יתכן שהפונקציה הנ"ל כבר אינה קמורה, ובעלת נקודות מינימא רבות. כמו כן הדבר מסכן את יציבות התהליך האיטרטיבי שתואר לעיל.

בציור 6.9 נראה תהליך השחזור עבור תמונה סינתטית. בציור מובאת התמונה המקורית, התמונה המקולקלת, שחזור עם מישקול קבוע (הגישה הישנה) ושחזור עם משיקול תלוי מקום. שגיאת השחזור בני המקרים היא - ללא משקלות - $MSE=27.06$, ועם משקלות - $MSE=21.001$. ערכים אלו התקבלו ע"י החסרת התמונה המקורית מתוצאת השחזור וסיכום ריבועי הפרשים. לצורך זה הושמטה רצועה של 5 פיקסלים מקצוות התמונה, מכיוון שממילא שם פעולותינו לא מוצלחות בכל שיטה. רואים שמושג שיפור משמעותי ע"י הכנסת המשקלות לשחזור. חשוב להבהיר כי השיפור הפוטנציאלי טוב אף יותר כיוון שניתן היה להגדיל את המקדם של הרגולריזציה ולכפות חלקות חזקה יותר באזורים החלקים, כך שפוטנציאל גישה זו לא מוצא עד תום. ציור 6.10 מראה את מטריצת המשקל שנבחרה.





המשקלות נבנו ע"י סידרת הצעדים הבאה :

1. לקיחת התמונה המקולקלת והחלקתה ע"י מסנן 5 על 5. צעד זה נועד לדכא את הרעש בתמונה לפני חישוב נגזרות.
 2. חישוב הנגזרות האופקית והאנכית ע"י מסנן מהצורה $\begin{bmatrix} 1 & 0 & -1 \end{bmatrix}$ (בשני הצירים), וחישוב נורמת הגרדיאנט (סכום ריבועי שני הנגזרות). כעת יש לנו תמונה בה ערך גבוה מעיד על שפה, וערך נמוך מתייחס לאזור חלק.
- בניית משקל בכל נקודה ע"י הנוסחה $-0.01 + 10 / [\text{Grad} + 10]$. כך קיבלנו כי כאשר Grad הוא 0, המשקל הוא 0.99, וכאשר הוא נוטה לאינסוף המשקל הוא -0.01.

עד כה תיארנו כיצד נכליל את הנחת הגאוסיות שלנו ע"י הכנסת מומנט סדר שני מורכב יותר. נראה כעת גישה אחרת שמשמעותה בריחה מהנחת הגאוסיות אל חוקי פילוג אחרים. נחזור לבסיס - אנו לוקחים את התמונה X ומבצעים עליה פעולת גזירה כלשהי D . נסתכל על פיקסל מסויים ונניח כי ערך הנגזרת הוא g . בשיטה הקודמת, g עולה בריבוע והתוצאה היא "מחירו" של פיקסל זה. מדוע לעלות בריבוע ולא ברביעית? מדוע לא ערך מוחלט? וניתן כמובן להציע מגוון אפשרויות אחרות.

ננסה להבין מה המשמעות של בחירות שונות של החזקה בדרך הבאה. נניח שרצוננו לייצר פונקציה $f[k]$ (חד-מימדית) חלקה ודיסקרטית אשר תעבור מ-0 ל-1 ידוע לנו כי הפונקציה ב- $k=0$ היא 0, וב- $k=10$ היא 1. נציע שתי חלופות - לפונקציה f ונבחן את מחירן. חלופה ראשונה תהיה ש- f_1 תעלה בהדרגה 10 מדרגות בגובה 0.1 כל אחת, וחלופה שניה בה f_2 עולה ב- $k=5$ בבת אחת מ-0 ל-1. נניח כי פעולת הנגזרת היא פשוט הפרשים של דגימות סמוכות. לפי השיטה המקורית נקבל

$$\sum_{k=1}^{10} (f_1[k] - f_1[k-1])^2 = 0.1^2 + 0.1^2 + \dots + 0.1^2 = 10 \cdot 0.1^2 = 0.1$$

$$\sum_{k=1}^{10} (f_2[k] - f_2[k-1])^2 = 0^2 + \dots + 1^2 + \dots + 0^2 = 1$$

לכן ברור כי תמחור כזה יעדיף עליה הדרגתית. נניח כעת כי החלפנו את החזקה בערך מוחלט. נקבל:

$$\sum_{k=1}^{10} |f_1[k] - f_1[k-1]| = |0.1| + |0.1| + \dots + |0.1| = 10 \cdot 0.1 = 1$$

$$\sum_{k=1}^{10} |f_2[k] - f_2[k-1]| = 0 + \dots + |1| + \dots + 0 = 1$$

הפעם קיבלנו כי מחיר שני הגישות זהה. נמשיך במגמה זו ונציע שורש על הערך המוחלט ונקבל:

$$\sum_{k=1}^{10} |f_1[k] - f_1[k-1]|^{0.5} = |0.1|^{0.5} + |0.1|^{0.5} + \dots + |0.1|^{0.5} = 10 \cdot 0.1^{0.5} = 3.16$$

$$\sum_{k=1}^{10} |f_2[k] - f_2[k-1]|^{0.5} = 0 + \dots + |1|^{0.5} + \dots + 0 = 1$$

ובמקרה זה הפונקציה תעדיף עליה חדה על-פני עליה מתונה.

מדוע כל זה חשוב? ובכן, אנו מניחים שכאשר מבוצעת הפעולה \underline{DX} , ערכים (בערך מוחלט) בסביבת האפס מתייחסים לאזור נמוך, ולכן הגישה הריבועית יפה להם, כלומר נרצה שמעבר בין רמות יקרה בהדרגה. לעומת זאת, מערך מסוים ואילך כנראה שמדובר בשפות, ולכן נרצה שמעבר בין רמות יקרה באופן חד. לכן, נציע באופן כללי כי פונקצית המחיר לחלקותה של התמונה תהיה:

$$\sum_k \rho(\underline{DX}|_k)$$

כלומר, נבצע כמקודם פעולת גזירה \underline{D} , וכל ערך בתמונת התוצאה יוכנס לפונקציה $\rho(x)$. אם $\rho(x) = x^2$ חזרנו לגרסה המקורית. אם נבחר $\rho(x)$ שיתנהג כפונקציה ריבועית סביב הראשית, וימתן את עלייתו מנקודה מסוימת, הרי שנשיג את שרצינו. ישנן מספר אפשרויות - נציג שתיים מהן:

$$\rho(x) = \frac{x^2}{x^2 + \alpha} \quad \text{or} \quad \rho(x) = \begin{cases} -T(2x + T) & x < -T \\ x^2 & -T \leq x \leq T \\ T(2x - T) & x > T \end{cases}$$

הבחירה הראשונה ידועה בשם פונקציה קושי, והשנייה בשם פונקציה Huber-Markov. שימוש בפונקציות $\rho(x)$ מוכרת בסטטיסטיקה בשם שערך רובסטי (Robust Estimation), והפונקציות $\rho(x)$ זכו לשם M-functions.

הבעיה עם גישת השערך הרובסטי היא שעבור בחירה של פונקציות שעולות בקצב נמוך מליניארי (כמו הדוגמה הראשונה מבין השתיים במשוואה הקודמת), הפונקציה שמובאת למינימום כבר אינה קמורה, ואז מתקבלות תופעות של נקודות מינימה רבות, ובעיות התכנסות למינימה הגלובלית בהתאם לכך. בשל כך, בחירת פונקציה ערך מוחלט $\rho(x) = |x|$ זוכה לפופולריות רבה.

עבור שימוש בפונקציה המחיר החדשה, הפונקציה הכוללת שישי להביא למינימום היא:

$$f(\underline{X}) = [\underline{Y} - \mathbf{H}\underline{X}]^T [\underline{Y} - \mathbf{H}\underline{X}] + \frac{\sigma_N^2}{\sigma_D^2} \rho[\mathbf{D}\underline{X}]$$

כאשר לפי הסימון שלנו, ρ על וקטור הוא פעולת $\rho(x)$ על כל אחד מאיבריו, וסיכום התוצאות. מתוך ביטוי זה ניתן ללכת אחורה ולקבוע כי בעצם הנחנו כי פונקציה ההסתברות של מרחב התמונות נתונה ע"י:

$$P(\underline{X}) = C \cdot \exp\left\{-\frac{\rho[\mathbf{D}\underline{X}]}{2\sigma_D^2}\right\}$$

פונקציות פילוג בעלות מבנה זה קרויות פילוג Gibbs. המשותף לכולן הוא ההסתברות ל- \underline{X} יחסית הפוך דרך אקספוננט למידת האנרגיה שלו, והמעריך הוא סוג של אנרגיה.

האם יש קשר בין גישה זו לגישה הקודמת של בחירת מטריצת משקל? לשם גילוי קשר כזה, נכתוב את משוואת האיטרציה שיש להפעיל לשם מינימיזציה של הביטוי שנכתב למעלה:

$$\hat{\underline{X}}_{k+1} = \hat{\underline{X}}_k - \mu \frac{\partial f(\underline{X})}{\partial \underline{X}} \Big|_{\hat{\underline{X}}_k} = \hat{\underline{X}}_k - \mu \left[\mathbf{H}^T (\mathbf{H}\hat{\underline{X}}_k - \underline{Y}) + \frac{\sigma_N^2}{\sigma_D^2} \mathbf{D}^T \rho'[\mathbf{D}\hat{\underline{X}}_k] \right]$$

משמעות משוואה זו היא שיש לגזור את \underline{X} כמקודם, ואז להעביר כל פיקסל דרך נגזרת הפונקציה $\rho(x)$, ואת התוצאה שוב לגזור. לשם דוגמה - עבור $\rho(x) = 0.5x^2$ ברור כי $\rho'(x) = x$, ואז נקבל כי פעולת ה- ρ' משאירה את הערכים כמות שהם. עבור $\rho(x) = |x|$ הנגזרת תהיה $\rho'(x) = \text{sign}(x)$, כך שכל פיקסל לאחר נגזרת ישווה לאפס - שליליים יהפכו ל-1, וחיוביים ל+1. בהשוואה למשוואה האיטרטיבית במקרה של השימוש במשקל, שני הגורמים הנ"ל מתחלפים:

$$\mathbf{W} \mathbf{D} \hat{\mathbf{X}}_k \Leftrightarrow \rho'[\mathbf{D} \hat{\mathbf{X}}_k] \quad (\rho(x) \text{ גישת})$$

בשני המקרים התהליך מתחיל זהה - התמונה $\hat{\mathbf{X}}_k$ עוברת דרך פעולת הגזירה \mathbf{D} . בשני המקרים הפעולה הבאה היא פעולה על כל פיקסל בנפרד. האחת כופלת במשקל כלשהו, והאחרת מעבירה דרך פונקציה כלשהי. לכן, אם בגישת המשקלות לא היינו משתמשים במטריצת משקל קבועה לאורך כל התהליך האיטרטיבי, אלא משנים אותה בכל איטרציה להיות

$$\mathbf{W}|_{\text{pos}} = \frac{\rho'[\mathbf{D} \hat{\mathbf{X}}_k]_{\text{pos}}}{[\mathbf{D} \hat{\mathbf{X}}_k]_{\text{pos}}}$$

היינו מקבלים אותה תוצאה בדיוק. מה משמעות בחירה זו? לדוגמה עבור $\rho(x) = 0.5x^2$ יתקבל כי \mathbf{W} אינה אלא 1 לכל הפיקסלים. עבור $\rho(x) = |x|$ נקבל כי המשקל יחסי הפוך לגודל הנגזרת בערכה המוחלט, כיוון ש:

$$\mathbf{W}|_{\text{pos}} = \frac{\rho'[\mathbf{D} \hat{\mathbf{X}}_k]_{\text{pos}}}{[\mathbf{D} \hat{\mathbf{X}}_k]_{\text{pos}}} = \frac{\text{sign}(\mathbf{D} \hat{\mathbf{X}}_k|_{\text{pos}})}{\mathbf{D} \hat{\mathbf{X}}_k|_{\text{pos}}} = \frac{1}{|\mathbf{D} \hat{\mathbf{X}}_k|_{\text{pos}}}$$

אם נחשוב על מה שקיבלנו, הרי שלו יכולנו לקבוע את מטריצת המשקלות לא לפי \mathbf{Y} אלא לפי \mathbf{X} היינו צפויים לקבל תוצאה טובה יותר. לכן, כתחליף לחישוב \mathbf{W} מבוסס \mathbf{Y} , מציעה התוצאה שהתקבלה שנחשב את \mathbf{W} בכל איטרציה על סמך תוצאת האיטרציה הקודמת $\hat{\mathbf{X}}_k$, וזה קורב יותר לשימוש ב- \mathbf{X} המקורית.

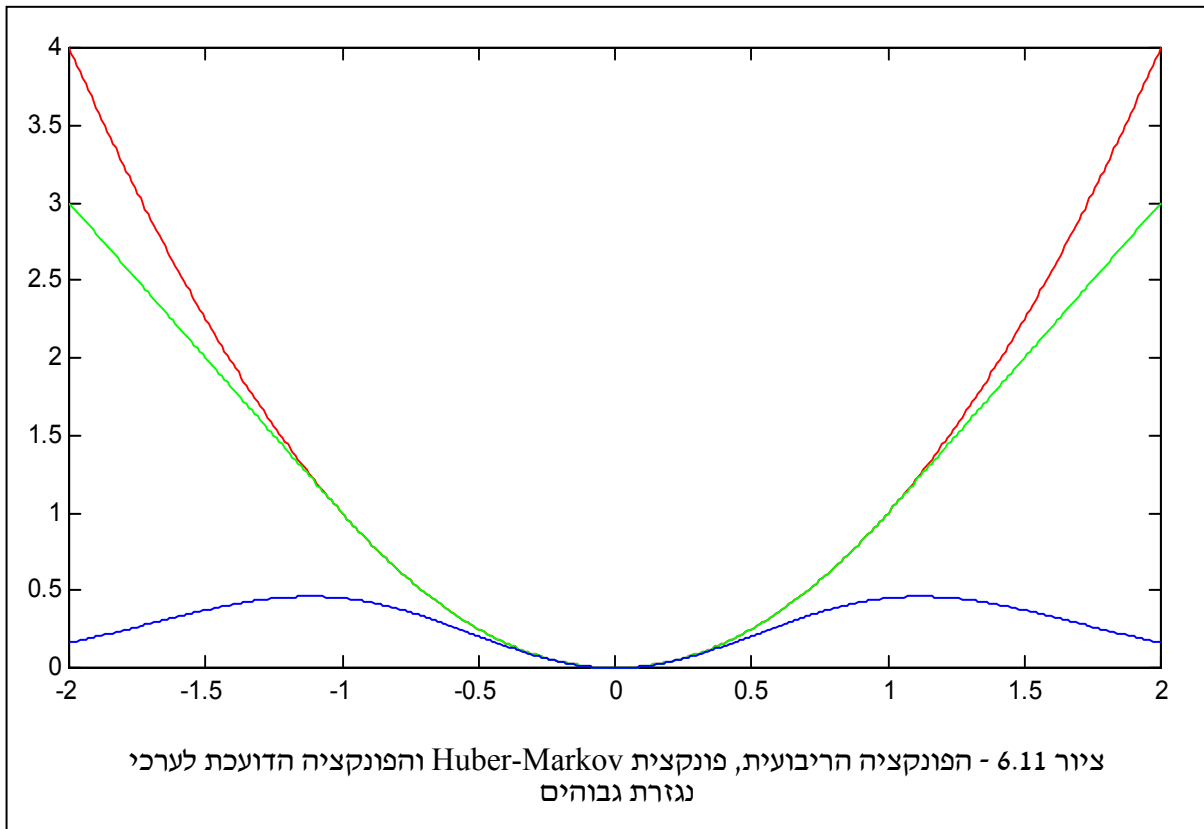
רעיון מעניין הוא הרעיון הבא: מדוע לא לעשות שימוש בפונקציות שאפילו יורדות עבור ערכים גבוהים. למשל, נציע את הפונקציה הבאה:

$$\rho(x) = x^2 \cdot \exp\{-\alpha x^2\}$$

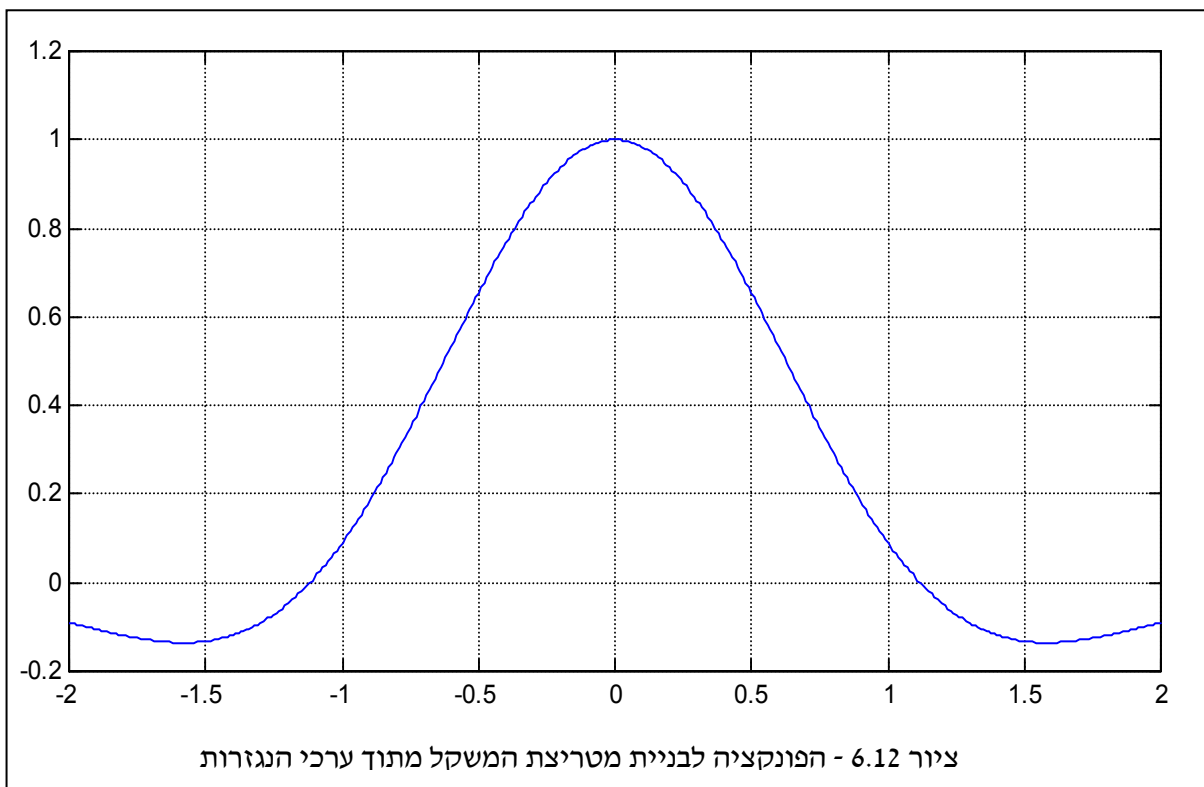
בחירה זו תתנהג כפונקציה ריבועית לערכים נמוכים, אך לאחר עליה מסוימת תתקבל ירידה לאפס יעבור ערכים גבוהים. ציור 6.11 מראה את הפונקציה הריבועית, פונקציית Huber-Markov, ובחירת פונקציה חדשה זו. בכל פונקציה שעולה מונוטונית אנו גובים מחיר שולי (שולי - מתייחס לנגזרת הפונקציה) חיובי.

עבור פונקציות יורדות נעודד היווצרם של שפות אם נגזרתם גדולה. עבור הבחירה הנ"ל, המשקל שמקביל לשימוש יהיה:

$$w(x) = \frac{\rho'(x)}{x} = 2 \exp(-\alpha x^2) [1 - \alpha x^2]$$



עבור הגרף הנ"ל מתקבל כי חישוב במשקל נעשה לפי הגרף הבא, המסגיר כי רעיון בחירת ערכים שליליים אינו נטול בסיס:



6.6 בעיות היפוך אחרות והקשר לבעיית השחזור הקלאסית

אם נסכם את שראינו עד כה, הרי שעבור בעיות בעלות המבנה $\underline{Y} = \mathbf{H}\underline{X} + \underline{N}$, בהן \mathbf{H} מערכת ליניארית כלשהי ידועה, \underline{X} תמונה רצויה שלא ניתנת למדידה, \underline{Y} תמונה אחרת שנמדד במקומה, ו- \underline{N} - וקטור רעש אקראי גאוסי לבן בעל עוצמה σ_N , הצעת שחזור לא רעה ל- \underline{X} היא מינימיזציה של הביטוי:

$$f(\underline{X}) = [\underline{Y} - \mathbf{H}\underline{X}]^T [\underline{Y} - \mathbf{H}\underline{X}] + \frac{\sigma_N^2}{\sigma_D^2} \rho[\mathbf{D}\underline{X}]$$

כאשר \mathbf{D} אופרטור גזירה ליניארי כלשהו, הפונקציה $\rho(x)$ היא פונקציית שערך רובסטית שנבחרת מראש, ו- σ_D פרמטר השולט בעוצמת החלקות שנרצה לכפות.

עד כה השתמשנו בתוצאה הנ"ל לשם פתרונה של בעיית השחזור הקלאסית של בניית שחזור לתמונה מטושטשת ורועשת. נשאלת השאלה - האם ניתן לעשות שימוש בתוצאה הנ"ל לבעיות שחזור אחרות? התשובה היא כן, ואנו נציג מספר בעיות כאלה.

האפשרות הראשונה היא ניקוי תמונה מרעש. במקרה זה, המודל נותר כמות שהוא למעט העובדה ש- $\mathbf{H} = \mathbf{I}$. לכן, במקרה זה, עלינו להביא למינימום את

$$f(\underline{X}) = [\underline{Y} - \underline{X}]^T [\underline{Y} - \underline{X}] + \frac{\sigma_N^2}{\sigma_D^2} \rho[\mathbf{D}\underline{X}]$$

וכמובן התהליך שמתקבל פשוט יותר בהשוואה לזה המקורי (כיוון שאנו פטורים מהפעלת \mathbf{H} בתהליך האיטרטיבי).

אפשרות שניה היא הגדלת תמונה. בהינתן תמונה \underline{Y} , נשאל כיצד להגדילה פי r בכל ציר. אפשרות אחת טריוויאלית היא שכפול פיקסלים, כך שכל פיקסל יוחלף בבלוק של r על r פיקסלים זהים לו. גישה זו מניבה תמונה באיכות ירודה למדי. בפרק 3 בו דנו בשחזור מדגימה תיארונו הגדלה ע"י אינטרפולציה ביליניארית. התוצאה בגישה זו טובה יותר, אך עדיין, באזור שפה בתמונה נקבל מריחה לא נעימה של המעבר, ושוב תתקבל תמונה באיכות ירודה.

אפשרות חדשה היא להתייחס לבעיית ההגדלה כבעיית שחזור. אנו נניח כי הייתה תמונה גדולה ויפה \underline{X} אשר עברה תהליך של טשטוש \mathbf{H} ודילול \mathbf{S} לקבלת התוצאה הירודה \underline{Y} . לכן, הבעיה אותה נרצה לפתור תהיה מינימיזציה של:

$$f(\underline{X}) = [\underline{Y} - \underline{S}\underline{H}\underline{X}]^T [\underline{Y} - \underline{S}\underline{H}\underline{X}] + \frac{\sigma_N^2}{\sigma_D^2} \rho[\underline{D}\underline{X}]$$

באופן זה עשינו שימוש בעובדה שדילול היא פעולה ליניארית הניתנת לתיאור כמכפלת מטריצה בוקטור. מטריצה זו מלבנית ולא ריבועית - אם התמונה בכניסתה היא 100 על 100, ובמוצאה 25 על 25, גודל המטריצה יהיה 625 שורות על 10,000 עמודות. פעולת הדילול היא פעולה ספרבילית וניתן לפיכך לכותבה כמכפלת קרונקר בין שתי פעולות חד-מימדיות.

באשר לדרך בחירת האופרטורים \underline{H} ו- \underline{S} להגדלת תמונה, \underline{S} ייקבע ישירות לפי יחס ההגדלה אותו רצוננו להשיג. \underline{H} יכול להיבחר כ- \underline{I} אם אין רצוננו באפקט של חידוד. בחירת קיומו של טשטוש תגרום לתוצאה להיות חדה יותר. במקרים רבים ניתן לקבוע מדוע התמונה קטנה מידי מתוך שיקולים פיזיקליים, ואז \underline{H} יכולה להיבחר בהתאמה לדרך היוצרה של התמונה.

בעיה אחרת לגמרי אשר גם היא נופלת לקטגוריה של משפחת הבעיות הנ"ל היא בעיית הסופר-רזולוציה. נתונה לנו מצלמה ירודה באיכותה ורצוננו לצלם יעד כלשהו. היעד מיוצג כתמונה אידיאלית \underline{X} , ואילו מה שהמצלמה מניבה היא התמונה \underline{X} לאחר סידרת פעולות מזיקות - הזזה, טשטוש, דילול, ולבסוף רעש. במקום לצלם פעם אחת אנו נצלם M פעמים, בהזזות שונות ולכן נקבל:

$$\begin{aligned} \underline{Y}_1 &= \underline{S}\underline{H}\underline{F}_1 \underline{X} + \underline{N}_1 \\ \underline{Y}_2 &= \underline{S}\underline{H}\underline{F}_2 \underline{X} + \underline{N}_2 \\ &\vdots \\ \underline{Y}_M &= \underline{S}\underline{H}\underline{F}_M \underline{X} + \underline{N}_M \end{aligned}$$

הזזת תמונה (כל עיוות גיאומטרי שהוא!) ניתנת לתיאור כפעולה ליניארית ולכן כמטריצה. זאת כיוון שאם ניקח צירוף ליניארי של שתי תמונות ונזיז את שילובן, יהיה זה כמו להזיז כל אחת מהן ואז לצרפן ליניארית.

ב- M הפעמים בהם צילמנו עשינו שימוש באותה מצלמה ירודה באיכותה, ולכן \underline{H} ו- \underline{S} זהים לכל התמונות שנוצרו. לעומת זאת, תוך כדי צילום זזנו ולכן האוריינטציה הגיאומטרית בין התמונות השתנתה. זו מיוצגת כאן ע"י האופרטורים הליניאריים \underline{F} . כך התקבל סט תמונות \underline{Y} . נוכל לקחת את כל השוואות הנ"ל ולצרפן לאחת לקבלת:

$$\begin{bmatrix} \underline{Y}_1 \\ \underline{Y}_2 \\ \vdots \\ \underline{Y}_M \end{bmatrix} = \begin{bmatrix} \underline{S}\underline{H}\underline{F}_1 \\ \underline{S}\underline{H}\underline{F}_2 \\ \vdots \\ \underline{S}\underline{H}\underline{F}_M \end{bmatrix} \underline{X} + \begin{bmatrix} \underline{N}_1 \\ \underline{N}_2 \\ \vdots \\ \underline{N}_M \end{bmatrix} = \tilde{\underline{H}}\underline{X} + \tilde{\underline{N}}$$

ושוב, אנו רואים אותו מבנה. לכן, שחזור תמונת הסופר-רזולוציה ייעשה ע"י מינימיזציה של הביטוי

$$f(\underline{X}) = [\underline{Y} - \tilde{\mathbf{H}}\underline{X}]^T [\underline{Y} - \tilde{\mathbf{H}}\underline{X}] + \frac{\sigma_N^2}{\sigma_D^2} \rho[\underline{DX}] = \sum_{k=1}^M \|\underline{Y} - \mathbf{SHF}_k \underline{X}\|^2 + \frac{\sigma_N^2}{\sigma_D^2} \rho[\underline{DX}]$$

בדומה לנאמר עבור הגדלת תמונה, על מנת להפעיל תהליך סופר-רזולוציה, עלינו לדעת את האופרטורים המעורבים. ידיעת אופרטורי התנועה כרוכה שתהליך שערך תנועה בין התמונות הנתונות - נושא שיידון בפרקים מתקדמים יותר. כמקודם, S יכול להיקבע ישירות מתוך יחס ההגדלה הרצוי. H ייקבע ע"י הבנת מאפייני המצלמה בה נעשה שימוש, או פשוט בבחירה שרירותית לפי מידת החידוד שרצוננו להשיג בתוצאה הסופית.

פרק 7 – התמרות דיסקרטיות בדו-מימד

143	פרק 7 – התמרות דיסקרטיות בדו-מימד
143	7.1 התמרות ליניאריות - בסיס
148	7.2 התמרת ה-DFT הדו-מימדית ותכונותיה
151	7.3 התמרת ה-DCT הדו-מימדית ותכונותיה
156	7.4 התמרת Hadamard דו-מימדית ותכונותיה
160	7.5 התמרת Haar ומבוא ל-Wavelet
168	7.6 התמרות מבוססות תוכן תמונה
171	7.7 התמרות Over-Complete

כבר פגשנו במסגרת פרקים קודמים דיון בהתמרות פוריה (רציף ודיסקרטי), ואנו שבים במסגרת פרק זה לנושא זה בשל חשיבותו הרבה. הרעיון הבסיסי שילווה אותנו לכל אורכו של הדיון כאן הוא שבהינתן תמונה בגודל N על M כמטריצת ערכי רמות אפור, מבלי ששמנו לב נכפה עלינו ייצוג מאוד מסוים של התמונה מעל הבסיס הטריטוריאלי $e_{[j,k]}$ (תמונה שכולה אפס למעט האיבר ה- $[j,k]$ השווה 1). למה שבחירה זו תחשב טובה יותר מאחרות? מדוע לדבוק בה, למעט פשטותה? אנו נראה כי אמנם קיימות התמרות לבסיסים אחרים, עימם ניתן לפשט בעיות, לחסוך בחישובים ועוד.

עוד עובדה שתרחף מעל רוב פרק זה כאקסיומה היא שאנו מעוניינים לעבוד עם התמרות ליניאריות בלבד, בהן התמונה תיוצג כצירוף ליניארי של תמונות בסיס כלשהו. ברור כי בחירה זו מצמצמת באופן דרסטי את מקבץ האפשרויות, אך מאידך, מבטיחה לנו טיפול מתמטי קל יותר. אנו נתחיל את הדיון בבניית הקשר בין בחירת בסיס אלטרנטיבי והתמרה ליניארית.

7.1 התמרות ליניאריות - בסיס

התמרה ליניארית קונונציונלית הינה פעולה הניתנת לתיאור כמכפלה במטריצה. אם אנו עוסקים בתמונה V בגודל N על M , נוכל לסדרה לוקטור ארוך באורך NM ואז תמונת ההתמרה U נתונה ע"י:

$$\underline{U} = A\underline{V}$$

כשאנו מניחים ש- A מטריצה ריבועית בגודל $[MN \times MN]$. במקרה זה נקבל כי תמונת ההתמרה זהה בגודלה לתמונת המקור. אנו נדרוש כי A תהיה הפיכה על מנת לאפשר התמרה הפוכה מ- \underline{U} ל- \underline{V} . במקרים רבים אנו נלך צעד נוסף במגמת הפשטת ההתמרה ונשתמש רק בהתמרות שביצוען קדימה-אחורה קל עד טריוויאלי - התמרות יוניטריות, עבורן מתקיים:

$$A^{-1} = A^H = [A^T]^*$$

הזכרנו כי התמרה מהווה ייצוג עפ"י בסיס חדש. מהו הקשר בין בסיס חדש זה והמטריצה הנבחרת A ? נניח כי נתון לנו סט תמונות בסיס \underline{b}_k איתו אנו רוצים לעבוד. נניח בשלב זה כי מספר פונקציות הבסיס הוא בדיוק MN (כמספר הדגימות בתמונת המקור \underline{V}), וכן נניח כי בסיס זה פורס את המרחב. לכן נייצג את \underline{V} כצירוף ליניארי של פונקציות בסיס אלה ונקבל:

$$\underline{V} = \sum_{k=1}^{MN} u_k \underline{b}_k = B\underline{U}$$

כאשר B היא מטריצה המכילה את \underline{b}_k כעמודותיה, והוקטור \underline{U} הוא מקדמי הצירוף הליניארי. כיוון שהבסיס שלם (פורס את המרחב) ובעל MN אלמנטים, קיים רק צירוף ליניארי אחד שייתן שיויון. צירוף זה מתקבל ע"י:

$$\underline{U} = B^{-1}\underline{V} = A\underline{V}$$

ומכיוון שהבסיס שלם נקבל כי המטריצה B לא סינגולרית והפיכה. למעשה, קיבלנו כי ההתמרה שיש לבצע על \underline{V} לקבלת \underline{U} היא ההיפוך של B , כלומר B מייצג את ההתמרה ההפוכה. קשר פשוט יותר יתקבל אם נעשה שימוש בוקטורים \underline{b}_k אורתונורמליים, כלומר כאלה שמקיימים

$$\forall k, j \quad \underline{b}_k^T \underline{b}_j = \delta(k - j)$$

אז יתקבל כי המטריצה B יוניטרית, ולכן

$$\underline{U} = B^{-1}\underline{V} = B^H\underline{V} = A\underline{V}$$

ויצא כי שורותיה של המטריצה A מכילות (לאחר פעולת צמוד קומפלקסי) את הוקטורים המהווים את הבסיס החדש.

להתמרה יוניטרית תכונה חשובה אותה אנו מכירים בשם "משפט פרסבל" עבור המקרה הפרטי של התמרת פוריה. אם נחשוב לרגע על התמרה יוניטרית כפעולה גיאומטרית המקבלת וקטור ויוצרת ממנו וקטור חדש, אזי, התמרה יוניטרית לא משנה את עוצמתו (הנמדדת בנורמה) של הוקטור בתהליך, אלא פועלת כסיבוב ברב מימד. קל לראות זאת אלגברית ע"י הקשר:

$$\underline{U} = \underline{A}\underline{V} \Rightarrow \|\underline{U}\|^2 = \|\underline{A}\underline{V}\|^2 = \underline{V}^H \underline{A}^H \underline{A} \underline{V} \stackrel{\substack{\uparrow \\ \underline{A}^H \underline{A} = \underline{I}}}{=} \underline{V}^H \underline{V} = \|\underline{V}\|^2$$

במקרים מסוימים יש עניין בהתמרות בהן יש יותר פונקציות בסיס מאשר נקודות בתמונה. בסיסים אלו נקראים Over-Complete. במקרה כזה, המטריצה B המכילה את וקטורי בסיס אלה כעמודותיה, ולכן יש לה יותר עמודות משורות, ולכן, הוקטור \underline{U} יהיה ארוך יותר מ- \underline{V} . במקרה כזה הקשר המתאר את ההתמרה ההפוכה הוא:

$$\underline{V} = \sum_{k=1}^{MN} u_k \underline{b}_k = \underline{B}\underline{U}$$

להתמרה - תהליך חישוב \underline{U} מתוך \underline{V} - מורכבת יותר הפעם. כיוון שזו מערכת ליניארית בת NM משוואות ויותר מ- MN נעלמים (איברי \underline{U}), ישנם אינסוף אפשרויות של פתרונות \underline{U} , ונשאלת השאלה את מי מהם לבחור. נוכל להעדיף את ה- \underline{U} בעל הנורמה המינימלית, \underline{U} בעל חלקות מירבית בין איבריו, \underline{U} הדליל ביותר האפשרי ועוד. אנו נעסוק לקראת סופו של הפרק בבעיה זו ונציע לה פתרונות מעניינים. נקודה מעניינת היא שלמרבית המקרים לא ניתן יהיה להציג את בניית \underline{U} מתוך B ו- \underline{V} כפעולה ליניארית, דהיינו מכפלה במטריצה.

נחזור כעת להתמרות ריבועיות בהן מספר הבסיסים כמספר הדגימות בתמונה הנתונה. ראינו תהליך בו התחלנו עם התמרה כלשהי, עברנו להתמרות ליניאריות, ופישטנו הלאה ע"י שימוש בהתמרות יונטריות. הפשטה אפשרית אחרת היא השימוש בהתמרות פרידות (ספרביליות). בהתמרות אלו, במקום לעסוק ב- MN תמונות בסיס דו-מימדיות, אנו נניח כי יש לנו שתי משפחות של וקטורים חד-מימדיים:

$$\{\underline{c}_k\}_{k=1}^M \in \mathfrak{R}^M, \quad \{\underline{d}_k\}_{k=1}^N \in \mathfrak{R}^N$$

כלומר, M וקטורים באורך M כל אחד, ו- N וקטורים אחרים באורך N כל אחד. משני סטים אלו נבנה הבסיס הדו-מימדי לפי:

$$b_{[j,k]} = \underline{c}_j \underline{d}_k^T \Rightarrow \underline{b}_k = \underline{c}_j \otimes \underline{d}_k$$

הביטוי הראשון בונה את \underline{b}_k כתמונה בגודל M על N ע"י הכפלת וקטור עמודה בוקטור שורה. הביטוי השני השקול לו עושה שימוש בסידור לקסיקוגרפי לפי עמודות, ויוצר את \underline{b}_k ע"י מכפלת קרונקר. את מלוא הבסיס B ניתן לייצר ע"י הפעולה:

$$B = C \otimes D$$

נראה דוגמה פשוטה להבהרת הרעיון הזה.

דוגמה 7.1: נתונים לנו שני סטים של בסיסים חד-מימדיים באורך 3 ו-4 בהתאמה. בסיסים אלו מתוארים כעמודות המטריצות C ו- D :

$$C = \begin{bmatrix} 1 & 1 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \quad D = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

נבנה בעזרת שני בסיסים אלו בסיס לתמונות דו-מימדיות בגודל 3 (שורות) על 4 (עמודות). יהיו לנו 12 תמונות בסיס בנות 12 אלמנטים כל אחת. לדוגמה, וקטור בסיס ראשון יהיה:

$$\underline{b}_1^T = [1 \quad -1 \quad 0] \otimes [1 \quad 0 \quad 1 \quad 0] = [1 \quad -1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 1 \quad -1 \quad 0 \quad 0 \quad 0 \quad 0]$$

באופן דומה, בניית מלוא הבסיס B תיתן:

$$B = C \otimes D = \begin{bmatrix} 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & -1 & 1 \\ 1 & 1 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 1 & -1 \end{bmatrix}$$

מדוע בסיס פריד (ספרבילי) נחשב פשוט יותר? בשל התכונה הבאה (המתוארת על ההתמרה ההפוכה):

$$\underline{V} = \underline{B}\underline{U} = [\underline{C} \otimes \underline{D}]\underline{U} = [\underline{C} \cdot \underline{U} \cdot \underline{D}]_{CS}$$

משוואה זו אומרת שבמקום לכפול את הוקטור \underline{U} במטריצת ההתמרה המלאה, ההתמרה ניתנת להשגה ע"י פעולה על התמונה במבנה דו-מימדי U , כשכופלים משמאל במטריצה C ומימין במטריצה D . הפעלת C משמאל מבצעת התמרה חד-מימדית על העמודות של U , ולאחר מכן הפעלת D מימין משלימה בביצוע התמרה חד-מימדית על השורות של התוצאה. השוואת כמות המכפלות תחשוף מדוע זה כה אטרקטיבי: בגישה הישירה עלינו לכפול מטריצה בגודל $[MN \times MN]$ בוקטור בגודל MN . לכן, סך המכפלות הוא M^2N^2 . בגישה הספרבילית, הפעלת C ב- U כרוכה ב- M^2N מכפלות (מטריצת התוצאה היא בגודל M על N , וכל איבר בה חושב ע"י M מכפלות), והכפלה נוספת של התוצאה ב- D דורשת N^2M מכפלות נוספות (מטריצת התוצאה היא בגודל M על N וכל איבר בה דורש N מכפלות). היחס בין כמויות החישובים הוא:

$$\text{Ratio} = \frac{M^2N^2}{NM(M+N)} = \frac{MN}{M+N}$$

לדוגמה, עבור $M=N=1$, יש יתרון לגישה הישירה (היחס הוא 0.5), עבור $M=N=2$ היחס הוא 1 - שתי הגישות זהות במורכבותן. עבור $M=N=3$ היחס הוא 1.5, ולמעשה, מערך זה ומעלה יש יתרון הולך ומתחזק לטובת ביצוע התמרה במבנה ספרבילי על פני ביצועה הישיר. עבור הבחירה $M=N$ נקבל כי היחס הנ"ל הוא פשוט $0.5N$ כלומר עולה ליניארית עם N .

עד כה עסקנו בביצוע ההתמרה ההפוכה מ- \underline{U} ל- \underline{V} . מה באשר להתמרה הישירה עבור המקרה הספרבילי? נעשה שימוש בתכונה הבאה למכפלת קרונקר ונקבל את הקשר הבא:

$$[\underline{C} \otimes \underline{D}]^{-1} = [\underline{C}^{-1} \otimes \underline{D}^{-1}] \Rightarrow \underline{B}^{-1}\underline{V} = \underline{U} = [\underline{C}^{-1} \otimes \underline{D}^{-1}]\underline{V} = \underline{C}^{-1}\underline{V}\underline{D}^{-1}$$

כלומר, קיבלנו כי גם ההתמרה הישירה היא ספרבילית, ולכן גם בה קיים אותו מרכיב של חסכון בחישובים. אם בנוסף להיותה ספרבילית, ההתמרות C ו- D יוניטריות, נקבל כי גם פעולת ההיפוך פשוטה, וההתמרה הכוללת קלה מאוד לביצוע. במקרה זה, ההתמרה הכוללת גם היא יוניטרית.

ראינו אם כך סיבות מצוינות להעדיף התמרות ריבועיות, יוניטריות ופרידות. ואמנם, אלה הן ההתמרות הפופולריות - התמרת ה- DFT , התמרת ה- DCT , התמרת ה- $HAAR$, והתמרת ה- $Wavelet$. אבל, כפי שנאמר כבר בסעיף זה, ישנם מקרים בהם נרצה התמרה כללית יותר בשל סיבות ייחודיות.

7.2 התמרת ה-DFT הדו-מימדית ותכונותיה

במסגרת פרק 2 כבר פגשנו את התמרת ה-DFT הדו-מימדית. במסגרת סעיף זה נחזור ונציגה, והפעם באוריינטציה של התמרה יוניטרית, ספרבילית דו-מימדית. ייחודה וחסרונה המרכזי של התמרה זו, בהשוואה להתמרות אחרות, היא עובדת היותה קומפלקסית. ברוב מוחלט של המקרים נעסוק בתמונות מעל הממשיים, והמעבר לערכי התמרה קומפלקסיים יהווה עבורנו מטרד.

התמרת ה-DFT הדו-מימדית צומחת מתוך התמרת DFT חד-מימדית. התמרה זו מוגדרת באופן הבא: לוקטור $v[m]$ מעל התמך $\Omega_v = \{m \mid 0 \leq m \leq M-1\}$, התמרת ה-DFT החד-מימדית נתונה ע"י

$$V[k] = F_{1D-DFT} \{v[m]\} = \frac{1}{\sqrt{M}} \sum_{m=0}^{M-1} v[m] \exp\left\{-j2\pi mk\right\}$$

עבור התמך $\Omega_v = \{k \mid 0 \leq k \leq M-1\}$. ההתמרה ההפוכה מוגדרת ע"י

$$v[m] = F_{1D-DFT}^{-1} \{F[k]\} = \frac{1}{\sqrt{M}} \sum_{k=0}^{M-1} F[k] \exp\left\{j2\pi mk\right\}$$

עבור תמונה $f[m,n]$ מעל התמך $\Omega_f = \{[m,n] \mid 0 \leq m \leq M-1, 0 \leq n \leq N-1\}$, התמרת ה-DFT הדו-מימדית נתונה ע"י ביצוע התמרה על שורותיה של התמונה, ולאחר מכן על עמודותיה. סדר זה יכול להתחלף ללא שינוי, כיוון שהפעולה המבוצעת פרידה. לכן, בשלב ראשון מתקבל:

$$F[m,\ell] = F_{1D-DFT} \{f[m,n]\} = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} f[m,n] \exp\left\{-j2\pi\left(\frac{n\ell}{N}\right)\right\}$$

ובשלב שני מתקבל:

$$\begin{aligned} F[k,\ell] &= F_{1D-DFT} \{F[m,\ell]\} = \frac{1}{\sqrt{M}} \sum_{m=0}^{M-1} F[m,\ell] \exp\left\{-j2\pi\left(\frac{mk}{M}\right)\right\} = \\ &= F_{2D-DFT} \{f[m,n]\} = \frac{1}{\sqrt{MN}} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} f[m,n] \exp\left\{-j2\pi\left(\frac{mk}{M} + \frac{n\ell}{N}\right)\right\} \end{aligned}$$

באופן דומה, ניתן לראות כי ביצוע ההתמרה ההפוכה בשני שלבים של התמרה הפוכה חד-מימדית מניבים את הקשר הבא:

$$f[m, n] = F_{DFT}^{-1} \{F[k, \ell]\} = \frac{1}{\sqrt{NM}} \sum_{k=0}^{M-1} \sum_{\ell=0}^{N-1} F[k, \ell] \exp \left\{ j2\pi \left(\frac{mk}{M} + \frac{n\ell}{N} \right) \right\}$$

כאשר התמך במישור תדר נתון ע"י $\Omega_F = \{[k, \ell] | 0 \leq k \leq M-1, 0 \leq \ell \leq N-1\}$, כך שלעניינינו, יוצא כי רק MN ערכים מגדירים במלואה את ההתמרה. כל שמבדיל בין ההתמרות קדימה ואחורה הוא סימנו של המעריך המרוכב. ההגדרות המוצעות שונות במקצת מאילו המופיעות בפרק 2, והבדל זה הינו בקבוע לפני ההתמרה אשר מבטיח כי מדובר בהתמרה יוניטרית.

כשדנים ב-DFT מקובל להשתמש בסימון $W_N^k = \exp\{-j2\pi k/N\}$. אז ההתמרה וההתמרה ההפוכה נתונות ע"י ייצוג מטריצי מהצורה (נתייחס להתמרה חד-מימדית):

$$\begin{bmatrix} V[0] \\ F[1] \\ \vdots \\ F[M-1] \end{bmatrix} = \frac{1}{\sqrt{M}} \begin{bmatrix} W_M^0 & W_M^0 & \dots & W_M^0 \\ W_M^0 & W_M^1 & \dots & W_M^{M-1} \\ \vdots & \vdots & \ddots & \vdots \\ W_M^0 & W_M^{M-1} & \dots & W_M^{(M-1)^2} \end{bmatrix} \begin{bmatrix} v[0] \\ v[1] \\ \vdots \\ v[M-1] \end{bmatrix} \Rightarrow \underline{F} = \mathbf{W}_M \underline{f}$$

מטריצת ההתמרה \mathbf{W} הינה מטריצה יוניטרית, ולכן, היפוכה קיים ופשוט (נתון ע"י שחלוף וצמוד). ההתמרה הדו-מימדית ניתנת אף היא לתיאור מטריצי וקטורי. למעשה ישנן שתי אפשרויות לייצגה. בדרך הראשונה, כשאנו מתייחסים ל- f כמטריצת הפיקסלים בגודל M שורות על N עמודות, הכפלתה משמאל במטריצה \mathbf{W}_M תביא לביצוע התמרה על עמודותיה של התמונה. הכפלת התוצאה מימין במטריצה \mathbf{W}_N משלימה ע"י ביצוע התמרה על שורות התמונה, וכך קיבלנו את F - תמונת ההתמרה הדו-מימדית. לכן:

$$F = \mathbf{W}_M f \mathbf{W}_N$$

בגישה השניה, נמיר את התמונה למבנה וקטורי ע"י סידור לקסיקוגרפי לפי עמודות ונקבל וקטור \underline{f} באורך MN . וקטור זה יוכפל באופן הבא:

$$\underline{F} = \mathbf{W}_{2D-DFT[M,N]} \underline{f} = [\mathbf{W}_M \otimes \mathbf{W}_N] \underline{f}$$

מטריצה זו בגודל MN על MN מבצעת את פעולת ההתמרה במלואה על הוקטור f מניבה את תמונת ההתמרה כשגם היא במבנה וקטור ע"י סידור לקסיקוגרפי לפי עמודות. קשר זה הכרנו עוד בפרק 2 כשראינו כי פעולות ליניאריות דו-מימדיות פרידות ניתנות לתיאור כמכפלת קרונקר של שתי פעולות חד-מימדיות. ברור כי המטריצה הכוללת מהווה אף היא מטריצה יוניטרית (ניתן להוכיח זאת ע"י שימוש בתכונות מכפלת קרונקר).

נסקור מספר תכונות חשובות של ה-2D-DFT:

1. **מחזוריות:** בהינתן תמונה $f[m,n]$ מעל התמך $\Omega_f = \{[m,n] | 0 \leq m \leq M-1, 0 \leq n \leq N-1\}$, התמרת ה-DFT הדו-מימדית הנתונה כ- $F[k,\ell]$ מוגדרת מעל התמך $\Omega_F = \{[k,\ell] | 0 \leq k \leq M-1, 0 \leq \ell \leq N-1\}$. לכל נקודה דיסקרטית $[k,\ell]$ מחוץ לתחום זה מתקבלת מחזוריות לפי הקשר:

$$\forall p, q \in \mathbf{Z}, F[k, \ell] = F[k + pM, \ell + qN]$$

2. **ליניאריות, יוניטריות, וספרביליות:** פעולת ה-2D-DFT המתוארת ע"י המטריצה $\mathbf{W}_{2D-DFT[M,N]}$ מהווה התמרה ליניארית, ריבועית, יוניטרית, ופרידה. כל תכונות אלה קלות לאבחנה מתוך כל שנאמר עד כה. מטריצה זו הינה גם מטריצה סימטרית.

3. **יישום מהיר:** ביצוע פעולת ה-DFT הדו-מימדית על תמונה בגודל M על N יכולה להיעשות ע"י M פעולות FFT חד-מימדיות לאורך N , ולאחר מכן, N פעולות FFT חד-מימדיות לאורך M . כיוון שפעולת FFT על N איברים דורשת $N \log_2 N$ פעולות (כל פעולה היא חיבור ומכפלה), סך הפעולות הנדרש להתמרה מלאה הוא:

$$MN \log_2 N + NM \log_2 M = (NM) \log_2 (NM)$$

זאת בהשוואה לביצוע ישיר של ההתמרה ע"י $M^2 N^2$ פעולות. נשים לב שלו לקחנו את התמונה והמרנו אותה לסידור לקסיקוגרפי, פעולת FFT חד-מימדית הייתה דורשת אותה כמות של פעולות.

4. **לכסון מטריצות בלוק סיבוביות:** מטריצת התמרת ה-DFT החד-מימדית מלכסנת מטריצות בעלות מבנה סיבובי (ראה פרק 2 להגדרה מדויקת של מבנה זה). בשל כך, מטריצת התמרת ה-DFT הדו-מימדית מלכסנת מטריצות בלוק סיבוביות.

5. **סימטרייה צמודה:** לתמונה ממשית $f[m,n]$ ההתמרה $F[k,\ell]$ מקיימת את הקשרים הבאים:

$$F[k, \ell] = F^*[M-k, N-\ell] \quad \forall 0 \leq k < M, 0 \leq \ell < N$$

כתוצאה מכך, למרות שההתמרה מכילה NM ערכים קומפלקסיים (כלומר שווה ערך ל- $2MN$ ערכים ממשיים), הרי שרק MN מתוכם בלתי תלויים. ראינו כבר מקבילה של תכונה זו עבור התמרות הפורייה שנסקרו בפרק 2.

6. **משפט הקונבולוציה:** בהינתן שתי תמונות $f_1[m, n]$ ו- $f_2[m, n]$ בגודל זהה M על N , תוצאת ה-DFT הדו-מימדי של הקונבולוציה הציקלית ביניהם זהה לתוצאת המכפלה של התמרותיהן. כלומר, אם:

$$z[m, n] = \sum_{k=0}^{M-1} \sum_{j=0}^{N-1} f_1[k, j] f_2[(m-k) \bmod M, (n-j) \bmod N]$$

$$. Z[k, \ell] = F_1[k, \ell] F_2[k, \ell] \text{ אזי}$$

7. **משפט פרסבל:** בהינתן שתי תמונות $f_1[m, n]$ ו- $f_2[m, n]$ בגודל זהה M על N , תוצאת מכפלתם הפנימית זהה המקום ובתדר, דהיינו:

$$\sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f_1[m, n] f_2^*[m, n] = \sum_{k=0}^{M-1} \sum_{\ell=0}^{N-1} F_1[k, \ell] F_2^*[k, \ell]$$

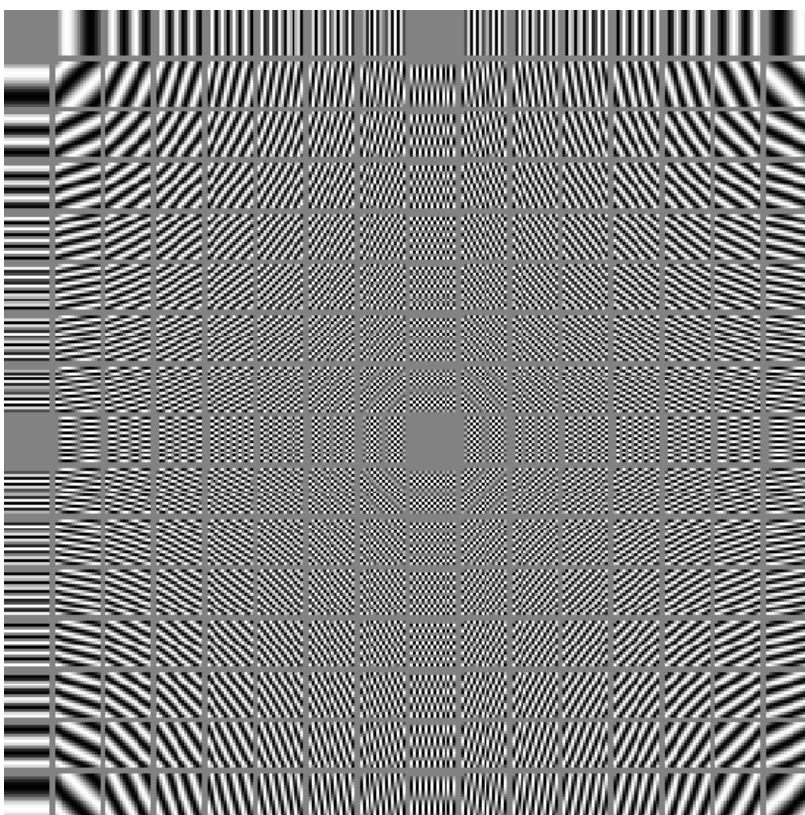
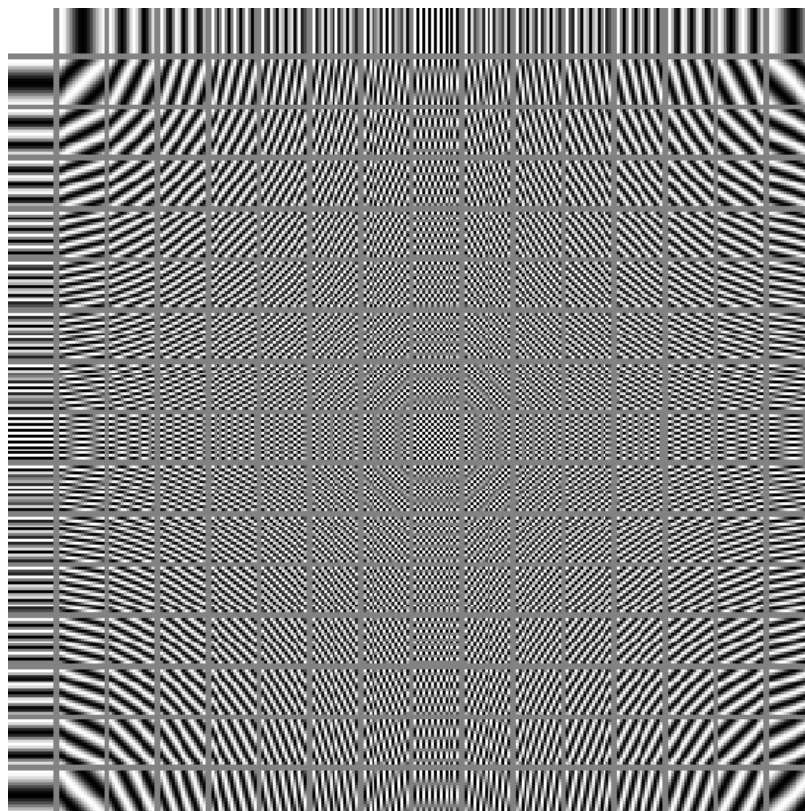
תוצאה זו לא אמורה להפתיע אותנו כיוון שכפי שנאמר כבר בתחילת פרק זה, כל התמרה יוניטרית ממלאת תכונה זו, ובפרט ה-2D-DFT.

8. **וקטורי הבסיס להתמרה:** התמרת ה-DFT הדו-מימדית מציעה שימוש ב- MN תמונות בסיס בגודל M על N כל אחת, לצורך ייצוגה של תמונה נתונה. על מנת לקבל את תמונת הבסיס ה- k , עלינו לבנות וקטור מלאכותי \underline{f} בו כל הערכים 0 למעט האיבר ה- k שיהיה 1, ולבצע התמרת פוריה הפוכה עליו, לקבלת \underline{f} . מכאן נובע באופן פשוט כי עמודותיה של המטריצה $\mathbf{W}_{2D-DFT[M, N]}^H = \mathbf{W}_M^H \otimes \mathbf{W}_N^H$ הם וקטורי הבסיס של ההתמרה.

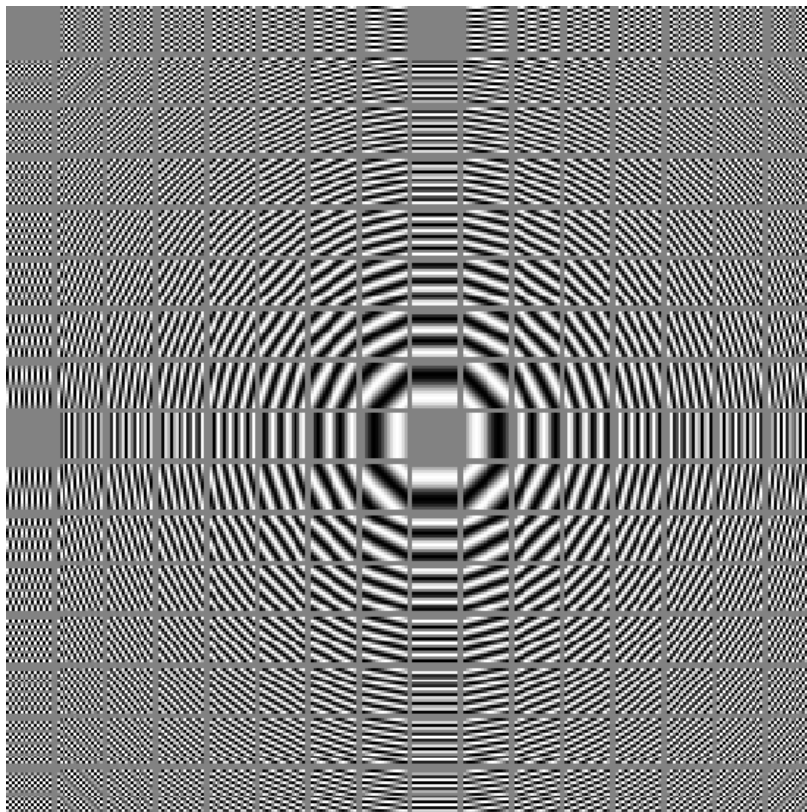
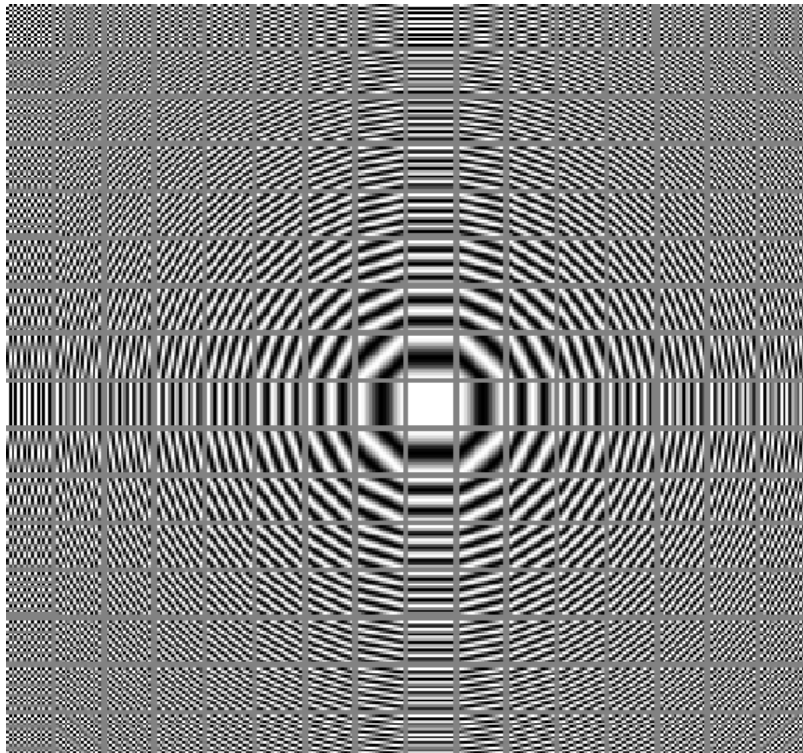
דוגמה 7.1: נציג את תמונות הבסיס של התמרת ה-2D-DFT לתמונות בגודל 16 על 16. תהיינה 25 תמונות בסיס, אך כיוון שכל תמונת בסיס כזו היא קומפלקסית, יש להציג אותה כשתי תמונות עבור החלק הממשי והחלק המדומה. ציור 7.1 מראה את משפחת הבסיסים הזו.

ניתן לראות כי תמונת הבסיס הראשונה קבועה, וממנה התדר הולך ומתחזק כשהוא הרמוני (פונקציית קוסינוס או סינוס) טהור, ועם מגוון כיוונים. מעבר לחצי (אחרי האינדקס 9) מתקבלת ירידה שוב.

נוח לראות את תמונות הבסיס לאחר פעולת הזזה למרכז, כך שהתדרים עולים כשמתרחקים באופן רדיאלי מהמרכז - ציור 7.2 נותן ראייה זו. כפי שניתן לראות, סביב נקודת המרכז התדר עולה ובמגוון זוויות לפי כיוון ההתקדמות. באופן זה כל שפה בתמונה יכולה להיות מיוצגת.



ציור 7.1 - וקטורי הבסיס של התמרת ה-2D-DFT. למעלה - החלק הממשי, למטה - החלק המדומה



ציור 7.2 - וקטורי הבסיס של התמרת ה-2D-DFT לאחר פעולת הזזה למרכז (FFTSHIFT). למעלה - החלק הממשי, למטה - החלק המדומה

7.3 התמרת ה-DCT הדו-מימדית ותכונותיה

התמרת ה-DCT החד-מימדית לוקטור $v[m]$ באורך M והתמרה ההפוכה לה נתונות ע"י שני

הקשרים הבאים:

$$V[k] = \begin{cases} \sqrt{\frac{1}{M}} \sum_{m=0}^{M-1} v[m] \cos\left[\frac{\pi(2m+1)k}{2M}\right] & k = 0 \\ \sqrt{\frac{2}{M}} \sum_{m=0}^{M-1} v[m] \cos\left[\frac{\pi(2m+1)k}{2M}\right] & k \neq 0 \end{cases}$$

$$v[m] = \sqrt{\frac{1}{M}} V[0] + \sqrt{\frac{2}{M}} \sum_{k=1}^{M-1} V[k] \cos\left[\frac{\pi(2m+1)k}{2M}\right]$$

גם הפעם מדובר בהתמרה ליניארית אך ללא ערכים קומפלקסיים. ניתן, בדומה ל-DFT, להציע מבנה מטריצי לתיאור ההתמרה, וקל לראות מה יהיה תוכנה של מטריצה מעין זו, אשר תסומן ב- C_M . מטריצה זו אינה סימטרית, אך היא ממשית ויוניטרית. לכן, היפוכה נתון פשוט ע"י פעולת השחלוף (ללא צמוד קומפלקסי).

בדומה ל-DFT, התמרה דו-מימדית לתמונה f בגודל M על N תיעשה ע"י ביצוע התמרה לשורות

ולעמודות, כלומר ביצוע פעולה פרידה. לכן:

$$F = C_M f C_N$$

או בגישה של תמונה במבנה וקטורי ע"י סידור לקסיקוגרפי לפי עמודות:

$$\underline{F} = C_{2D-DCT[M,N]} \underline{f} = [C_M \otimes C_N] \underline{f}$$

ומטריצה זו הינה שוב יוניטרית, המייצגת התמרה פרידה. נסקור בקצרה וללא הוכחות את תכונותיה של התמרה זו.

1. **ליניאריות, יוניטריות, וספרביליות:** פעולת ה-2D-DCT המתוארת ע"י המטריצה $\mathbf{W}_{2D-DCT[M,N]}$ מהווה התמרה ליניארית, ריבועית, יוניטרית, ופרידה. כל תכונות אלה קלות לאבחנה מתוך כל שנאמר עד כה.

2. **יישום מהיר:** התמרת ה-DCT קשורה באופן הדוק לפעולת ה-DFT. כתוצאה מכך, גם ביצוע ה-DCT הדו-מימדי על תמונה בגודל M על N יכול להיעשות באופן מהיר ע"י שימוש בפעולות FFT. סך הפעולות הדרוש זהה לביצוע התמרת DFT דו-מימדי בגודל זהה.

3. **לכסון מטריצות בלוק מרקוביות:** מטריצת התמרת ה-DCT החד-מימדית מלכסנת במדויק מטריצות תלת-אלכסוניות, בעלות המבנה הבא:

$$Q = \begin{bmatrix} 1-\alpha & -\alpha & 0 & & \\ -\alpha & 1 & -\alpha & 0 & \\ & \ddots & \ddots & \ddots & \\ & & 0 & -\alpha & 1 & -\alpha \\ & & & 0 & -\alpha & 1-\alpha \end{bmatrix}$$

בשל תכונה זו, ניתן להראות כי מטריצות טופליץ מרקוביות הבנויות באופן הבא:

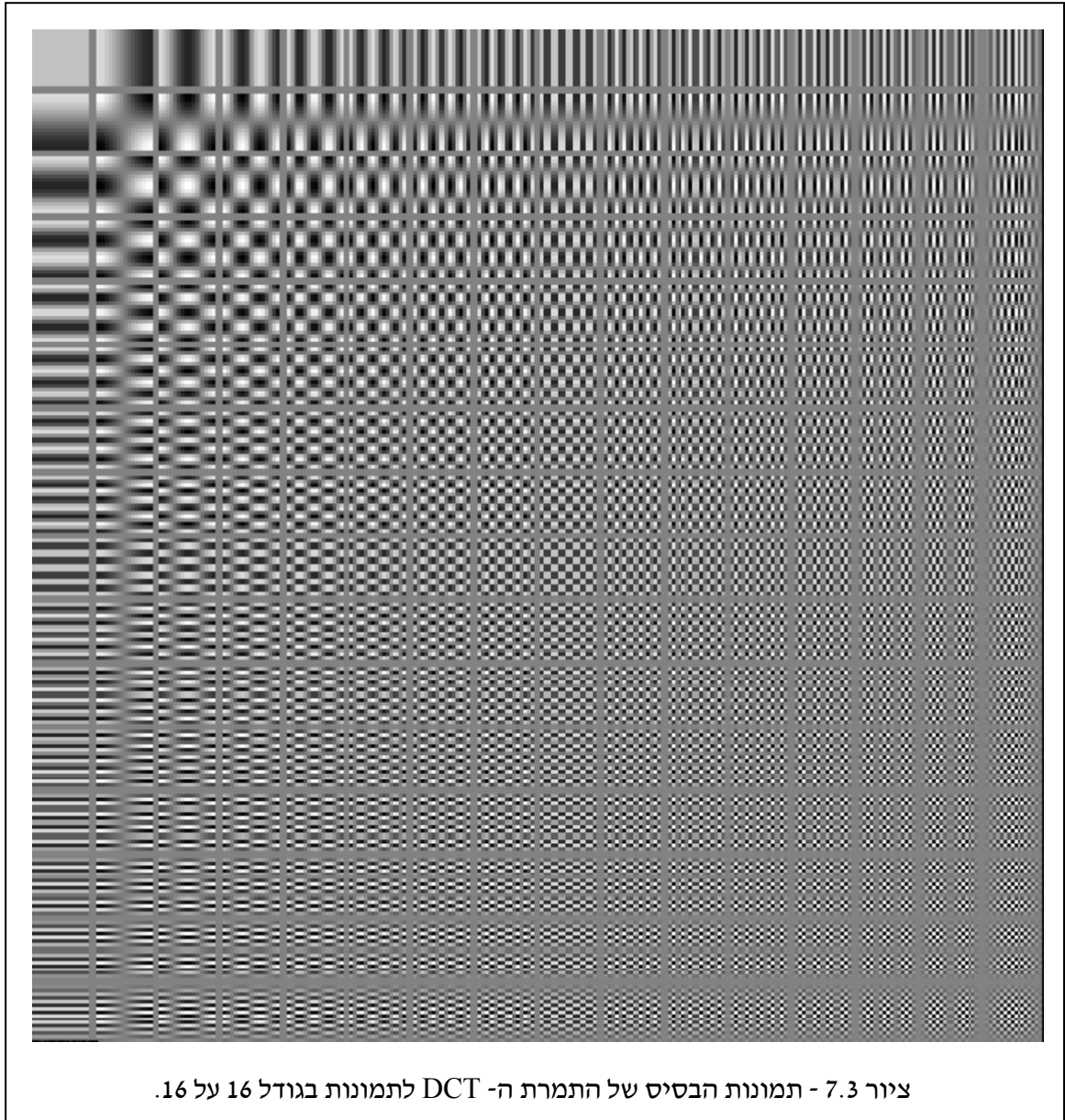
$$R = \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{M-1} \\ \rho & 1 & \rho & \rho^2 & \dots \\ & \ddots & \ddots & \ddots & \\ & \rho^2 & \rho & 1 & \rho \\ \rho^{M-1} & & \rho^2 & \rho & 1 \end{bmatrix}$$

מתלכסנות ע"י התמרת ה-DCT בקירוב טוב, אם $\rho \rightarrow 1$. הסיבה לכך היא שעבור $\rho \rightarrow 1$ מתקבל כי היפוכה של R היא בדיוק טוב Q, ולכן ברור שה-DCT מהווה מטריצה מלכסנת. תכונה זו חיונית כיוון שכפי שנראה בהמשך, מטריצת טופליץ מרקובית מתארת התנהגות אופיינית של אותות בטבע, ולכסונה חשוב לצורך ייצוג נאות או דחיסה. אנו נשוב לתכונה חשובה זו כשנדבר על התמרת KLT.

בשל כל האמור לעיל, כאשר למטריצה יש מבנה בלוק טופליץ מרקובי, מתקבל כי התמרת ה-DCT הדו-מימדית מלכסנת אותה בקירוב טוב, וזה נכון כאשר $\rho \rightarrow 1$.

4. **קונבולוציה:** בשל הקשר ההדוק בין ה-DCT ל-DFT, ניתן להראות כי ניתן בעזרת DCT לבצע קונבולוציה מהירה יותר, בדומה לדרך בה הדבר מוצע ב-DFT. עם זאת, הקרשים מורכבים יותר, ואין זה הצד החזק של ה-DCT.

5. **וקטורי הבסיס להתמרה:** ציור 7.3 מראה את וקטורי הבסיס של התמרת ה-DCT הדו-מימדית לתמונות בגודל 16 על 16. ניכר הדמיון הרב להתמרת ה-DFT.



7.4 התמרת Hadamard הדו-מימדית ותכונותיה

התחלנו את הדיון בהתמרות ספציפיות בהתמרת DFT, וראינו כי היא כופה עלינו חשבונות עם מספרים קומפלקסיים. התמרת ה-DCT הקלה בעיה זו בהציעה התמרה יוניטרית ממשית. כעת נמשיך קו זה של הפשטה ונציע התמרה בה אין כלל מכפלות במספרים כלשהם - כל איברי המטריצה הממירה יהיו 1 או -1. זוהי התמרת Hadamard, המוכרת גם בשם התמרת Hadamard-Walsh.

כמקודם, נתחיל בתיאור ההתמרה החד-מימדית לוקטור באורך M , ונניח הפעם כי M מהווה חזקה שלמה של 2 (דרישה זו מוכרת לנו גם בהקשר ל-DFT, שם היא חשובה לשם יישום יעיל של ה-

(FFT). בניגוד לדרך בה נקטנו בדיון בהתמרות קודמות, הפעם נבנה ישירות את מטריצת ההמרה H .
התהליך הוא רקורסיבי - נתחיל במטריצה בסיסית בגודל 2 על 2 :

$$H_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

התמרה זו תקיפה לוקטורים באורך 2, וניכר כי היא יוניטרית. לשם בניית התמרה לאורך 4 נבנה את H_2 ע"י:

$$H_2 = H_1 \otimes H_1 = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$$

בשל דרך בנייתה, ברור כי התמרה זו יוניטרית. התהליך יכול להמשיך שוב ושוב, ומתקבל כי :

$$H_n = H_{n-1} \otimes H_1$$

והתוצאה היא מטריצה המכילה 1 ו- (-1), שהינה יוניטרית וסימטרית. בשל כך, ההתמרה ההפוכה נתונה ע"י אותה מטריצה ממש.

ברור כי התמרה זו לתמונות בגודל M על N (שוב, בהנחה כי אלו הם חזקות שלמות של 2) מתקבלת ע"י מכפלת קרונקר בין שתי מטריצות התמרה חד-מימדיות. עובדה זו מבטיחה שההתמרה הכוללת לתמונה תהווה מטריצה סימטרית, יוניטרית, בעלת איברים 1 ו- (-1), ולמעשה, זהה לחלוטין להתמרה לוקטור חד-מימדי באורך NM .

כשמסתכלים על שורה מסוימת במטריצת ההמרה H , מספר החלפות הסימן מתאר מקבילה לתדר. מספר החלפות הסימן מוכר בשם Sequency, וניתן לראות כי שורות המטריצה H אינן מסודרות בסדר Sequency מונוטוני. מקובל לעיתים לקחת את מטריצת ההמרה שנבנתה ולסדר את שורותיה כך שאמנם יתקבל Sequency גדל מונוטונית.

דוגמה 7.2: נראה מהי מטריצת Hadamard לוקטורים באורך 8, וכיצד לסדרה. המטריצה נתונה ע"י :

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1- & 1 & 1- & 1 & 1- & 1 & 1- & 1 \\ 1- & 1- & 1 & 1 & 1- & 1- & 1 & 1 \\ 1 & 1- & 1- & 1 & 1 & 1- & 1- & 1 \\ 1- & 1- & 1- & 1- & 1 & 1 & 1 & 1 \\ 1 & 1- & 1 & 1- & 1- & 1 & 1- & 1 \\ 1 & 1 & 1- & 1- & 1- & 1- & 1 & 1 \\ 1- & 1 & 1 & 1- & 1 & 1- & 1- & 1 \end{pmatrix}$$

ה- Sequency לפי שורות הוא - 0, 7, 3, 4, 1, 6, 2, ו- 5. המטריצה לאחר סידורה תהיה :

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1- & 1- & 1- & 1- & 1 & 1 & 1 & 1 \\ 1 & 1 & 1- & 1- & 1- & 1- & 1 & 1 \\ 1- & 1- & 1 & 1 & 1- & 1- & 1 & 1 \\ 1 & 1- & 1- & 1 & 1 & 1- & 1- & 1 \\ 1- & 1 & 1 & 1- & 1 & 1- & 1- & 1 \\ 1 & 1- & 1 & 1- & 1- & 1 & 1- & 1 \\ 1- & 1 & 1- & 1 & 1- & 1 & 1- & 1 \end{pmatrix}$$

נסקור בקצרה וללא הוכחות את תכונותיה של התמרה זו.

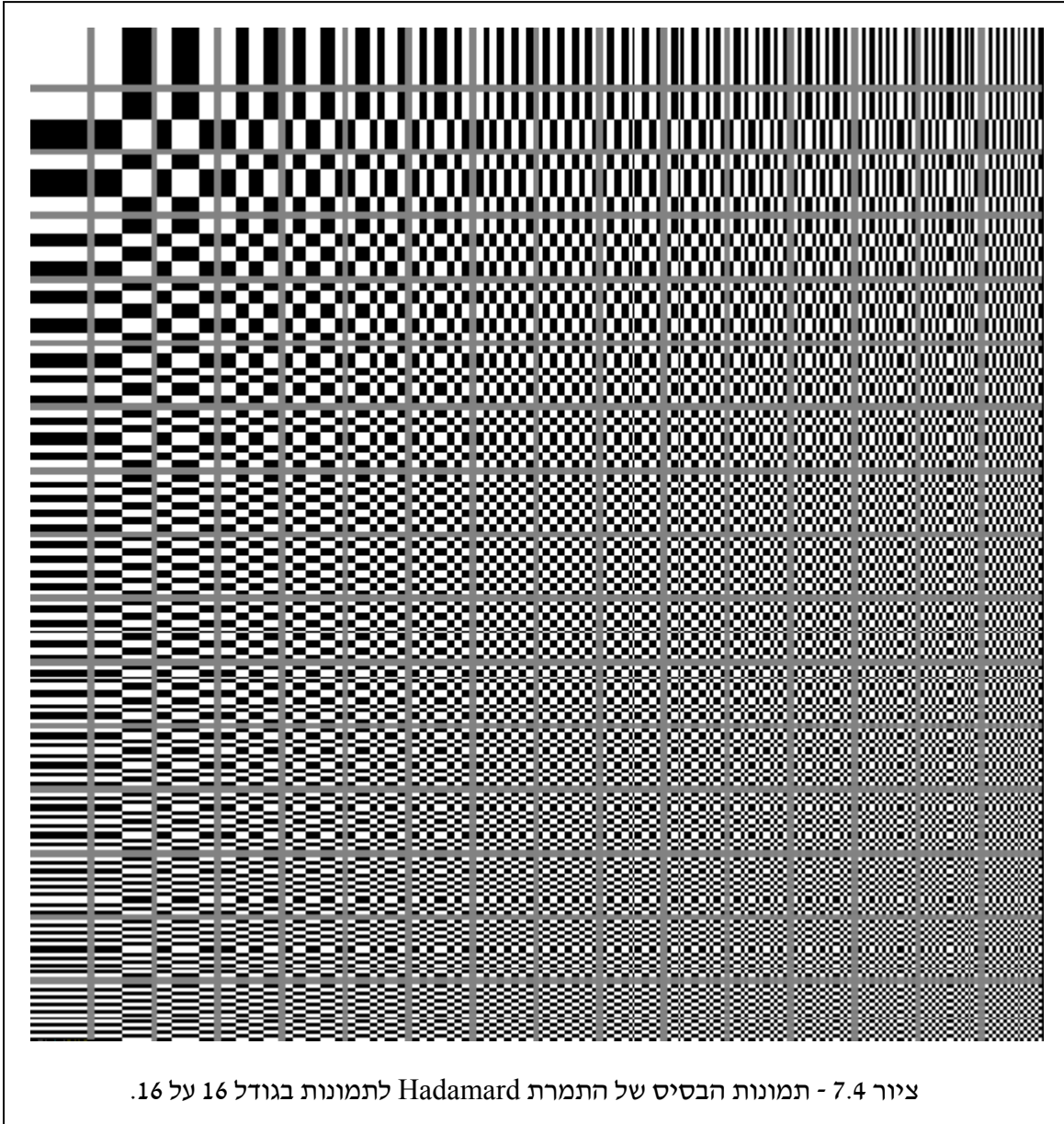
1. **ליניאריות, יוניטריות, וספרביליות:** פעולת ה- 2D-Hadamard המתוארת ע"י המטריצה $\mathbf{H}_{2D-Had.[M,N]}$ מהווה התמרה ליניארית, ריבועית, יוניטרית, ופרידה. כל תכונות אלה קלות לאבחנה מתוך כל שנאמר עד כה. יתרה מזו, התמרה דו-מימדית זו זהה בדיוק להתמרת הדמרד חד-מימדית על וקטורים באורך MN . המטריצה $\mathbf{H}_{2D-Had.[M,N]}$ סימטרית ומורכבת מאיברי $1+$ ו- $1-$ בלבד.

2. **יישום מהיר:** גם להתמרה זו יישום מהיר, בדומה למבנה FFT, אך מהיר אף יותר בשל שימוש בחיבורים בלבד.

3. **לכסון מטריצות בלוק מרקוביות:** בדומה להתמרת ה- DCT, מטריצת התמרת ההדמרד מקרבת באופן סביר ליכסון של מטריצות בעלות מבנה בלוק טופליץ עם ערך קורלציה קרוב לאחת, $\rho \rightarrow 1$.

4. **וקטורי הבסיס להתמרה:** ציור 7.4 מראה את וקטורי הבסיס של התמרת ההדמרד הדו-מימדית לתמונות בגודל 16 על 16. ניכר הדמיון הרב להתמרת ה- DCT. תכונה מעניינת היא שאם לוקחים את

בסיס ה-DCT לאחר פעולת sign (כלומר שליליים ל-1 וחיוביים ל-1+), מקבלים התמרה קרובה מאוד להתמרת הדמרד, עובדה המצביעה על הקשר והדמיון הקרוב ביניהן.



7.5 התמרת Haar ומבוא ל-Wavelet

מבין כל המטריצות היוניטריות שיכולות לשמש אותנו כהתמרה, ראינו בתחילה מטריצה בה הערכים מרוכבים (DFT), שיפרנו ע"י מציאת התמרה ממשית (DCT), והקלנו את המורכבות החישובית לאין ערוך ע"י שימוש במטריצה בה האיברים הם $1+$ או $1-$. האם ניתן להמשיך במגמת הפשטה זו? התשובה היא כן! נציע כעת התמרה אשר גם בה הערכים הם $1+$ או $1-$, אך מרבית איברי מטריצת ההתמרה יהיו אפס, ובכך נשיג התמרה הדורשת פחות חישובים. התמרה זו היא התמרת Haar. נתחיל בהצגתה של התמרה זו לחד מימד ונראה כיצד ממנה נובעת התמרה דו-מימדית.

חשוב להבהיר כי יתרונה של התמרת Haar אינו מתמצא בהיותה בעלת מטריצת התמרה דלילה (בה יש הרבה אפסים). התמרה זו מהווה מקרה פרטי פשוט של משפחה רחבה של התמרות הקרויות Wavelet, אשר נמצאו להיות יעילות בשל מגוון סיבות לטיפול באותות טבעיים. במסגרת זו נציג את רק את היסודות לתחום רחב ועשיר זה.

כמקודם, נתחיל בתיאור ההתמרה החד-מימדית לוקטור באורך M זוגי. נתחיל במטריצה בסיסית בגודל 2 על 2. מטריצה זו זהה להתמרת Hadamard:

$$D_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \Rightarrow D_1^{-1} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

התמרה זו תקיפה לוקטורים באורך 2, וניכר כי היא יוניטרית. משמעות התמרה זו היא שבהינתן שני מספרים a ו- b הבונים את הוקטור הנכנס להתמרה, ההתמרה תהיה:

$$\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} a+b \\ a-b \end{bmatrix} = \begin{bmatrix} c \\ d \end{bmatrix} \Rightarrow \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} c+d \\ c-d \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix}$$

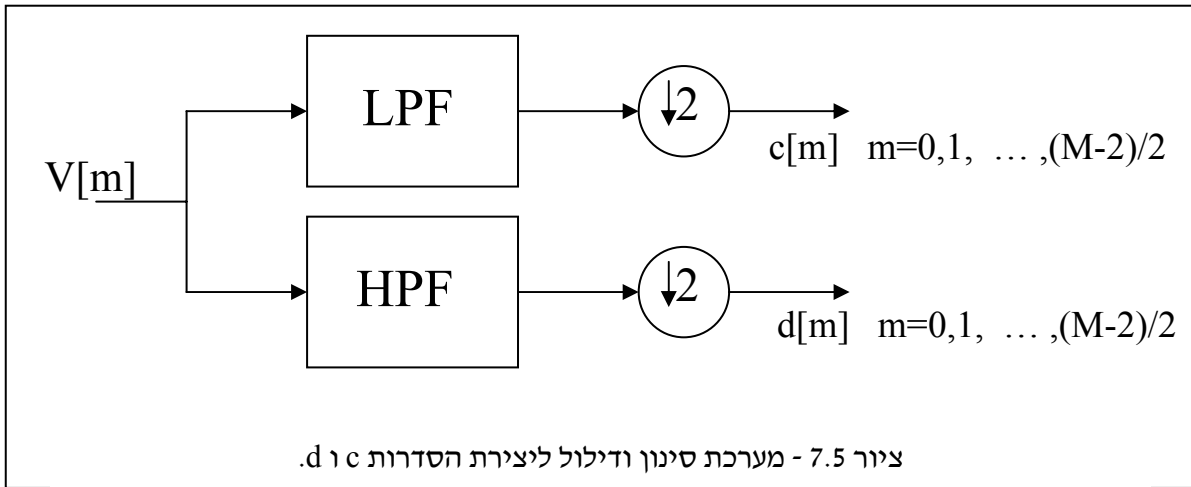
כלומר, צמד המספרים הוחלף בהתמרה בסכום ובהפרשם (ובהכפלה בקבוע הנובע מרצונו בהתמרה יוניטרית). ההתמרה ההפוכה ביצעה אותו דבר בדיוק.

בהינתן סידרה בת M (זוגי) איברים, נוכל לחלקה לזוגות סמוכים, ולהמיר כל צמד בסכום והפרשם כפי שבוצע לצמד הקודם. אם נסמן את סידרת הכניסה ב- $v[m]$, נקבל את סידרת המספרים הבאים:

$$c[\frac{m}{2}] = \frac{1}{\sqrt{2}} [v[m] + v[m+1]] \quad m = 0, 2, 4, \dots, M-2$$

$$d[\frac{m}{2}] = \frac{1}{\sqrt{2}} [v[m] - v[m+1]] \quad m = 0, 2, 4, \dots, M-2$$

וכך התקבלה סידרת מספרים בת M איברים המייצגת באופן מלא את סידרת המקור. למעשה, המערכת המתוארת בציור 7.5 שקולה לתהליך שתואר.



במערכת זו, ערוץ ה-LPF מחשב ממוצעים מקומיים לפי:

$$c[m] = \frac{1}{\sqrt{2}} [v[m] + v[m+1]] \quad m = 0, 1, 2, 3, 4, \dots, M-2, M-1$$

אך כיוון שאנו נזקקים רק לזוגיים שבהם, ישנה פעולת דילול ביחס 2, שזורקת את כל התוצאות במקומות אי-זוגיים. באופן דומה, ערוץ ה-HPF מחשב הפרשים מקומיים לפי:

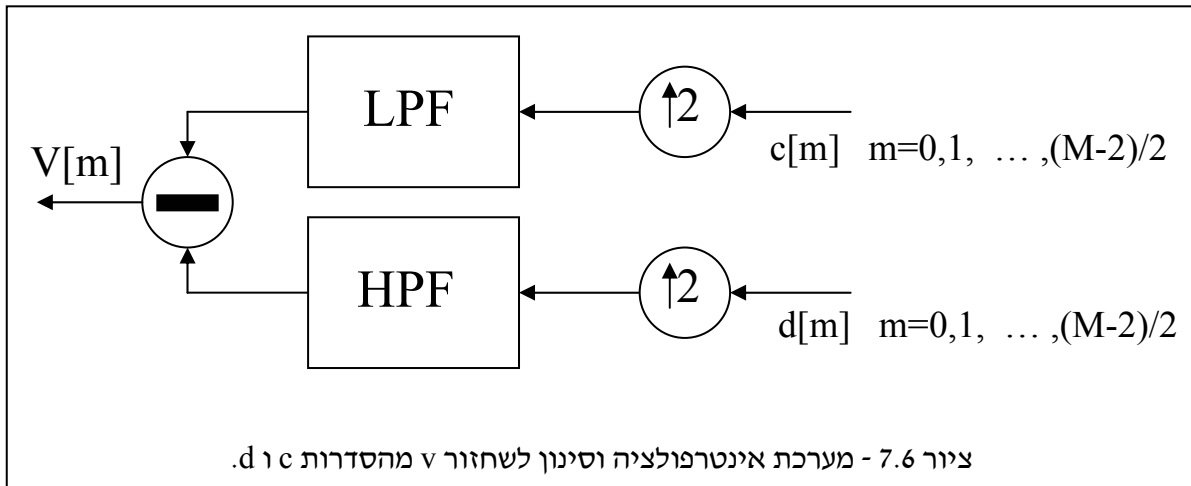
$$d[m] = \frac{1}{\sqrt{2}} [v[m] - v[m+1]] \quad m = 0, 1, 2, 3, 4, \dots, M-2, M-1$$

ושב דילול הוא זה שמסיר מחצית המספרים (מהמקומות האי-זוגיים). על מנת לקבל את הסדרות c ו-d ולחזור למקור $v[m]$, עלינו לקחת כל צמד של ערכים מתאימים, ולבצע עליהם סיכום והפרש, כמקודם, ולשלב את הערכים לפי הסדר הנכון. ניתן לראות כי סידרת הפעולות הבאה משיגה אותו אפקט:

1. קחגי את שתי הסדרות c ו-d והכנסגי אפס בין כל שתי ערכים שלהן (איבר ראשון יהיה גם 0).
2. העברגי את הסדרה המוגדלת c דרך LPF זהה לזה בו נעשה שימוש קודם.
3. העברגי את הסדרה המוגדלת d דרך HPF זהה לזה בו נעשה שימוש קודם.
4. חסרגי את התוצאות, וקבל את v.

נמחיש תהליך זה דרך דוגמה:

גם את תהליך השחזור ניתן לתאר כסכימת בלוקים הדומה לציור 7.5. סכימת השחזור מתוארת בציור 7.6.



כיוון שבכל התהליך הנ"ל בוצעו פעולות ליניאריות על איברי $v[m]$, נוכל לייצגן כמכפלת הוקטור \underline{V} במטריצה. עבור וקטור בעל ארבעה איברים, מטריצה זו תהיה הבאה:

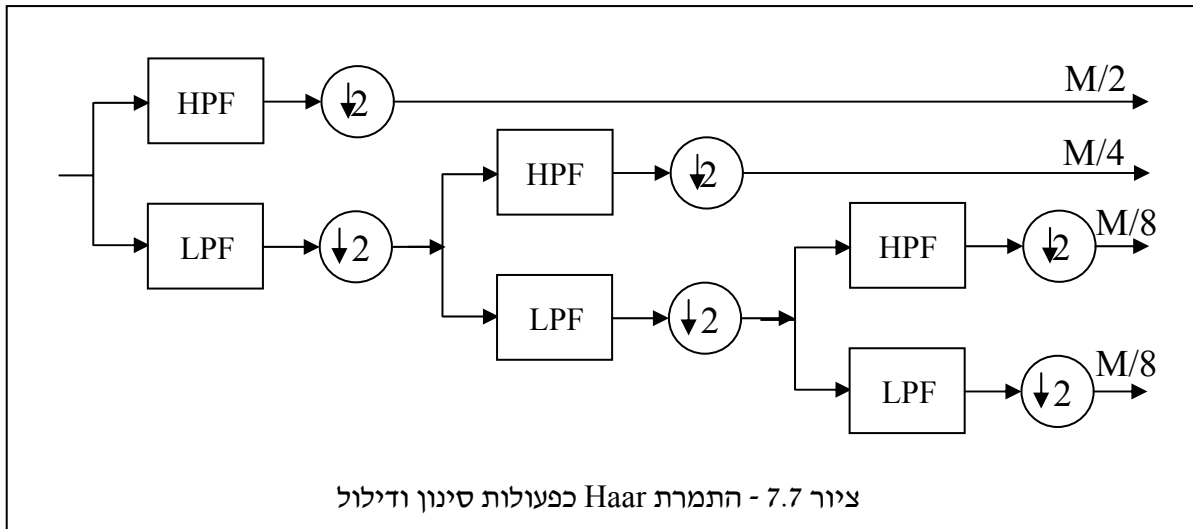
$$D_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

כך ששתי השורות הראשונות בונות את הממוצעים (LPF) ושתי השורות האחרונות בונות את ההפרשים (HPF). במטריצה זו רק חצי מהאיברים שונים מאפס. באופן דומה, לוקטור בעל שמונה איברים נקבל:

$$D_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix}$$

וקל לראות כי מטריצה זו יוניטרית, ובה רק רבע מהאיברים שונים מאפס. חשוב להבהיר - זו עדיין לא התמרת Haar.

עד כה התמקדנו במרכיב החשוב הראשון של ההתמרה - ההפרדה לסדרות LPF ו-HPF מדוללות. נציג כעת את המוטיב השני המרכזי והוא נושא המולטי-רזולוציה. הרעיון בעיקרו פשוט: את סידרת הפרשים d נותיר ללא שינוי, ועל סידרת הממוצעים c נבצע אותו תהליך כמקודם, כלומר בניית שתי סדרות של ממוצעים והפרשים עבודה. ברור כי כל סידרה כזו כבר תהיה באורך רבע מהאורך המקורי, וסך כל האיברים יסתכם ל- M . תהליך זה יחזור רקורסיבית כל פעם על סידרת ה-LPF, עד לקבלת סדרות באורך 1. התוצאה היא התמרת Haar. סכימת בלוקים של דרך בנייתה מתוארת בציור 7.7.



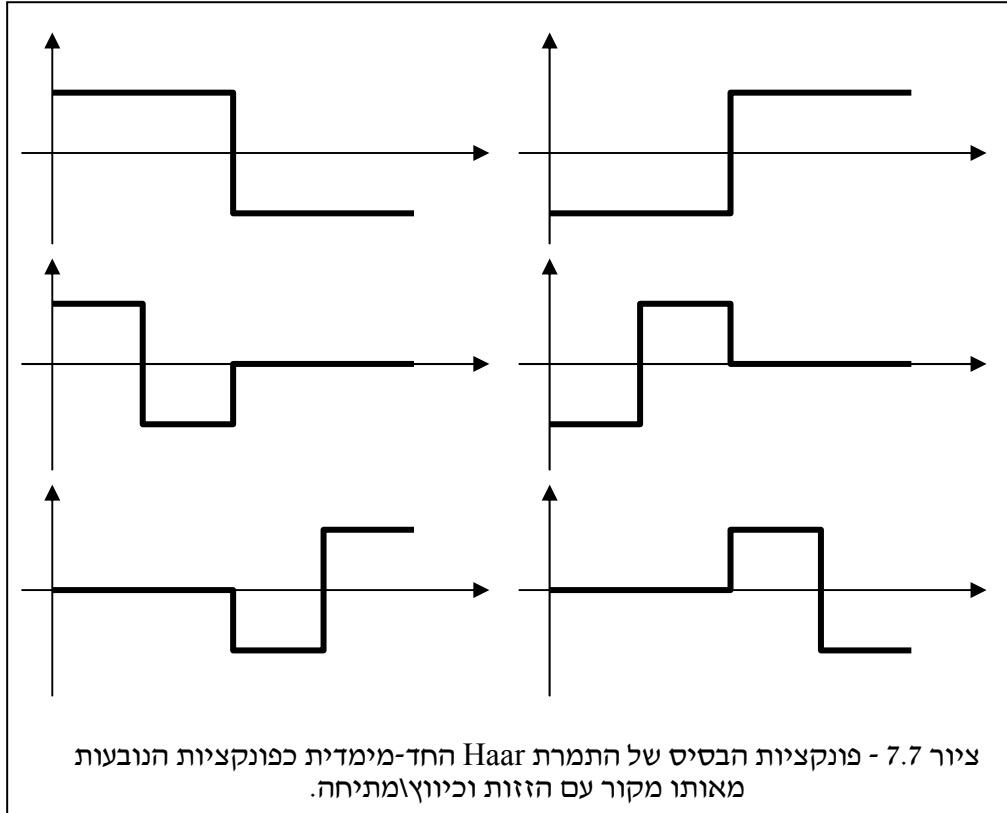
לדוגמה, התמרת וקטור בעל 4 אלמנטים יתבצע ע"י המטריצה הבאה:

$$D_2^H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 & 0 \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ \sqrt{2} & -\sqrt{2} & 0 & 0 \\ 0 & 0 & \sqrt{2} & -\sqrt{2} \end{bmatrix}$$

המטריצה הימנית במכפלה היא המטריצה שנכתבה קודם. מטריצה זו מוכפלת במטריצה שמתירה את שתי השורות התחתונות ללא שינוי, ומבצעת התמרה על שני האיברים הראשונים כפי שתואר קודם. ניתן באופן דומה לבנות את המטריצה המתייחסת להתמרת וקטור באורך 8 אלמנטים:

$$D_3^H = \frac{1}{\sqrt{8}} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ \sqrt{2} & \sqrt{2} & -\sqrt{2} & -\sqrt{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sqrt{2} & \sqrt{2} & -\sqrt{2} & -\sqrt{2} \\ 2 & -2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & -2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & -2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & -2 \end{bmatrix}$$

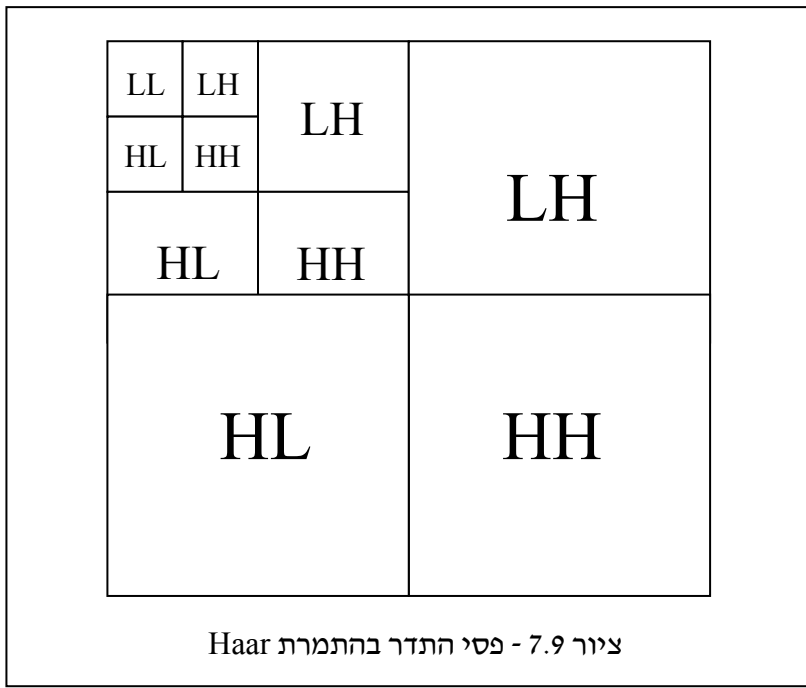
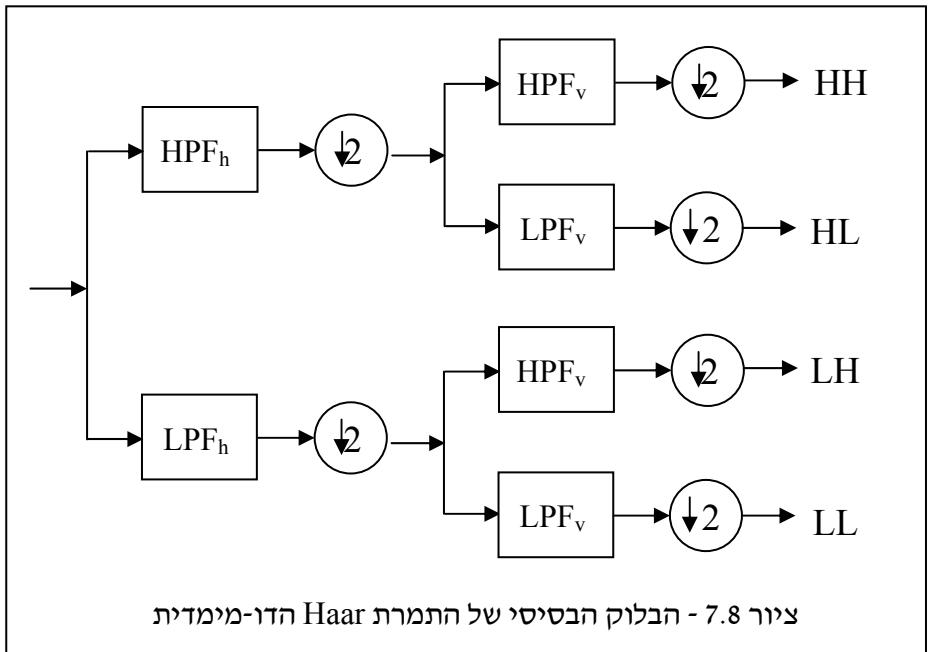
תכונה מעניינת וחשובה של מטריצת ההתמרה היא ששורותיה (למעט הראשונה) מהוות תולדה של אותו גל המעוות ע"י הזזה ומתיחה. גלים אלו נקראים "גלונים" וזהו התרגום לשם Wavelet. ציור 7.8 מראה גלים אלו כגרפים.



באשר לדו-מימד, כמו בכל המקרים הקודמים, ההתמרה על תמונה בגודל M על N מתקבלת ע"י הפעלתה באופן ספרבילי. זוהי התמרה יוניטרית קלה במיוחד הניתנת לביצוע ע"י $O\{MN\}$ פעולות (בניגוד לכל ההתמרות הקודמות הדורשות $O\{MN \cdot \log\{MN\}\}$). ביצוע ההתמרה בדרך יעילה נעשה ע"י מערכת הסינונים והדילולים שתוארה קודם לכן, עם הבלוק הבסיסי המתואר בציור 7.8.

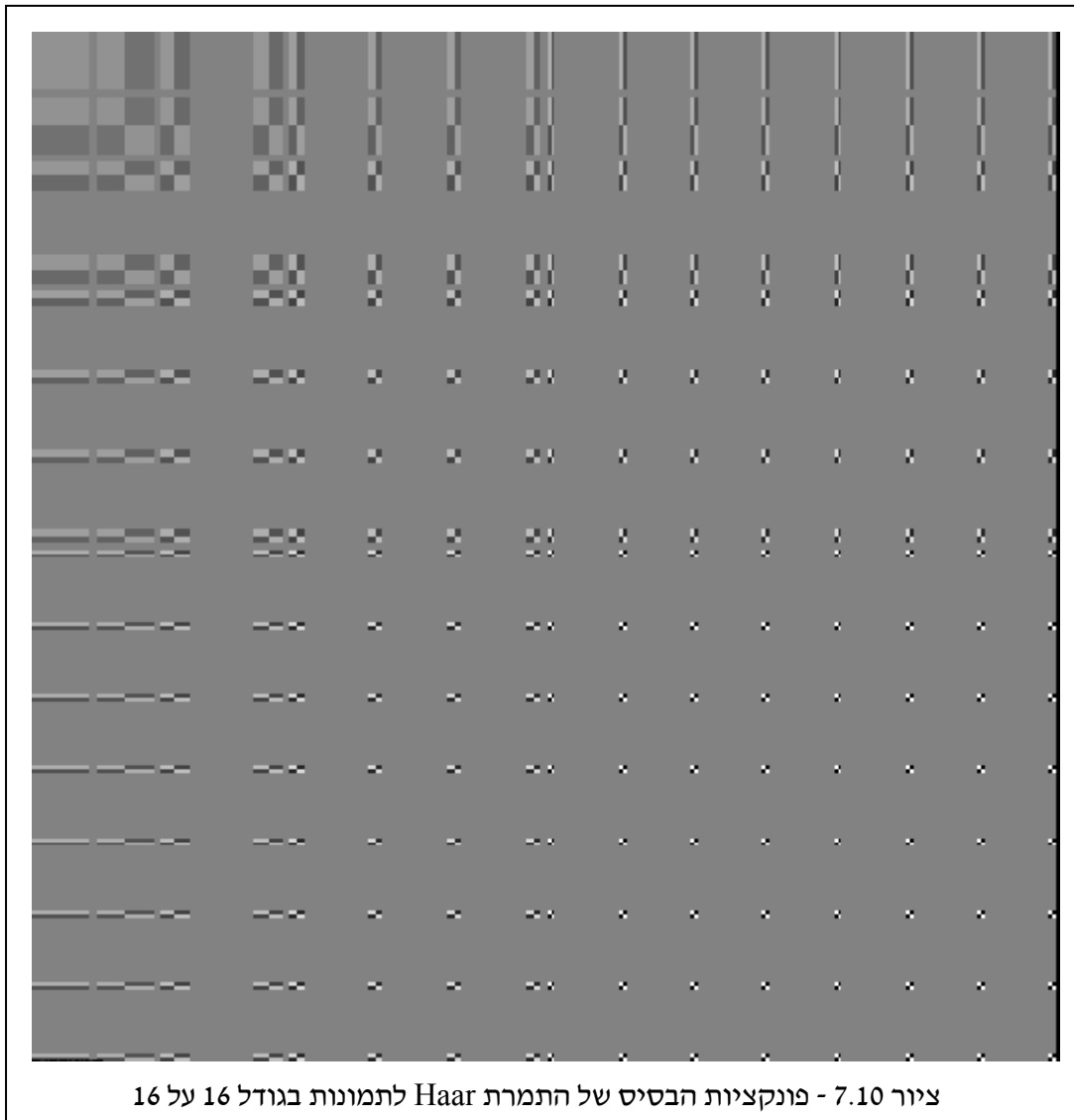
בבלוק זה, לאחר ביצוע סינון LPF ו-HPF על שורות (ולכן הסימון Horizontal-h), כל מוצא עובר גם סינון LPF ו-HPF על עמודות. אפקטיבית, מתקבל כי תחום התדר מתחלק לארבעה אזורים שווים המסומנים ב-LH, HL, LL ו-HH. באזור ה-LL המתייחס לתדרים הנמוכים מתבצעת שוב חלוקה לארבעה אזורים וחוזר שוב ושוב. ציור 7.9 מראה כיצד בסופו של דבר נראית החלוקה להתמרה בת שלושה שלבים (המתייחסת לתמונות בגודל 8 על 8).

אם נתייחס לבלוק LH הגדול, בלוק זה נבנה ע"י סינון תמונת המקור במסנן LPF אופקי ו-HPF אנכי. כיוון שכעת התוצאה ממלאת רק רבע תחום בתדר, היא מדוללת ביחס 2:1 בכל ציר (כלומר צורכת רבע מכמות הפיקסלים בתמונת המקור). לכן, לתמונה של 8 על 8, בלוק זה מוחזק במטריצה של 4 על 4. באופן דומה, בלוק ה-LH הפנימי יותר מוחזק ע"י 2 על 2, וכו'.



פונקציות הבסיס של התמרה דו-מימדית זו מתוארים בציור 7.10. ניכר כי פונקציות אלה שונות בתכלית מכל ההתמרות אותן פגשנו קודם. הייחוד בפונקציות בסיס אלה הוא עובדת היותן דלילות, כלומר תמונות בהן מעט איברים שונים מאפס. למעשה, אנו רואים כאן גלונים (Wavelet) הדומים זה לזה עד כדי מיקום וגודל.

בכל האמור עד כה התמקדנו בהתמרת Haar. התמרה זו מהווה מקרה פרטי של משפחת התמרות ה-Wavelet. בהתמרת Wavelet כללית מוחלף המסנן [1 -1] אשר שימש ל-HPF במסנן אחר, אורך



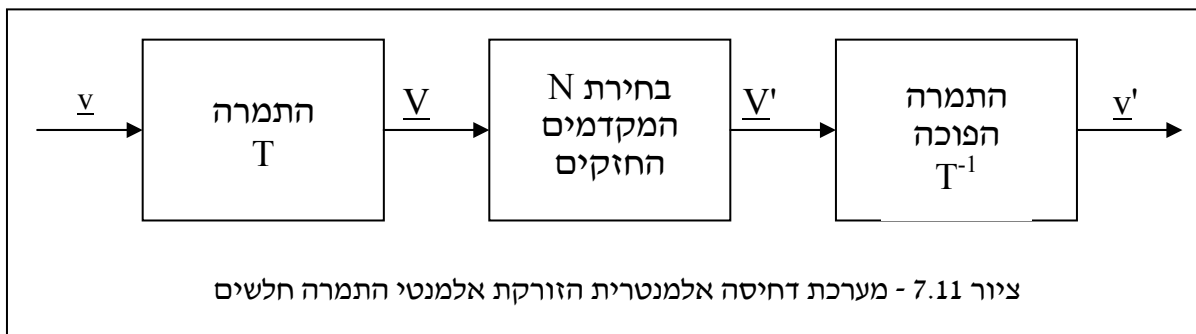
יותר בדרך כלל, וכזה שיכול לבצע הפרדה טובה יותר של פסי התדר. מסנן זה הנקרא "פונקציית האס" יוצר את כל מאפייני ההתמרה, וממנו נגזרות כל פונקציות הבסיס של ההתמרה. תכנון מסנן HPF כזה שייתן כי ההתמרה הכוללת היא אורתו-נורמלית אינה בעיה קלה כלל, וישנן פונקציות מסוימות ידועות היכולות להוות פתרון.

לסיכום, התמרת Haar שונה באופייה מההתמרות הקודמות בהן עסקנו. זוהי התמרה יוניטרית וממשית. התמרה זו ניתנת לביצוע מהיר (ב- $O\{M\}$ פעולות) בשל היותה דלילה. בנוסף לקלות ביצועה הנומרי, יש לה תכונה אטרקטיבית חשובה שלא נידונה כאן - היכולת לטפל באותות המשנים את אופייהם הספקטרילי במקום באופן נאות.

7.6 התמרות תלויות תמונה

עד כה עסקנו בהתמרות קבועות שאינן תלויות בתמונה עליה הן מבוצעות. נשאלת השאלה - האם ניתן להרוויח משהו מבניית התמרות הקשורות לתמונה בה הן מטפלות. נציג אם כך את הבעיה הבאה - נתון לנו סט של תמונות ורצוננו לתכנן את ההתמרה הטובה ביותר עבורן. ברור כי מיד עולה השאלה - מהו קריטריון הטיב בו נשתמש? במסגרת זו נציג תכונה חדשה בה לא עסקנו עד כה - ייצוג קומפקטי ע"י חלק מבסיסי ההתמרה. תכונה זו תשמש אותנו לקביעת טיבה של התמרה.

הרעיון הוא הבא - נקבל וקטור (אשר יכול להיות תמונה לאחר סידור לקסיקוגרפי) \underline{v} שאורכו M , ונבצע עליו התמרה T , לקבלת \underline{V} . לאחר ההתמרה ניקח את M איברי הוקטור \underline{V} ומתוכם נבחר את N (ברור כי $0 < N \leq M$) האיברים החזקים, ונאפס את האחרים. פעולה זו תסומן ב- $I_N \underline{X}$ כלומר, הוקטור מוכפל במטריצה בגודל M על M הדומה למטריצת יחידה I , למעט העובדה שרק N איברים ראשונים על האלכסון הם 1. את התוצאה שתסומן ב- \underline{V}' נזין להתמרה הפוכה של T , ונקבל וקטור המתיימר להיות קירוב של \underline{v} , ולכן יסומן ב- \underline{v}' . ציור 7.11 מראה תהליך זה כסכימת בלוקים. התמרה מוצלחת תיחשב התמרה שלכל N שנבחר תיתן כי \underline{v} ו- \underline{v}' קרובים ככל האפשר.



בעיה אחרת לכאורה שונה היא הבעיה הבאה - נתייחס לסכימה בציור 7.11 ונדרוש כי לאחר איפוס $M-N$ האיברים החלשים, וקטור התוצאה \underline{V}' יהיה בעל האנרגיה המירבית האפשרית. ברור כי אם T תיבחר להיות יוניטרית, אנרגיה זו תהיה קטנה מאנרגיית וקטור המקור \underline{v} , ולכן דרישה זו דומה במקצת לדרישה הקודמת כיוון ש- T תנסה להיות כזו שתרכז את האנרגיה של הוקטורים המוזנים לה במקדמים הראשונים של ההתמרה.

נציג שתי בעיות אלו באופן פורמלי - סט התמונות (למעשה וקטורים) עבורן מתוכננת ההתמרה

האופטימלית יסומן ב- $\{\underline{v}_1, \underline{v}_2, \dots, \underline{v}_L\}$. עבור N כלשהו נקבל כי:

$$\underline{V}' = I_N T \underline{v}, \quad \underline{v}' = T^{-1} I_N T \underline{v}$$

לכן, הבעיה הראשונה אותה הגדרנו תהיה:

$$\text{Minimize}_T \varepsilon_1^2\{T\} = \sum_{k=1}^L \left| T^{-1} I_N T \underline{v}_k - \underline{v}_k \right|^2$$

כשמשמעות הביטוי היא שלכל אחד מהוקטורים עבורם נרצה את ההתמרה, על הפרש בין כניסת המערכת למוצאה להיות קטן ככל האפשר. עבור הבעיה השניה שהגדרנו נקבל את בעיית האופטימיזציה הבאה:

$$\text{Maximize}_T \varepsilon_2^2\{T\} = \sum_{k=1}^L \left| I_N T \underline{v}_k \right|^2$$

כשכאן אנו דורשים אנרגיה מרבית לאחר ההתמרה והקיצוץ.

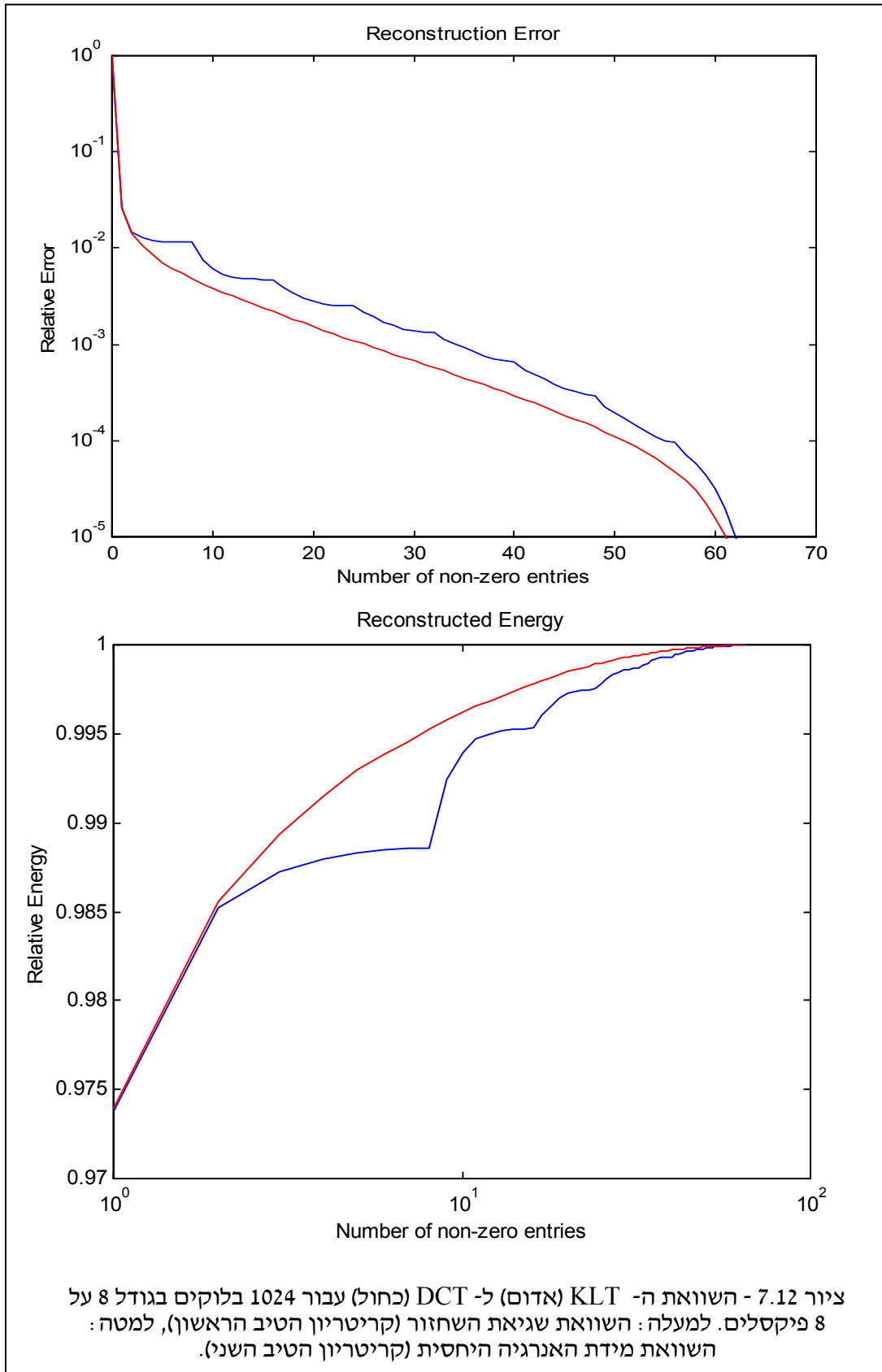
התקבלו שתי בעיות אופטימיזציה מאוד מעניינות אשר לא נפתור במסגרת זו! מסתבר כי הפתרון לשתי הבעיות זהה. ההתמרה האופטימלית בשני המובנים הנ"ל היא המטריצה היוניטרית המלכסנת את המטריצה הבאה:

$$R = \sum_{k=1}^L \underline{v}_k \underline{v}_k^H$$

כאשר יש לסדר את עמודותיה לפי סדר יורד של הערכים העצמיים. לכן, קיבלנו התמרה יוניטרית חדשה התלויה בתמונות עליהן היא מבוצעת. כאשר אנסמבל התמונות איסופי ומייצג את חוק פילוג התמונות, התמרה זו קרויה התמרת KLT - על שם של Karhunen ו- Loeve אשר הציעו אותה לראשונה. למעשה, כאשר מדובר באנסמבל כזה, המטריצה R היא מטריצת האוטוקורלציה של הוקטור האקראי \underline{v} .

האם התמרה זו מעשית? מסתבר כי במקרים רבים לא! הצורך לחשב התמרה לתמונה/ות מהווה מעמסה כבדה ומטרידה. בדחיסת תמונות, שם נעשה שימוש ניכר בהתמרות, גישה זו גם אינה אפשרית כיוון שיש לשדר למקלט גם את ההתמרה בה נעשה שימוש, וזה הורס את יחס הדחיסה. שאלה טבעית המתעוררת אם כן היא האם למשפחות מסוימות של תמונות ניתן לקרב את ה-KLT ע"י התמרה פשוטה ומוכרת? למזלנו התשובה לכך היא חיובית. לתמונות טבעיות, עבור בלוקים קטנים יחסית (עד 16 על 16 פיקסלים), התמרת ה-DCT מהווה קירוב טוב מאוד להתמרת ה-KLT. אין זו הפתעה כיוון שכבר אמרנו כי DCT מלכסן בקירוב מטריצות בלוק-מרקוביות. מסתבר כי לבלוקים קטנים זהו בדיוק המבנה שיש למטריצה R. בשל האמור לעיל, ה-DCT מהווה התמרה פופולרית מאוד בעיבוד תמונה בכלל ובדחיסת תמונות בפרט.

בציור 7.12 מובאת השוואה בין ה-DCT ל-KLT. השוואה זו מתייחסת ל-1024 בלוקים של תמונות בגודל 8 על 8 שנלקחו מהתמונה Lena. בהשוואה זו מוצגים שני קריטריוני הטיב אותם הצגנו, וניכר כי על-פי שניהם ה-KLT טוב יותר, אך גם ניכרת קירבה גדולה בין השניים.



7.7 התמרות Over-Complete

בפתיחת פרק זה העלנו את האפשרות לעשות שימוש בהתמרות ליניאריות בהן יש יותר פונקציות בסיס מאשר נקודות בתמונה. מסנן את כמות הפיקסלים (אורך וקטור הכניסה להתמרה) ב- M , ואת מספר הבסיסים כ- N , וכן, $N > M$. בסיסים מסוג זה נקראים Over-Complete. ברור כי במקרה כזה וקטורי הבסיס תלויים ליניארית, וישנה יתרות. נסמן ב- B מטריצה המכילה את וקטורי בסיס אלה כעמודותיה - למטריצה זו יש יותר עמודות משורות (היא מלבנית אופקית). בהינתן וקטור \underline{v} באורך M , התמרתו תהיה הוקטור \underline{V} באורך N . במקרה כזה הקשר המתאר את ההתמרה ההפוכה הוא:

$$\underline{v} = \sum_{k=1}^{MN} V_k b_k = B\underline{V}$$

כלומר, ביצוע ההתמרה ההפוכה פשוט למדי - איברי \underline{V} מהווים את מקדמי הצירוף הליניארי של פונקציות הבסיס לבנייתה של \underline{v} . הסוגיה המרכזית היא כיצד לקבוע את \underline{V} , כלומר לבצע את ההתמרה הישירה (בניגוד להפוכה). מסתבר כי תהליך זה מורכב יותר הפעם. אם נתבונן בקשר הנ"ל, הוא מייצג מערכת ליניארית בת M משוואות ו- N נעלמים, ולכן ישנם אינסוף אפשרויות של פתרונות \underline{V} . כיוון שיש אינסוף נעלמים, מהו הפתרון מביניהם שאותו נעדיף? נראה מספר אפשרויות ונדון בהם.

בחירה אפשרית אחת היא להעדיף את הפתרון \underline{V} שיהיה גם בעל אנרגיה מינימלית. סיבה לבחירה זו היא העדפה של ערכים נמוכים באיברי \underline{V} משיקולים שונים. במקרה כזה יהיה עלינו לפתור את בעיית האופטימיזציה הבאה:

$$P_1 : \underset{\underline{V}}{\text{Minimize}} \|\underline{V}\|^2 \quad \text{Subject To: } B\underline{V} = \underline{v}$$

פתרון בעיה זו ייעשה ע"י כופלי לגרנז' ומתקבל:

$$\begin{aligned} L\{\underline{V}, \underline{\lambda}\} &= \|\underline{V}\|^2 + \underline{\lambda}^T [B\underline{V} - \underline{v}] \\ \Rightarrow 2\underline{V} + B^T \underline{\lambda} &= 0 \Rightarrow \underline{V} = -\frac{1}{2} B^T \underline{\lambda} \\ \Rightarrow -\frac{1}{2} B B^T \underline{\lambda} &= \underline{v} \Rightarrow \underline{V} = B^T [B B^T]^{-1} \underline{v} \end{aligned}$$

בשורה ראשונה כתבנו את פונקצית הלגרנז'יאן כצירוף של פונקצית המחיר והאילווצים המוכפלים בוקטור כופלי לגרנז'. בשורה שניה גזרנו לפי \underline{V} והשוונו לאפס, וכך התקבל ביטוי לפתרון התלוי בכופלי הלגרנז'. בשורה תחתונה הצבנו את התוצאה לאילוץ לחילוצו של וקטור כופלי הלגרנז' וכך התקבל

הפתרון המלא לבעיה. התוצאה שהתקבלה היא בפועל היפוך מוכלל של B (Pseudo-Inverse). נשים לב כי ההתמרה ליניארית - כלומר בניית \underline{V} מתוך \underline{v} נעשית ע"י הכפלה במטריצה כלשהי.

אפשרות שניה לביצוע ההתמרה היא העדפת וקטור \underline{V} חלק. סיבה לבחירה זו יכולה להיות תהליך דחיסה שיבוצע עליו לאחר מכן (שיהיה אולי מוצלח יותר אם הוקטור יהיה חלק). בעיית האופטימיזציה הפעם תהיה:

$$P_2 : \underset{\underline{V}}{\text{Minimize}} \|\underline{D}\underline{V}\|^2 \quad \text{Subject To : } \underline{B}\underline{V} = \underline{v}$$

כאשר D אופרטור גזירה דיסקרטי. נניח כי D בנויה כך שאינה סינגולרית. פתרון בעיה זו ע"י כופלי לגרנז' נותן:

$$\begin{aligned} L\{\underline{V}, \underline{\lambda}\} &= \|\underline{D}\underline{V}\|^2 + \underline{\lambda}^T [\underline{B}\underline{V} - \underline{v}] \\ \Rightarrow 2\underline{D}^T \underline{D}\underline{V} + \underline{B}^T \underline{\lambda} &= 0 \Rightarrow \underline{V} = -\frac{1}{2} [\underline{D}^T \underline{D}]^{-1} \underline{B}^T \underline{\lambda} \\ \Rightarrow -\frac{1}{2} \underline{B} [\underline{D}^T \underline{D}]^{-1} \underline{B}^T \underline{\lambda} &= \underline{v} \Rightarrow \underline{V} = [\underline{D}^T \underline{D}]^{-1} \underline{B}^T \left[\underline{B} [\underline{D}^T \underline{D}]^{-1} \underline{B}^T \right]^{-1} \underline{v} \end{aligned}$$

קיבלנו הפעם הכללת התוצאה הקודמת, כשהמטריצה $\underline{D}^T \underline{D}$ מהווה מעין מטריצת משקל. גם הפעם ההתמרה ליניארית ויש מטריצה המייצגת אותה.

אפשרות אחרונה שאותה נזכיר היא האפשרות לבנות התמרה דלילה. הרעיון הוא לחפש מבין הפתרונות את הפתרון בו מרבית איברי ההתמרה הם אפס. סיבה טובה להעדיף וקטורים דלילים הינה כיוון שזהו ייצוג קומפקטי ונוח, וייצוג זה גם טוב לדחיסה כפי שניתן לראות מהדוגמה הבאה. ישנה טענה רווחת כי מערכת הראייה האנושית (במוח האדם) מבצעת התמרות מעין אלו לשם ייצוג טוב של המידע הנקלט במוח.

דוגמה 7.2: אם נניח כי לוקטורי כניסה באורך 64 אברים בני 8 סיביות כל אחד נעשה שימוש בבסיס בן 2^{20} פונקציות בסיס, ונצליח (ע"י קירוב) למפות כל וקטור כניסה לוקטור בסיס אחד ויחיד (כלומר, כל יתר מקדמי הצירוף הליניארי הם 0), הרי שלצורך שידור וקטור הכניסה נזדקק ל- 20 סיביות בלבד, במקום 512 במקור (יחס דחיסה של 0.04)

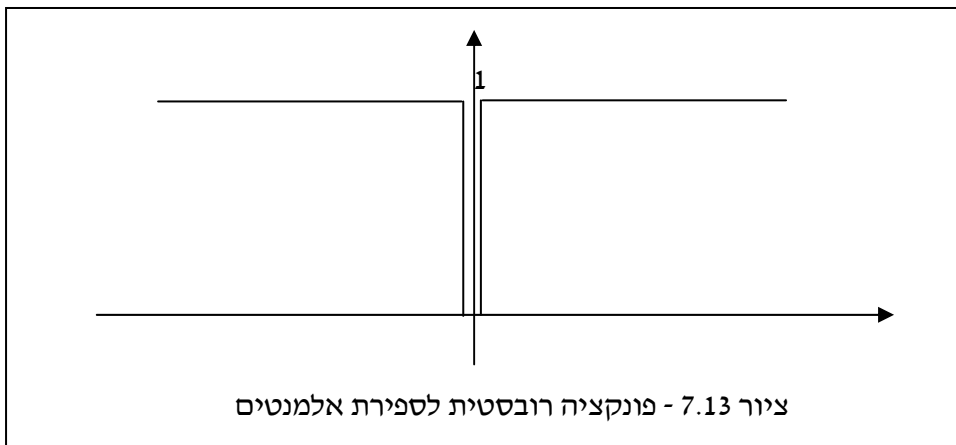
בטרם ניגש לשאלה כיצד מוצאים את ההתמרה הדלילה ביותר, חשוב להדגיש רעיון בסיס חשוב שנובע מהאמור לעיל. כשהצגנו את ה-KLT, טענו כי הוא האופטימלי במובן של ייצוג קומפקטי, אך שם הגדרת המושג "ייצוג קומפקטי" הייתה שונה בתכלית. בעוד ששם רצינו שלחלק מוקטור ההתמרה נצליח להעביר ככל האפשר יותר מהמידע הנחוץ, כאן ההגדרה מתרכזת בהתמרות מסוג Over-Complete, והדגש הוא על דלילות וקטור הייצוג. ניסיונות נומריים למצוא את ההתמרה האופטימלית במובן חדש זה

הראו כי מתקבלים בסיסים דמויי Wavelet, ותכונה זו מהווה חיזוק משמעותי לטענה כי התמרת ה-Wavelet אטרקטיבית לייצוג תמונות.

נחזור לבעיה של בניית התמרה דלילה. עלינו לכתוב פונקציה מחיר ש-"תמדוד" את מידת הדלילות של הוקטור \underline{V} . סוגיה זו מחזירה אותנו לשערוך רובסטי ולפונקציות $\rho(x)$. בהינתן וקטור \underline{V} , ברור כי הפונקציה

$$\sum_{k=1}^N \rho(\underline{V}_k)$$

תהווה אומדן טוב למידת דלילותה אם $\rho(x)$ נראית כבצורה 7.13.



לכן, בעיית האופטימיזציה המוצעת היא:

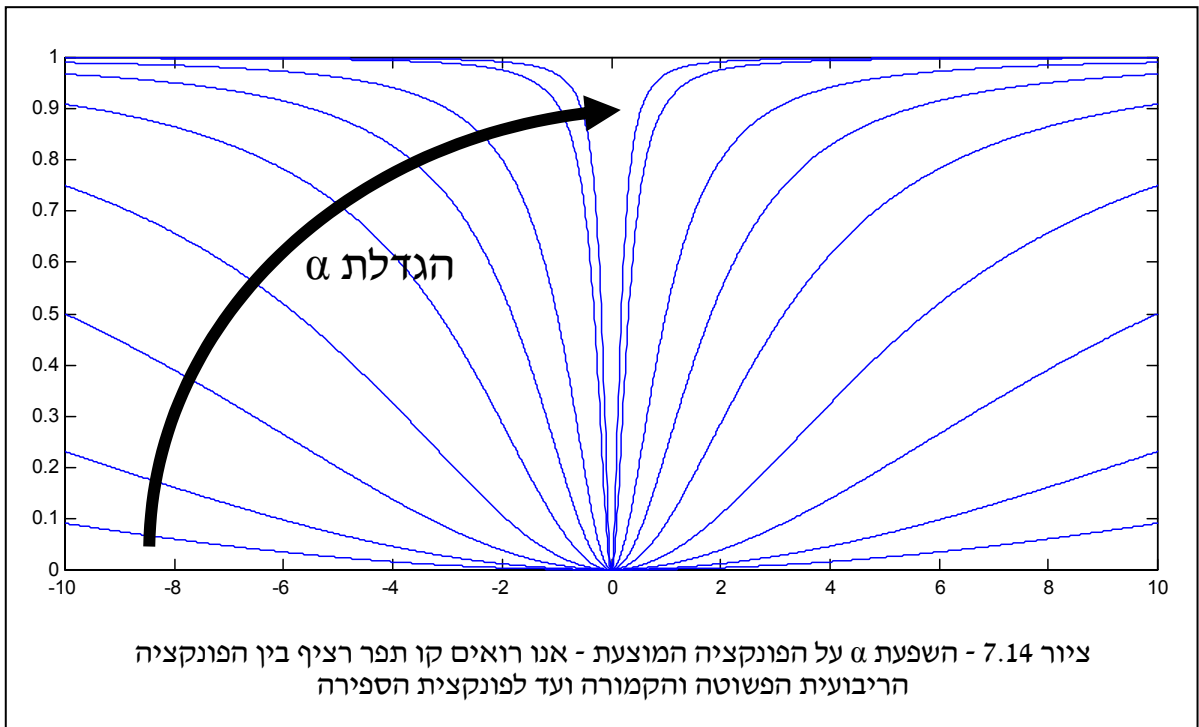
$$P_3 : \underset{\underline{V}}{\text{Minimize}} \sum_{k=1}^N \rho(\underline{V}_k) \quad \text{Subject To : } B\underline{V} = \underline{y}$$

הפונקציה המוצעת קשה מאוד לפתרון איטרטיבי בשל היותה מאוד לא קמורה. לכן, בסבירות גבוהה יתקבל כי ניפול לפתרון שיהווה נקודת מינימום לוקלית של הפונקציה הני"ל. על מנת להתגבר על בעיה זו ניתן לעשות שימוש בפונקציה $\rho(x) = |x|$. בחירה זו תהפוך את הבעיה לקמורה, ולכן לבעלת פתרון יחיד קל יחסית לקבלה, וללא נקודות מינימום לוקליות. מסתבר כי בחירה זו גם מביאה פתרון דליל וגם ניתנת לייצוג כבעיית תכנות ליניארי (זוהי משפחה של בעיות אופטימיזציה בעלות מבנה ייחודי המאפשר פתרון יעיל יחסית).

אם הבחירה הני"ל של $\rho(x)$ אינה משביעת רצון, ניתן להציע תהליך הקרוי Graduated Non Convexity (GNC) לשם פתרון הבעיה תוך ניסיון לחמוק מנקודות מינימום מקומיות. הרעיון הוא לעשות שימוש בפונקציה מהצורה:

$$\rho(x) = \frac{\alpha x^2}{1 + \alpha x^2}$$

כאשר פותרים תחילה עם α קטן מאוד. עבור אפס נקבל את הפתרון של בעיית המינימום נורמה - P_1 . לאחר הפתרון נגדיל מעט את הערך של α ונפתור שוב, וחוזר חלילה. מכיוון שהתחלנו בבעיה קמורה לאט ונעוותים אותה לפונקציה שאינה קמורה, אנו משיגים אתחול טוב בכל שלב עבור השלב הבא, ולכן יש סיכוי טוב שהפתרון יגלוש לכיוון של תוצאה דלילה מקסימלית. נשים לב כי עבור $\alpha \rightarrow \infty$ אנו משיגים פונקציה שסופרת אלמנטים שונים מאפס. ציור 7.14 ממחיש זאת.



פרק 8 – דחיסת תמונות

175	פרק 8 – דחיסת תמונות
175	8.1 יסודות בתורת האינפורמציה
181	8.2 יתרונות בתמונות וניצולן
184	8.3 קידוד אנטרופיה - אלגוריתם הופמן
188	8.4 קידוד עם שגיאות - גישת החיזוי
191	8.5 קידוד עם שגיאות - גישת ההתמרה

בפרק זה אנו נכנסים לתחום של דחיסת תמונות. כבר בפרק ראשון עמדנו על הבעייתיות של נפחן האופייני של תמונות. תמונה ממוצעת בגודל של 1000 על 1000 פיקסלים בצבע צורכת נפח של 3MB. הרעיון המרכזי מאחורי מגוון השיטות והאלגוריתמים שיידונו בפרק זה הוא שנפח זה אינו מחויב המציאות. בדחיסה אנו מבחינים בין דחיסה ללא אובדן, בה לאחר דחיסה ופריסה מתקבלת אותה תמונה בדיוק, ובין דחיסה אובדנית, בה מתירים שגיאות קלות לשם השגת רווח ביחס הדחיסה. כפי שנראה במסגרת פרק זה, ניתן בפחות מהנפח המלא לאחסן את אותה תמונה בדיוק, וניתן בהרבה פחות לאחסן תמונה דומה לה מאוד. השאלה היא כמובן איך עושים זאת, ועל שאלה זו ננסה לענות במסגרת פרק זה.

8.1 יסודות בתורת האינפורמציה

תורת האינפורמציה הינה תורה חדשה יחסית אשר פותחה לראשונה ע"י Shannon בשנות הארבעים של מאה זו. בבסיסה, תורה זו באה לכמת מידע, ולאפיין מערכות בהקשר של העברת מידע. תורה זו מהווה את התשתית התיאורטית לנושא דחיסת מידע בכלל ודחיסת תמונות בפרט. ברור כי במסגרת זו לא ניתן יותר מסקירה ראשונית לא ממצה בשל נפחו העצום של תחום זה וריבוי התוצאות בו.

נתחיל עם מושג האנטרופיה - מידת אי-ודאותו של מקור מידע. נתון לנו מקור המייצר מספרים אקראיים דיסקרטיים x חסרי תלות ביניהם (i.i.d) עם חוק פילוג $P(x)$. האנטרופיה של מקור זה מוגדרת ע"י הביטוי:

$$H\{x\} = \sum_x P(x) \log_2 \frac{1}{P(x)} \left[\frac{\text{bits}}{\text{symbol}} \right]$$

משמעות גודל זה היא כמות הסיביות התיאורטית הנחוצה לייצוג מדויק של תו ממקור זה במוצע. נראה מספר דוגמאות להמחשת מידה זו.

דוגמה 8.1: מקור שמייצר מספרים שלמים בתחום [1,16] בהסתברות זהה - 1/16 יהיה בעל אנטרופיה

$$H(x) = 16 \cdot \frac{1}{16} \log_2 16 = 4 \text{ bits/symbol}$$

תוצאה זו אינטואיטיבית לנו כיוון שעבור 16 אפשרויות, ניתן לייצג כל אחת ב-4 סיביות, כפי שהתקבל בפועל.

דוגמה 8.2: עבור אותו מקור כשהפעם ההסתברות נתונה כ- $P(x) = x/136$ (המספר 136 מנרמל את ההסתברות לקבלת סכום הסתברויות שווה 1), האנטרופיה תהיה:

$$H(x) = \frac{1}{136} \left[1 \cdot \log_2 \frac{136}{1} + 2 \cdot \log_2 \frac{136}{2} + \dots + 16 \cdot \log_2 \frac{136}{16} \right] \cong 3.76 \text{ bits/symbol}$$

תוצאה זו מעניינת כיוון שעדיין יש לנו 16 אפשרויות, אך לטענת התיאוריה, די ב-3.76 סיביות לתיאורן באופן מלא. ניתן למצוא בדבר הגיון אם נחשוב על הדברים באופן הבא - כאשר כל ההסתברויות זהות, אי הודאות היא מירבית, כלומר, אין לנו כל דרך לדעת משהו בטרום קיבלנו את התו. עבור הסתברות לא אחידה, עוד בטרום נקבל את התו נוכל לומר כי ישנה נטייה חזקה לכיוון הערכים הגדולים, בהיותם בעלי הסתברות גבוהה יותר. לכן, מידת אי-הודאות המתקבלת נמוכה יותר.

דוגמה 8.3: נניח כי ל-16 התווים האפשריים, הסתברותו של 1 היא 1, ושל כל היתר 0. במקרה כזה האנטרופיה תהיה:

$$H(x) = 1 \cdot \log_2 1 + 15 \cdot 0 \log_2 0 = 0 \text{ bits/symbol}$$

אין כאן כל הפתעה: במקרה זה, לפני שנקבל את התו נדע בדיוק מה יקרה, ולכן אין כל אי-ודאות.

למעשה, דרך דוגמה זו המחשנו שתי תכונות חשובות של פונקציית האנטרופיה:

1. למקור בעל N תווים, האנטרופיה H חסומה מלמעלה לפי $H \leq \log_2 N$. שיוון מתקיים כאשר ההסתברות זהה ושווה 1/N לכל תו.

2. למקור בעל N תווים, האנטרופיה H חסומה מלמטה לפי $-H \geq 0$. שיויון מתקיים כאשר תו אחד בעל הסתברות 1, וכל יתר התווים בעלי הסתברות 0.

משפט הצפינה של Shannon קובע כי עבור מקור בעל אנטרופיה $H(x)$, לא ניתן לדחוס מידע ממנו בפחות מהאנטרופיה ללא הכנסת שגיאות. יתרה מזו: לכל ε חיובי וקטן כרצוננו, קיים **בעקרון** אלגוריתם דחיסה אשר ייתן דחיסת המקור לקצב של $H(x) + \varepsilon$. תוצאה זו קובעת כי האנטרופיה מהווה אומדן לביצועיו של אלגוריתם דחיסה, אם רצוננו לדחוס ללא שגיאות. משפט זה של Shannon לא מציע את אלגוריתם הדחיסה עצמו, אלא מוכיח את קיומו באופן תיאורטי. בהמשך נכיר דרכים מעשיות למימוש טענה זו.

חשוב להבהיר כי מושג האנטרופיה אינו מוגבל לטיפול בתווים בודדים. לדוגמה, עבור מקור המייצר מספרים בתחום $[1, 16]$, ניתן להתייחס לצמדי תווים כ-"סופר-תווים" חדשים, ואז יש לנו 256 סוגי תווים אפשריים, וניתן לכל צירוף לקבוע את הסתברותו, לחשב אנטרופיה כמקודם וכו'. תהליך זה חשוב ונחוץ כאשר יש קשר סטטיסטי בין התווים. נמחיש זאת דרך דוגמה.

דוגמה 8.4: עבור המקור שתואר בדוגמה 8.1, לקיחת צמדי תווים תביא לכך שיהיו 256 תווים אפשריים שווי הסתברות. האנטרופיה תהיה:

$$H([x_1, x_2]) = 256 \cdot \frac{1}{256} \log_2 256 = 8 \text{ bits/symbol}$$

נזכור כי הפעם תו אחד הוא בעצם שני תווים, ולכן אין כל שינוי ביחס לטיפול בתווים בודדים.

דוגמה 8.5: נתון מקור המייצר את המספרים $1+$ ו- $1-$ לפי חוק הפילוג הבא:

$$P(x_k / x_{k-1}) = \begin{cases} 0.8 & x_k = x_{k-1} \\ 0.2 & x_k \neq x_{k-1} \end{cases}$$

כלומר, לדוגמה, אם התו הקודם היה $1-$, סבירות גבוהה יש להיותו של התו הבא $1-$ גם כן. בטיפול בתווים בודדים נקבל כי $P(-1) = P(+1) = 0.5$ כיוון ש:

$$\begin{aligned} P(+1) &= P(+1/+1)P(+1) + P(+1/-1)P(-1) \\ P(-1) &= P(-1/+1)P(+1) + P(-1/-1)P(-1) \\ \Rightarrow \begin{bmatrix} P(+1/+1) - 1 & P(+1/-1) \\ P(-1/+1) & P(-1/-1) - 1 \end{bmatrix} \begin{bmatrix} P(+1) \\ P(-1) \end{bmatrix} &= \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ \Rightarrow P(+1) = P(-1) &= 0.5 \end{aligned}$$

לכן האנטרופיה תהיה:

$$H(x) = 0.5 \log_2 2 + 0.5 \log_2 2 = 1 \text{ bits/symbol}$$

עבור טיפול בצמדי תווים נקבל כי חוק ההסתברות הוא $P([+1,+1]) = P([-1,-1]) = 0.4$ ו-
 $P([-1,+1]) = P(+1,-1) = 0.1$ כיוון ש:

$$P([x_k, x_{k-1}]) = P(x_k / x_{k-1})P(x_{k-1}) = \begin{cases} 0.8 \cdot 0.5 & [+1,+1] \\ 0.2 \cdot 0.5 & [+1,-1] \\ 0.2 \cdot 0.5 & [-1,+1] \\ 0.8 \cdot 0.5 & [-1,-1] \end{cases}$$

במקרה זה האנטרופיה תהיה:

$$H([x_1, x_2]) = 2 \cdot 0.4 \log_2 2.5 + 2 \cdot 0.1 \log_2 10 = 1.7219 \text{ bits/symbol}$$

כך שלתו יחיד נדרשים לפחות מסיבית אחת, וברור כי הושגה יעילות גבוהה יותר כי נוצל הקשר בין התווים. למעשה, ניתן להמשיך במגמה זו ולטפל בשלשות, רביעיות וכו', כשצפוי שיפור ככל שעובדים עם בלוקים גדולים יותר.

למעשה, יכולנו לנצל את הקשר הסטטיסטי בין התווים הסמוכים בצורה יעילה יותר לו פעלנו על ההסתברויות המותנות בצורה ישירה. נניח כי אנו נמצאים בתו ה-L וכבר קיבלנו אותו במפענת, ונניח כי הוא היה +1. לכן, בבואנו לדחוס (ובמקלט - לפרוס) את התו הבא, אנו יודעים כי בסבירות גבוהה הערך הבא הוא +1 גם כן. למעשה, האנטרופיה של התו הבא תהיה:

$$H(X_{L+1}) = 0.8 \log_2 \frac{1}{0.8} + 0.2 \log_2 \frac{1}{0.2} = 0.7219$$

כך שקיבלנו תוצאה טובה יותר מזו של העבודה בצמדים. גישה זו אינה גישה עבודה בבלוקים אלא גישה חלופית של חיזוי - על פי התו הנוכחי שכבר שודר (ונקלט), נדחס התו הבא, כשעושים שימוש בהסתברות המותנית.

למעשה, לו קידדנו את המקור הזה בבלוקים שאורכם שואף לאינסוף היינו מקבלים תוצאה זו בדיוק. כאשר אנו עובדים בצמד תווים אנו מנצלים את הקשר ביניהם, אך לא את הקשר לצמד הקודם או לצמד הבא, ולכן התוצאה פחות טובה. כאשר הבלוקים מתארכים אנו מצליחים טוב יותר לנצל את הקשרים בין התווים, כשאנו מאבדים רק את הקשרים בתפרים בין הבלוקים. אם הבלוקים באורך אינסופי - אין אובדן. ברור כי הגישה המותנית יעילה יותר לאין-ערוד.

נציג עוד מספר תכונות המאפיינות את פונקציית האנטרופיה, הנוגעות לטיפול בבלוקים:

$$1. \text{ קיים אי-השוויון הבא: } H([x_1, x_2, \dots, x_N]) \leq \sum_{k=1}^N H(x_k)$$

כשוויון מתקיים עבור חוק פילוג פריד - $P([x_1, x_2, \dots, x_N]) = P(x_1)P(x_2) \dots P(x_N)$. משמעות תכונה זו היא שתמיד יש רווח או לפחות אי-הפסד בעבודה על בלוקים במקום על תווים בודדים.

$$2. \text{ קיים אי-השוויון הבא: } H([x_1, x_2, \dots, x_N]) \geq \max_k \{H(x_k)\}$$

כשוויון מתקיים עבור מקרה בו כל המשתנים נקבעים באופן דטרמיניסטי מתוך אחד מהם. פירוש הדבר הוא שכאשר הינו בלוק של תווים אך למעשה רק אחד מהם אקראי וכל היתר נגזרים ממנו בצורה מדויקת, די לקודד תו אקראי זה. במקרה כללי, אם נחשב אנטרופיה לכל אחד מהתווים בנפרד, לא יתכן שהאנטרופיה המשותפת תהיה קטנה מהגרוע שבהם (גרוע במובן של מכיל הרבה אינפורמציה).

$$3. \text{ קיים אי-השוויון הבא: } H([x_1, x_2]) \geq 2 \cdot H(x_1/x_2)$$

כשוויון מתקיים עבור מקרה בו אין קשר סטטיסטי בין התווים.

עד כה עסקנו בהגדרת מידת האינפורמציה של מקור כאשר רצוננו לדחוסו ללא שגיאות. מה קורה כאשר אנו מוכנים להתיר שגיאות? מסתבר שגם ליישום זה הציע Shannon אומדן אינפורמציה, הידוע בכינויו פונקציה קצב-העיוות $R(D)$ - Rate-Distortion Function. הרעיון מכליל בצורה ישירה את מושג האנטרופיה ממקודם באופן הבא. נניח כי אנו מוכנים לקבל כי ייווצר מרחק של עד D בין סידרת המקור x והסדרה הנפרסת לאחר דחיסה y . פער זה מסומן ב- $d\{x, y\} \leq D$ וברור כי עלינו להגדיר מהי המטריקה בה אנו עושים שימוש. אזי, מידת האנטרופיה של המקור x תחת שגיאה של עד D תהיה $R(D)$. נציג מספר תכונות של פונקציה זו:

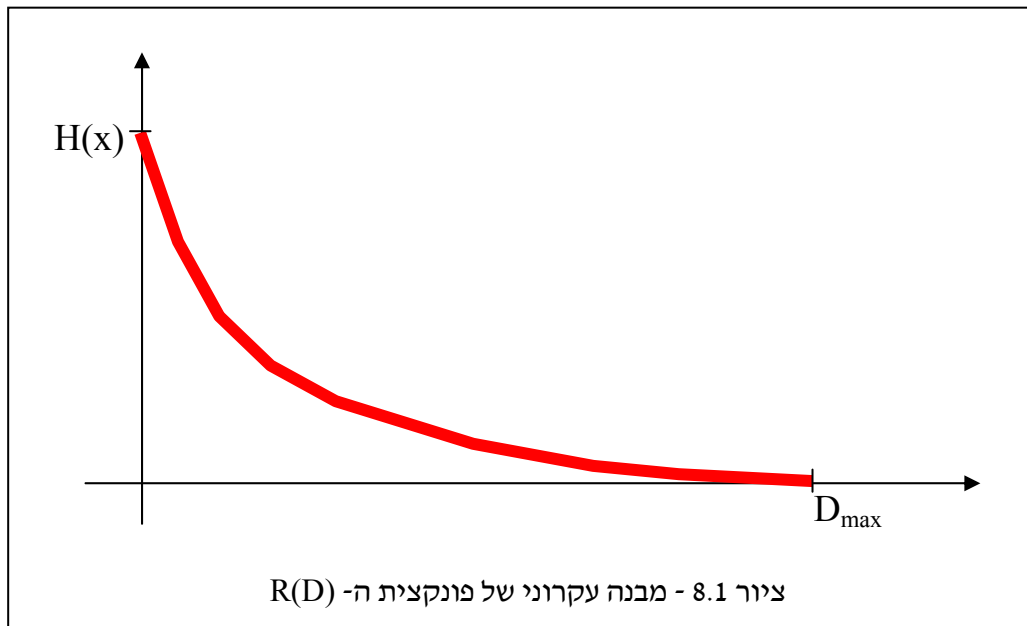
$$1. \text{ עבור } D=0 \text{ אנו בפועל לא מרשים שגיאה ולכן ברור כי } R(0) = H(x)$$

2. עבור D מעל סף מסוים התלוי בעוצמתו של x יתקבל כי אם פשוט נשלח את ממוצע המספרים לכל תו כניסה תתקבל שגיאה קטנה מ- D . במקרה זה נקבל כי $R(D) = 0$.

3. פונקציה ה- $R(D)$ מונוטונית לא עולה. עליה לא תיתכן כיוון שאז לשגיאה גדולה יותר נדמה כי נדרש קצב גבוה יותר - זה לא יתכן.

4. פונקציה ה- $R(D)$ היא פונקציה קמורה.

ציור 7.1 מראה כיצד נראית פונקציה ה- $R(D)$ באופן עקרוני.



חשוב להבהיר כי בעוד שאנטרופיה של מקור קלה לחישוב, חישובה של $R(D)$ (באופן אנליטי או נומרי) קשה מאוד, ולרוב בלתי אפשרי. חישובה של האנטרופיה דורש את חוק פילוגם של תווי המקור, חישוב ה- $R(D)$ דורש בנוסף את הגדרת המטריקה בין אותות הכניסה והיציאה ממערכת הדחיסה/פריסה.

דוגמה 8.6: נניח כי המקור שלנו מייצר מספרים שלמים בתחום $[0,127]$ בהסתברות אחידה. לכן האנטרופיה של מקור זה היא 7 סיביות לאות. כעת נניח כי אנו מתירים לקידוד לבצע סטייה של 1, כלומר,

$$D(x, y) = (x - y)^2 \leq 1$$

במילים אחרות, אנו מתירים לסטות ברמה אחת במקלט בהשוואה למשדר. נציע הצעה פשוטה שתגשים שגיאה זו תוך שידור פחות סיביות - ניקח את המספר ונחליף אותו במספר הזוגי הקרוב אליו מלמטה. כך יקרה כי מספרים זוגיים ייוותרו כמות שהם, ומספרים אי-זוגיים יומרו למספרים זוגיים. כיוון שכעת קיימים רק מספרים זוגיים, נדרשים 6 סיביות במקום 7. בתהליך דומה, עבור שגיאה של 2 נוכל לבחור את השכן הקרוב ביותר מבין המספרים המתחלקים ב-3 (0,3,6,9, ..., 126). כעת יש לנו רק 43 ערכים אפשריים שווי הסתברות ולכן האנטרופיה היא 5.246.

נוכל להמשיך תהליך זה ולהציע מקודד עם שגיאה למספרים אלו, ובעצם לקבל פונקציה המקרבת את ה- $R(D)$. הפונקציה המוצעת היא פונקציה החוסמת מלמעלה את פונקצית קצב-העיוות כיוון שאין כל הבטחה כי לא קיימת אסטרטגיה טובה יותר לשידור עם שגיאות במקרה המוצע עם פחות אנטרופיה.

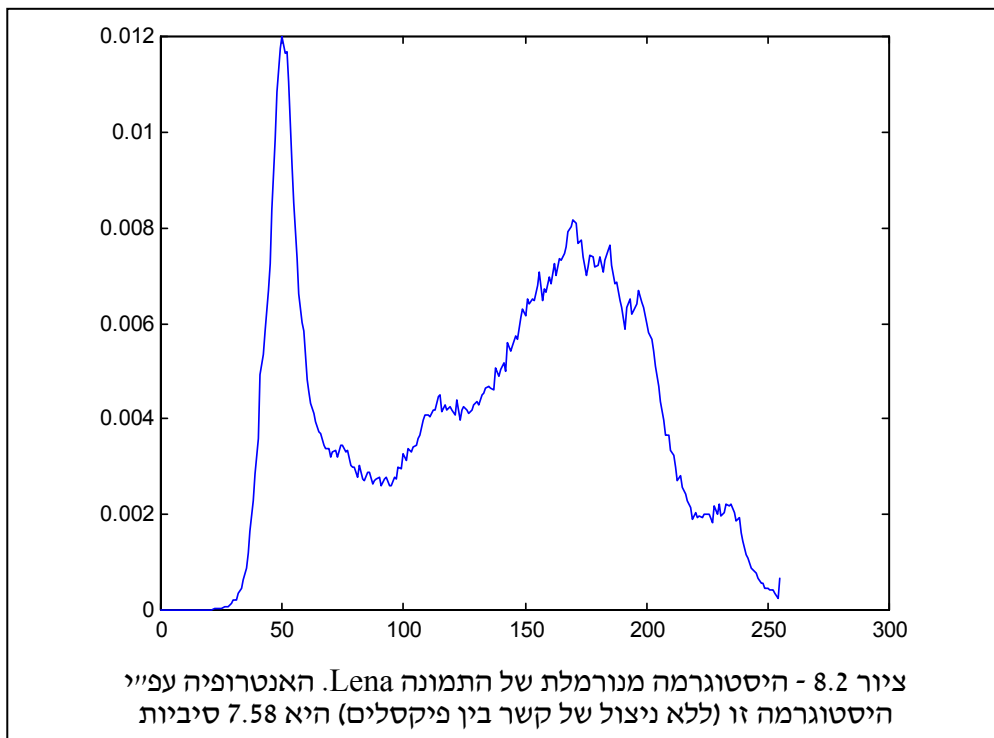
8.2 יתירויות בתמונות וניצולן

מהדיון בנושא האנטרופיה ופונקצית קצב-העיוות ברור כי דחיסה מצליחה בשל יתירות באות הנדחס. יתירות פירושה שהמידע מכיל כפילויות מיותרות הניתנות לניצול. עד כה ראינו מספר סוגי יתירויות עקרוניים. נזכיר אותם ונראה כיצד הם ישימים בתמונות:

1. יתירות הסתברותית - אפילו אם המקור שלנו מניב תווים מתוך סט נתון של N תווים אפשריים ובלי תלות ביניהם, אם לתווים שונים הסתברות שונה נקבל כי האנטרופיה נמוכה מ- $\log_2 N$.
אם נחזור לרגע לתמונות, ונזכור כי היסטוגרמה של תמונה היא פונקצית ההסתברות של רמות האפור בה, הרי שאם ההיסטוגרמה אינה אחידה, ישנה כאן יתירות הסתברותית הניתנת לניצול לשם קבלת דחיסה. למעשה, משיקולים אלו ברור כי ככל שההיסטוגרמה רחוקה מלהיות אחידה, כך התמונה דחיסה יותר.

דוגמה 8.7: ניקח את התמונה Lena בה יש 512 על 512 פיקסלים המיוצגים ע"י 8 סיביות כל אחד. עפ"י ההיסטוגרמה המנורמלת שלה נחשב את האנטרופיה. ההיסטוגרמה מתוארת בציור 8.2, והאנטרופיה המתקבלת היא:

$$H_{Lena}(\text{pixel}) = \sum_{k=0}^{255} h(k) \log_2 \frac{1}{h(k)} = 7.58 \text{ bits/pixel}$$



כלומר, רק מתוך ההיסטוגרמה ברור כי ניתן בפחות מ- 8 סיביות לייצג כל פיקסל. כמו כן אנו רואים כי הרווח אינו גדול - נוכל (תיאורטית בשלב זה) לחסוך כחצי סיבית לפיקסל.

2. יתירות בין תווים - ראינו כי כאשר עובדים על בלוקים של תווים וקיימת תלות סטטיסטית ביניהם, מתקבלת יעילות טובה יותר של הדחיסה. בהשאלה לתמונות נאמר כי כאשר אנו עומדים לקודד את הפיקסל ה- $[i,j]$ בתמונה, אם נצל את המידע שמספקים שכניו של פיקסל זה, נצליח טוב יותר בדחיסה. זאת כיוון שקיימת תלות בין פיקסלים סמוכים בתמונה - הם אמורים להיות בעלי רמות אפור קרובות, ולכן קיימת תלות סטטיסטית חזקה ביניהם.

דוגמה 8.8: נחזור לתמונה Lena. על מנת לעבוד בבלוקים נדרש לבנות היסטוגרמה של צמדים. היסטוגרמה זו תדרוש כ- 65000 תאים (256 בריבוע), וברור כי מלאכת בניית היסטוגרמה כזו לא תהיה נעימה. כמו כן ברור כי לבלוקים גדולים יותר נקבל היסטוגרמות בנפחים בלתי-אפשריים. מעבר לכך, אפילו אם נתעלם מהנפחים הנוצרים, הרי שבתמונה זו יש 512^2 פיקסלים, כלומר, בממוצע - רק פי ארבע יותר צמדים מכמות התאים לעבודה על צמדים. לכן, אין מספיק מידע בתמונה לשם בניית פונקציית הסתברות דו-מימדית.

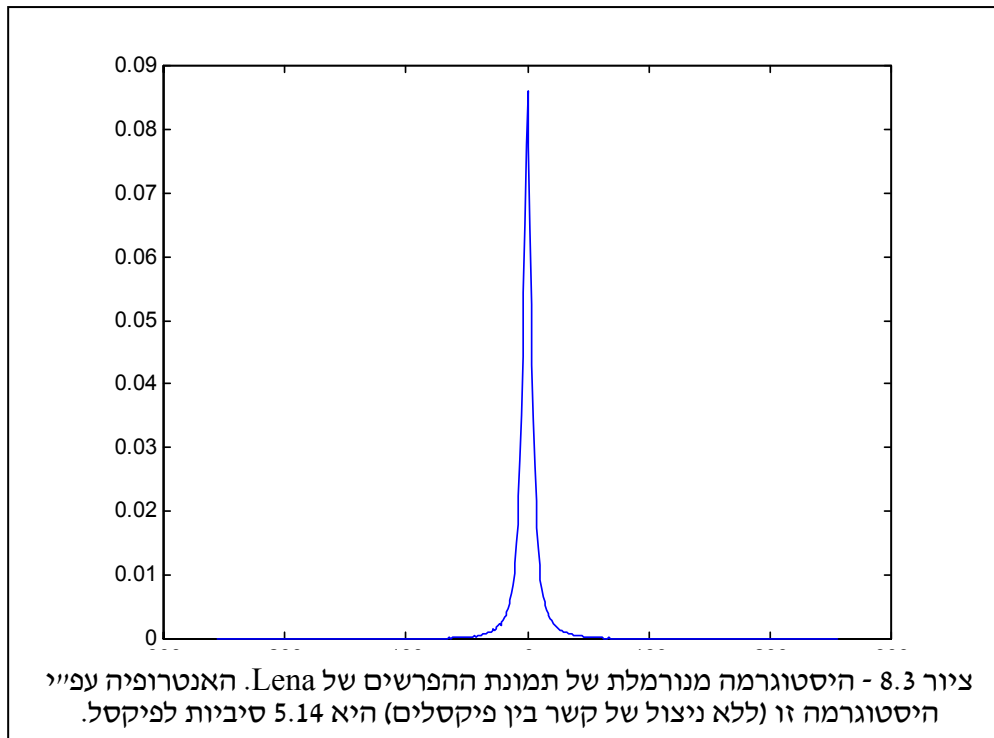
כתחליף לעבודה בבלוקים, ראינו כי ניתן להציע עבודה עם חיזוי. נמחיש רעיון זה על דחיסת התמונה הנתונה. נסדר את כל הפיקסלים בתמונה בעמודה בסידור לקסיקוגרפי, ונחליט כי במקום לשדר את הפיקסל $X[k]$, נשדר את ההפרש שלו משכנו הקודם - $X[k-1]$. במקלט כבר נקלט מידע על-פיו יודעים את $X[k-1]$, ולכן בהינתן ההפרש - $E[k]=X[k]-X[k-1]$, ניתן יהיה לקבוע את $X[k]$ עצמו. נתבונן אם כך בהיסטוגרמה של הוקטור $E[k]$ - היסטוגרמה זו נתונה בציור 8.3. ברור כי ערכי ההפרש נעים בתחום -255 ועד $+255$, כך שהתחום הדינאמי גדל. לעומת זאת, עיון בהיסטוגרמה זו מגלה כי רוב השגיאות מרוכזות קרוב לראשית (לאפס), והפילוג נראה פילוג לפלס בעל הצורה:

$$P(e) = \frac{\lambda}{2} \cdot \exp\{-\lambda|e|\}$$

מסתבר כי שגיאות חיזוי בתמונות בדרך כלל בעלות חוק פילוג כזה.

האנטרופיה במקרה זה היא 5.14 סיביות לפיקסל. ברור כי החיזוי המוצלח ביותר הוא לעשות שימוש בפונקציית ההסתברות המותנית, אך זו קשה לקבלה (היא גם משתנה במקום) וקשה לשימוש. כפי שאנו רואים, ע"י חיזוי פשוט למדי (אנו נראה בהמשך שניתן להציע טובים ממנו שיהיו מעשיים) קיבלנו אנטרופיה נמוכה בכ- 3 סיביות מהשמונה המקוריים.

כאשר נדון בדחיסת סדרות של תמונות אנו ננסה לנצל גם את הקשר ההדוק בין פיקסלים בין תמונות. מסתבר שקשר זה אף חזק יותר בהשוואה לקשר בין פיקסלים שכנים מאותה תמונה, כיוון שתמונות עוקבות בוודאק מאוד דומות. זהו בעצם המקור ליכולת הדחיסה החזקה כל-כך בוודאק.



3. יתירות פסיכו-ויזואלית - שתי היתריות הנ"ל מתייחסות למושג האנטרופיה. כעת ננצל את שלמדנו על פונקצית קצב-העיוות. מערכת הראיה שלנו אינה מושלמת, וניתן לנצל עובדה זו לשם השגת דחיסה טובה יותר.

לכן, אם רמות אפור בתמונה הנפרסת (לאחר דחיסה אובדנית) שונות במקצת מרמות האפור בתמונת המקור, במרבית המקרים אין אנו מסוגלים לזהות שינוי זה. כיוון שכך, במושגים של תורת האינפורמציה, אנו מתירים מרחק D כלשהו בין תמונת הכניסה ותמונת היציאה של תהליך הדחיסה-פריסה. כבר ראינו כי מרחק כזה מניב אנטרופיה קטנה יותר - כלומר - המידע מכיל אפקטיבית יותר יתירות.

לסיכום דיון זה על יתירויות, יתירות הסתברותית סקלרית מציעה חסכון קטן בייצוג האות. אם מוסיפים לה התייחסות להסתברויות בין פיקסלים, אנו בעצם עושים שימוש ביתירות מרחבית (או זמנית) בין פיקסלים, ואז ניתן לדחוס בערך פי שתיים. כאשר מכניסים למשחק את היתירות הפסיכו-ויזואלית ומתירים שגיאות קלות, יחס הדחיסה יכול לעלות לפי 10 ועד מאות - תלוי באיכות הרצויה ובאופי התמונה/האות המטופל.

עד כה עסקנו באנטרופיה תיאורטית. נשאלת השאלה - כיצד בונים מקודד מעשי המצליח להגשים את האנטרופיה או לפחות להתקרב לכך. באשר לקידוד ללא שגיאות בו אנו חותרים אל האנטרופיה, הכלי אותו נציג מיד הוא קוד הופמן. באשר לקידוד עם שגיאות, נזכור כי לרוב לא ניתן כלל לקבוע מהי האנטרופיה האפקטיבית. אנו נכיר מספר כלי יסוד המשמשים באלגוריתמי דחיסה אובדנית.

8.3 קידוד אנטרופיה - אלגוריתם הופמן

נניח כי נתון לנו מקור X המייצר תווים מעל אלף-בית של N תווים אפשריים (למשל - קוביה המייצרת מספרים בתחום $[0, 1]$ עד 6). עבור מקור זה ידוע לנו חוק ההסתברות $P(X)$, ועל פי פונקציה זו אנו יכולים לחשב את האנטרופיה התיאורטית $H(X)$.

רצוננו לבנות קוד אשר יגשים קצב שידור שיהיה האנטרופיה או קרוב אליה. לשם דוגמה, בהינתן תמונה בת מיליון פיקסלים כשכל פיקסל בעל רמת אפור בתחום $[0, 255]$, ברור כי ע"י הקצאת 8 סיביות לפיקסל בנינו קוד מעשי, אך אם האנטרופיה של רמות האפור (כסקלרים - אנו עדיין לא מדברים על בלוקים או על חיזוי) ידועה להיות 4.5 סיביות לפיקסל - כיצד נתכנן קוד שירד לרמה זו? לשם מתן תשובה, נציג דוגמה פשוטה וממנה נכיר שני עקרונות חדשים - קוד בעל אורך מילה משתנה, וקוד Prefix.

דוגמה 8.9: נתון מקור המייצג 6 תווים אפשריים $\{A, B, C, D, E, F\}$ עם ההסתברויות הבאות:

$$P\{A\} = 1/2, P\{B\} = 1/8, P\{C\} = 1/8, P\{D\} = 1/8, P\{E\} = 1/16, P\{F\} = 1/16$$

רואים כי סכום הסתברויות אלו הוא 1, כנדרש. בקידוד טריוויאלי נוכל להציע הקצאת 3 סיביות לכל תו ויהיה בכך בזבוז כיוון שיוותרו שני צירופים ללא שימוש. תוצאה זו צפויה כיוון שתיאורטית, עבור 6 תווים די ב- $\log_2 6 = 2.585$ סיביות לתו. במקרה שלנו האנטרופיה היא 2.125 סיביות לתו.

אם הגישה הטריוויאלית מקצה אורך זהה של מילות קוד לכל תו, גישה יעילה יותר תהיה להקצות מילים באורך שונה לתווים שונים - תו בעל הסתברות גדולה יקבל מילה קצרה, ותו בעל הסתברות נמוכה יקבל מילה ארוכה. בדוגמה שלנו - נשתמש בקוד הבא:

תו	הסתברות	קוד	אורך
A	1/2	0	1
B	1/8	100	3
C	1/8	101	3
D	1/8	110	3
E	1/16	1110	4
F	1/16	1111	4

האורכים אותם בחרנו הם בדיוק LOG לפי בסיס 2 של היפוך ההסתברות של כל תו. לכן, אורכה של המילה הממוצעת בקוד יהיה:

$$E[L] = \sum_{k=1}^N P\{X_k\}L_k = 1 \cdot \frac{1}{2} + 3 \cdot \frac{1}{8} + 3 \cdot \frac{1}{8} + 3 \cdot \frac{1}{8} + 4 \cdot \frac{1}{16} + 4 \cdot \frac{1}{16} = 2.125 \text{ bits/pixel}$$

וזהו בדיוק האנטרופיה, כך שיש לנו קוד מוצלח למדי. השאלה הנותרת היא האם הוא ניתן לפענוח? בקוד בו יש מספר קבוע של סיביות לכל תו, כל שעל המקלט לעשות הוא לפרק את סידרת הסיביות המגיעה לקבוצות שוות סיביות וכל קבוצה תייצג תו. כעת, כשמגיעה סידרת סיביות נדרש לפרקה לאורכים שונים, והשאלה אן זה אפשרי. נשים לב לתכונה הבאה המתקיימת בקוד שהוצע - אף מילת קוד קצרה אינה נהווה את הפתיח למילות ארוכות יותר. תכונה זו נקראת Prefix Code, והיא זו המבטיחה פענוח יחיד ונכון.

לדוגמה - נניח כי רצוננו לשדר את הסדרה - AABEACAADAFDBACA - נבנה עבור צירוף תווים אלו את סידרת הסיביות שתשודר - 0,0,100,1110,0,101,0,0,110,0,1111,110,100,0,101,0. סימני הפיסוק הם סימנים אותם המקלט לא יכול לקבל - עליו לפרק את הסדרה לבדו, ובזכות היות הקוד בעלת תכונת ה-Prefix הוא יכול. התו הראשון הוא 0, ולכן המילה היחידה האפשרית היא A. כך גם לסיבית השניה. בסיבית השלישית יש 1 ולכן זה יכול להיות B עד F. 1 כשלאחריו 0 יכולים להיות רק B ו-C, ו-1 כשלאחריו שני אפסים זה רק B. כך התהליך נמשך כשישנו רק פיסק אפשרי אחד - וזהו הפיסוק הנכון.

לסיכום דוגמה זו - ראינו כי קיים קוד המגשים את האנטרופיה בדיוק וזאת ע"י קוד בעל אורך מילה משתנה - מילים קצרות לתווים הסבירים יותר ומילים ארוכות לתווים נמוכי ההסתברות. כמו כן ראינו כי היות הקוד קוד Prefix מאפשרת פענוח יחיד ונכון של הסדרה. סוגיה שלא זכתה למענה עד כה - מאיפוא נולד הקוד שהוצע, וכיצד נוכל תכנן קוד למקרים אחרים?

קוד הופמן דן במקרה הכללי ומנסה להכליל את התוצאות הנ"ל. אנו דנים במקור X המייצר תווים מעל אלף-בית של N תווים אפשריים, ויש בידנו את חוק ההסתברות P(X), והאנטרופיה H(X). אלגוריתם הופמן מציע תהליך פשוט לפיו ייבנה קוד Prefix מוצלח. התהליך הוא תהליך של בניית עץ:

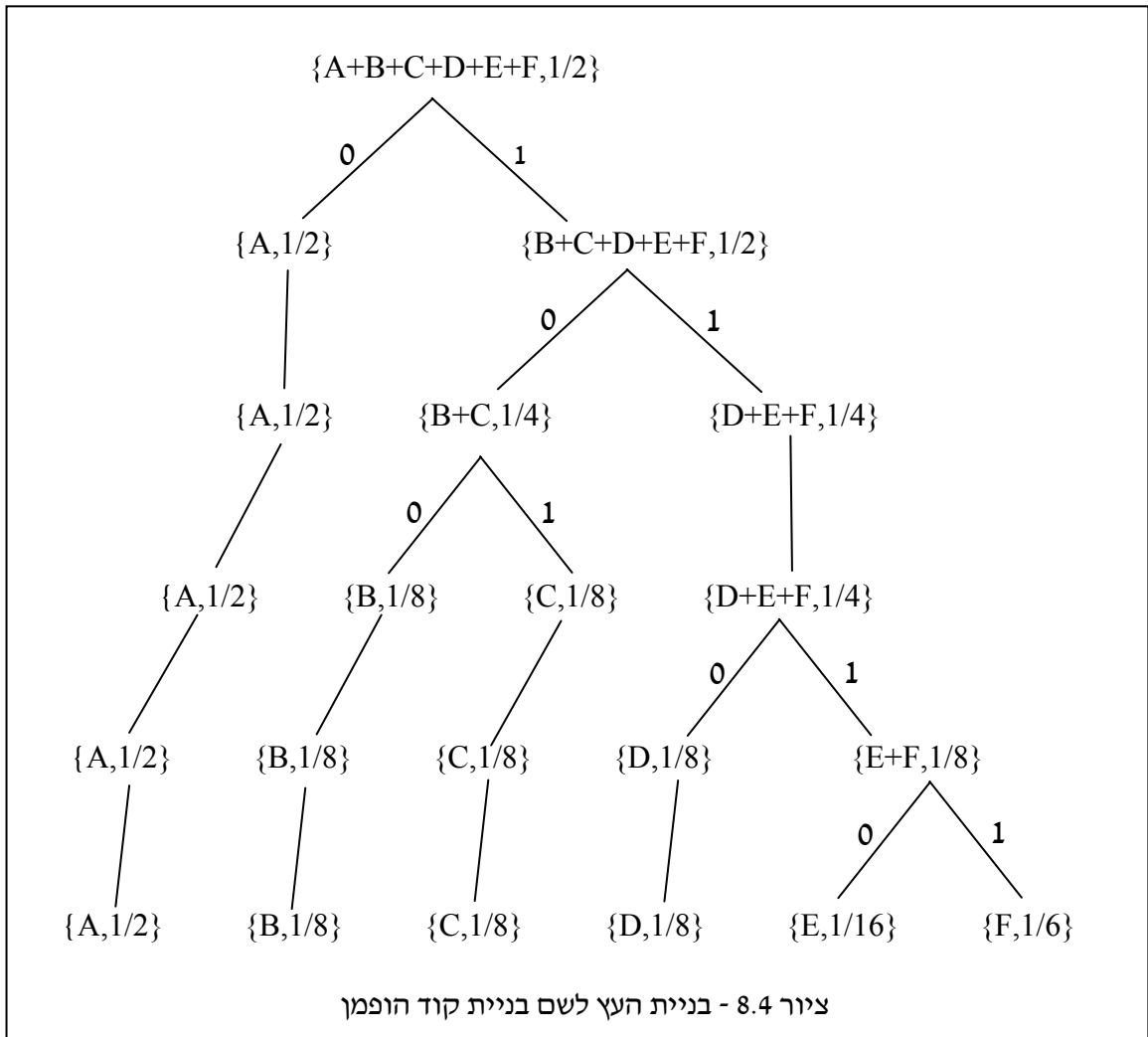
1. כתוב את סט N ההסתברויות בשורה. קבע $k=N$.
2. מתוך k התווים הקיימים, אחד את שתי התווים בעלי ההסתברויות הנמוכות ביותר. האיחוד ייעשה ע"י שני ענפי-עץ המתאחדים לצומת אחת. צומת זאת מגדירה תו חדש שהוא צירוף שני התווים שאוחדו, והסתברותו היא סכום הסתברותם. בסיום שלב זה מתקבל סט של $k-1$ תווים עם הסתברות לכל אחד, כשסכום ההסתברויות הוא 1.
3. אם $k>1$ חזור לסעיף 2. ואחרת - סיים.

התהליך שתואר בונה עץ. הקוד מתקבל ע"י הקצאת 0 ו-1 לכל לשני הענפים היוצאים מכל צומת. סיברת הסיביות המגדירה מסלול משורש העץ ועד לעלה היא מילת הקוד של התו המתייחס לעלה זה. נמחיש על תהליך זה דרך דוגמה.

דוגמה 8.9: נחזור לדוגמה הקודמת בה נתון מקור המייצג 6 תווים אפשריים - $\{A,B,C,D,E,F\}$ עם ההסתברויות $P\{A\}=1/2, P\{B\}=1/8, P\{C\}=1/8, P\{D\}=1/8, P\{E\}=1/16, P\{F\}=1/16$. ציור 8.4 מראה את העץ הנוצר.

בשורה התחתונה ישנה רשימה של כל התווים על הסתברויותיהם. בשלב ראשון מאוחדים E ו-F ביותם בעלי ההסתברות הנמוכה ביותר. נוצר תו חדש בנקרא E+F והסתברותו שמינית. בשלב שני יש ארבעה תווים בעלי הסתברות שמינית - בחרנו לאחד את D על (E+F). כך התהליך ממשיך עד לקבלת תו אחד בעל הסתברות 1, שהוא למעשה כל התווים כולם.

בשלב שני אנו מקצים סיביות לכל שני ענפי צומת. כך למשל, על מנת לצאת משורש העץ (למעלה) ועד התו E עלינו ללכת במסלול שמותווה ע"י ההחלטות 1, 1, 1, ולבסוף 0 - זוהי בדיוק המילה שהוצעה במסגרת דוגמה קודמת. למעשה - בדיקה קצרה מגלה כי עץ זה מניב במדויק את הקוד שהוצע בדוגמה קודמת.

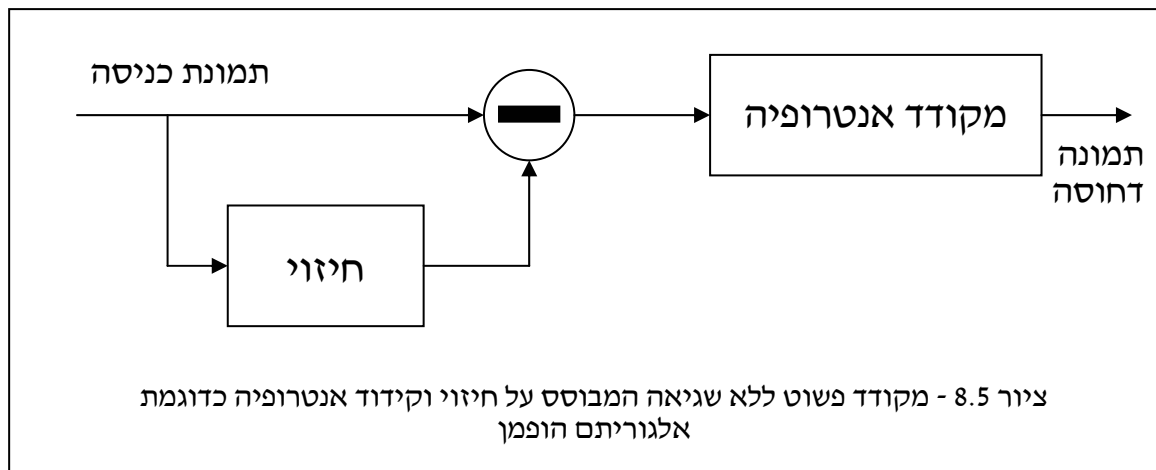


במקרה בו ההסתברויות מהוות חזקות שלמות שליליות של 2 (כמו הדוגמה הנ"ל) מתקבל כי קוד הופמן מגשים בצורה מדויקת את האנטרופיה, דהיינו, אורך המילה הממוצעת היא האנטרופיה. במקרה כללי יותר קוד הופמן ייתן אורך מילה שיהיה גבוה מהאנטרופיה בעד סיבית אחת לתו, כלומר:

$$H(X) \leq L_{\text{Average}} < H(X) + 1$$

כאשר נתון סט הסתברויות בו קוד הופמן סוטה בקרוב לסיבית אחת, ניתן להתגבר על כך ע"י עבודה בבלוקים. גם אם אין כל קשר סטטיסטי בין התווים, אם נתייחס לדוגמה לצמד תווים כ-"סופר-תו" חדש, הרי שנקבל סטייה של עד סיבית אחד ל-"סופר-תו". לכן, לתו בודד תתקבל סטייה של עד חצי סיבית. במקרה הכללי, עבודה על בלוק של N תווים תיתן סטייה של עד $1/N$ סיביות. ברור כי תרגיל זה בא עם מחיר - עבודה על בלוקים קשה יותר. לדוגמה - עבור מקור בין 100 תווים, קוד הופמן המקורי דורש בנייה של 100 מילות קוד ושימוש בהן. עבור צמדים נידרש לתכנון ואחסון של 10,000 מילות קוד, ותהליך הדחיסה/פריסה מסתבך בהתאם.

בהסתמך על כל האמור לעיל נוכל להציע אלגוריתם דחיסה פשוט ללא שגיאות לתמונה נתונה. אלגוריתם זה יקודד את ההפרשים בין פיקסלים סמוכים, וכבר ראינו כי האנטרופיה הסקלרית של מידע זה היא 5.14 סיביות. אם נשתמש בקוד הופמן נוכל לקבוע כי במקרה החמור ביותר נצליח להגיע ל-6.14 סיביות לפיקסל - דחיסה לא חזקה אך סבירה, ובמיוחד לנוכח פשטות האלגוריתם. אם נשתמש בחיזוי מוצלח יותר, וואו נקודד את שגיאות החיזוי בבלוקים נוכל לסחוט דחיסה טובה יותר. ציור 8.5 ממחיש בסכימת בלוקים את אלגוריתם הדחיסה המוצע.

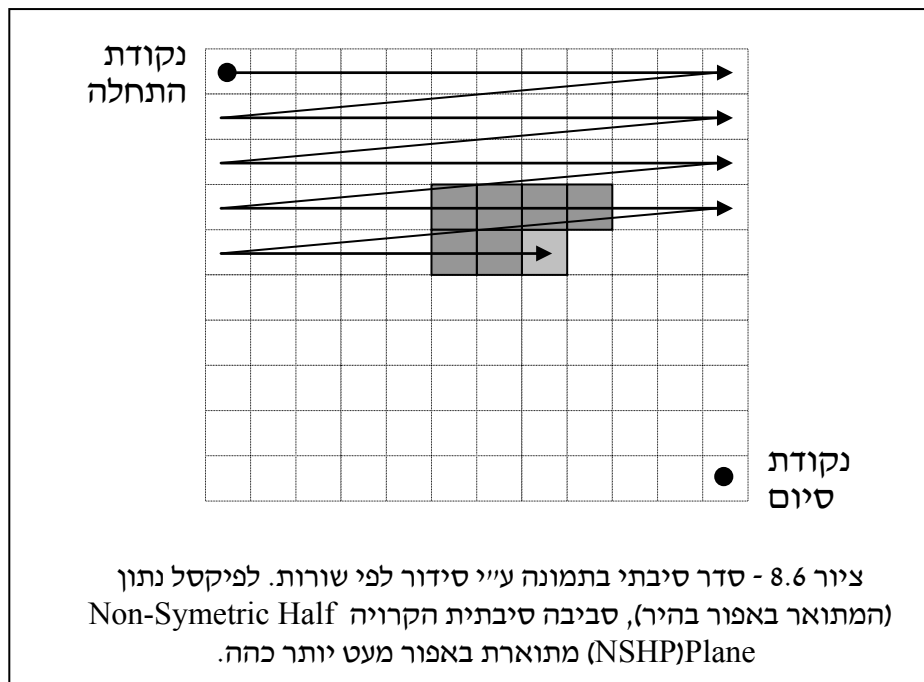


נקודה אחרונה בדיון זה על קודים בעלי אורך משתנה היא מידת רגישותם לשגיאות ערוץ. אם מופעל קוד הופמן על סידרת תווים ומתקבלת סידרת סיביות, די בשינוי של סיבית אחת בסדרה כדי לשבש את הפענוח של כל מה שיבוא אחריה. ניתן להילחם בתופעה זה במגוון דרכים, אך אנו לא ניכנס לנושא זה.

8.4 דחיסה עם שגיאות - גישת החיזוי

באופן כללי ישנן שתי גישות לביצוע דחיסה עם שגיאות - גישת החיזוי וגישת ההתמרה. אנו נציג את שתי גישות אלו במסגרת הדיון הבא. בשתי גישות אלו, המרכיב האחראי על אובדן המידע הוא קוונטיזר, כפי שכבר הוצג בפרק 4. ההבדל המהותי הוא בשאלה מהי הפעולה המקדימה שיש לבצע על הפיקסלים בטרם יוכנסו לקוונטיזר - חיזוי (כפי שכבר עשינו בסעיפים קודמים), או התמרה. שתי אפשרויות אלו בעלות מכנה משותף חשוב - שתיהן ממירות את הפיקסלים עתירי היתירות במספרים בעלי אנטרופיה נמוכה יותר, וקשר סטטיסטי דל בין האלמנטים השונים. כתוצאה מכך, המידע לאחר פעולת החיזוי/התמרה נוח יותר לדחיסה, ומאפשר עבודה בפיקסלים בודדים במקום בוקטורים.

נתחיל את הדיון בסכימה המבוססת על חיזוי. צעד ראשון לשם יישום של אלגוריתם דחיסה עם חיזוי הוא קביעת סדר סריקה לפיקסלים. החיזוי הינו חיזוי סיבתי - כלומר ערך החיזוי נבנה ע"י פיקסלים שכבר טופלו, וזאת על-מנת שגם המקלט יוכל לבנות את החיזוי ולצרפו לשגיאת החיזוי הנקלטת. בעוד שבאות חד-מימדי סיבתיות היא עובדה טבעית, בתמונה (אות דו-מימדית) אין זה כך. נניח לעת עתה כי התמונה מסודרת שורה לאחר שורה כמתואר בציור 8.6. בציור זה גם מתוארת סביבת פיקסלים סיבתית עפ"י סדר זה, אשר יכולה לשמש לשם בניית החיזוי. ישנן גישות שונות לסדר הפיקסלים אשר לא תידונה במסגרת זו.



בציור 8.5 הראינו כיצד פועל אלגוריתם עקרוני של דחיסה מבוססת חיזוי ללא שגיאות. על מנת להכניס שגיאות (שכזכור, צפויות לשפר את עומק הדחיסה), עלינו להכניס לפני קידוד האנטרופיה בלוק חדש - קוונטיזר. הקוונטיזר יוזן בשגיאת החיזוי אשר יכולה לקבל ערכים שליליים וחיוביים. במקום לקודד את השגיאות במדויק, הקוונטיזר יבצע "העגלה" של שגיאות אלו, וכתוצאה מכך תיווצר שגיאה בתהליך הדחיסה. בזכות "העגלה" זו, מקודד האנטרופיה יזדקק להרבה פחות סיביות לשידור שגיאות אלו, וכך יושג הרווח בדחיסה.

דוגמה 8.10: נניח כי החיזוי שלנו לפיקסל $X[k,j]$ הוא ממוצע של שכניו הסיבתיים לפי

$$\hat{X}[k,j] = \frac{1}{4}(X[k-1,j-1] + X[k-1,j] + X[k-1,j+1] + X[k,j-1])$$

בחיזוי זה עשינו שימוש בשלושת הפיקסלים המצויים מעליו, ובזה המצוי באותה שורה לשמאלו. בהנחה שכל פיקסל מקבל רמת אפור המיוצגת ע"י 8 סיביות, ערכי חיזוי זה ינועו בתחום -255 עד $+255$ ויקבלו כל ערך בתחום זה בקפיצות של 0.25 . לכן, יהיו 2041 ערכים אפשריים. דחיסה ללא שגיאה של אות שגיאה זה יחייב בנייה של קוד הופמן ל- 2041 ערכים אפשריים.

כתחליף, אם נפעיל קוונטיזר המעגל את השגיאה לערך השלם הזוגי, נקבל במוצא הקוונטיזר ערכי שגיאה חדשים באותו תחום אך עם 255 ערכים בלבד.

$$Q\{E\} = 2 \cdot \left\lfloor \frac{E+1}{2} \right\rfloor$$

קוונטיזר זה ימפה את הערכים $(-1,+1)$ לערך 0 , $(+1,+3)$ ל- 2 וכך הלאה. נמחיש רעיון זה על הפיקסלים בתמונה Lenna. נבנה לכל פיקסל את החיזוי כמתואר ונחשב את האנטרופיה ללא קוונטיזציה ועם קוונטיזציה כמתואר. לצורך חישוב האנטרופיה נזרוק את העמודות והשורות הקיצוניות, בהן החזאי נופל משטח התמונה. האנטרופיה ללא קוונטיזציה היא 4.86 סיביות לפיקסל, ואילו עם הקוונטיזציה מתקבלת אנטרופיה של 2.78 סיביות לפיקסל.

במקודד ללא שגיאה, הפיקסלים הנקלטים הם הפיקסלים המקוריים (ללא שגיאה מעצם הגדרתם ככאלה). לכן, המקלט מצליח לבנות חיזוי הזהה בערכו לחיזוי בו נעשה שימוש במשדר. בדחיסה עם שגיאה נוצרת בעיה - אם נבנה את החיזוי עפ"י הפיקסלים הסיבתיים המקוריים במשדר, ובמקלט ננסה לחקות תהליך זה ע"י בניית חיזוי על הערכים הנקלטים, נקבל ערכי חיזוי שונים. יתכן שהסטייה תהיה קטנה, אך זו תשפיע על הערכים המפוענחים הבאים ותיווצר תופעה של שגיאה מצטברת. כעבר מספר עשרות פיקסלים המפענח יניב ערכים שגויים לחלוטין וללא קשר לערכי המקור.

על מנת להתגבר על בעיה זו, על המשדר לבנות את החיזוי על פי ערכי רמות האפור כפי שיפוענחו במקלט. לכן, על המשדר להכיל בתוכו תת-מערכת שתהיה מקלט לכל דבר אשר מפענח את הערכים, ותת-מערכת זו היא אשר תניב את ערכי הפיקסלים על-פיהם ייבנה החיזוי. תהליך זה מתואר במלואו בציור 8.7. ציור זה מראה סכימה עקרונית בה האות המקורי מוזן לחזאי המנבא את הדגימה $X[k,j]$ על סמך שכניה הסיבתיים. חיזוי זה מסומן ב- $\tilde{X}[k,j]$. כאמור, על החיזוי במשדר להיות מבוסס על הערכים כפי ששודרו ולא עפ"י הערכים המקוריים. אנו נסמן את ערכי הדגימות הנקלטות במקלט ב- $\hat{X}[k,j]$. החיזוי הוא למשל פונקציה מהצורה:

$$P\{X[k, j]\} = \tilde{X}[k, j] = f\{\hat{X}[k-1, j-1], \hat{X}[k-1, j], \hat{X}[k-1, j+1], \hat{X}[k, j-1]\}$$

בשלב השני מחוסר החיזוי מהערך האמיתי ומתקבלת השגיאה

$$E[k, j] = X[k, j] - \tilde{X}[k, j]$$

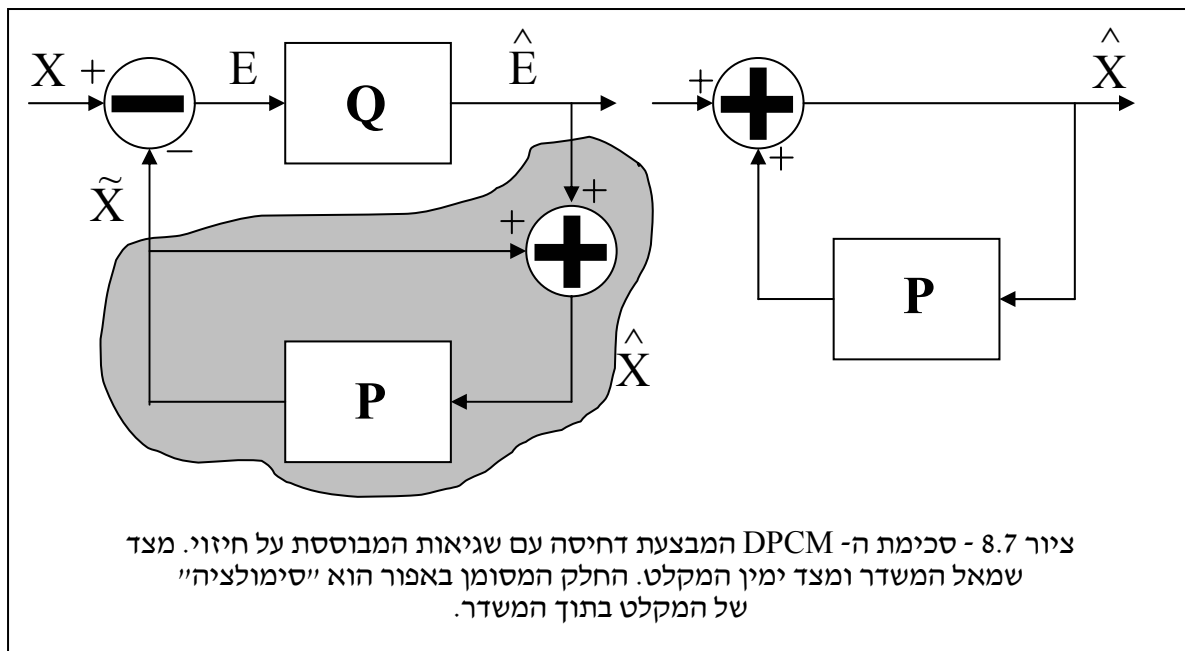
שגיאה זו מוזנת לקוונטיזר Q אשר ממיר את ערכי השגיאה בערכים מקורבים שלהם. זאת כיוון שאנו מוותרים על שידור השגיאות במדויק ורצוננו לשדר ערכים דומים להם. ברור כי הקוונטיזר היא פונקצית מדרגות, בה ככל שהמדרגות גדולות יותר, שגיאת הקוונטיזציה גדולה יותר. במוצא הקוונטיזר מתקבל

$$\hat{E}[k] = Q\{E[k]\}$$

גודל זה הוא המשודר החוצה, ולצורך כך נעשה שימוש במקודד אנטרופיה כדוגמת קוד הופמן. אם הקוונטיזר אינו מאבד מידע (כלומר מהווה מערכת יחידה), אנו מקבלים מערכת דחיסה ללא שגיאה. המקלט מקבל את $\hat{E}[k]$ ומחבר אליו את החיזוי $\tilde{X}[k, j]$. לכן:

$$\begin{aligned} \hat{X}[k, j] &= \hat{E}[k, j] + \tilde{X}[k, j] = \\ &= (E[k, j] + \tilde{X}[k, j]) - (E[k, j] - \hat{E}[k, j]) = \\ &= X[k, j] - (E[k, j] - \hat{E}[k, j]) \end{aligned}$$

כלומר, המוצא של המקלט הוא הערך האמיתי של הפיקסל עם תוספת שגיאה שהיא בדיוק שגיאת הקוונטיזציה. נשים לב כי המשדר מכיל בתוכו סימולציה של המקלט לשם בניית הערכים $\hat{X}[k, j]$.



דוגמה 8.11: אם נחזור לדוגמה 8.9, הרי שקיבלנו ירידה בשתי סיביות בקירוב לפיקסל ע"י קוונטיזציה ששגיאתה היא לכל היותר רמת אפור אחת - שגיאה שאנו לא נצליח להבחין בה. בהמשך לדוגמה זו, עבור הקוונטיזר הבא:

$$Q\{E\} = 4 \left\lfloor \frac{E+2}{4} \right\rfloor$$

יתקבל 0 עבור כניסות בתחום $(-2, 2]$, 4 עבור כניסות בתחום $(2, 6]$, וכו'. במקרה זה האנטרופיה המתקבלת היא 2.1 סיביות לפיקסל, והשגיאה המירבית הפעם היא 2 רמות אפור.

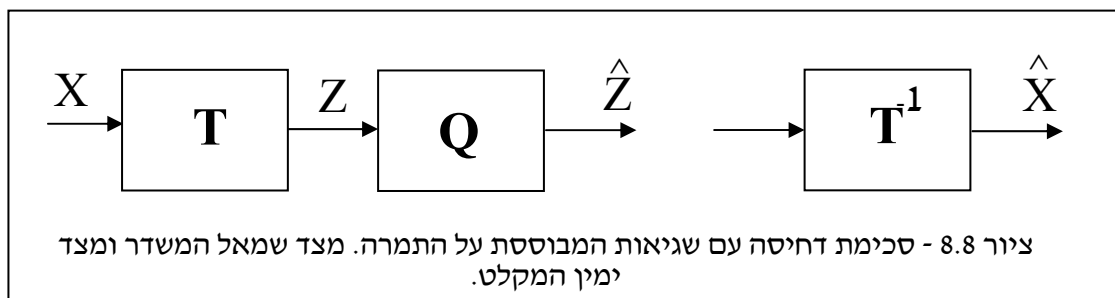
עד כה הנחנו קוונטיזר פשוט ויוניפורמי. למעשה, ניתן ליישם קוונטיזר כללי יותר בעל מדרגות ברוחב משתנה, אשר יתוכנן ע"י אלגוריתם מקס-לויד. קוונטיזר כזה ינסה ליצור מדרגות צרות יותר סביב הראשית כיוון שלערכים אלו הסתברות גבוהה. לכן, השגיאה הכוללת בין האות המשודר לאות הנקלט יהיה קטן יותר. ניתן גם להשתמש בשיקולים הנובעים מתוך מערכת הראיה בבואנו לתכנן את הקוונטיזר, אך אנו נעצור את הדין בנקודה זו.

אם נסכם, הרי שלמרכזו של האלגוריתם היה תהליך החיזוי והחסרתו אשר הניב תמונה חדשה בה הפיקסלים בעלי שתי תכונות חשובות:

1. אין קשר בין הפיקסלים - תכונה זו מאפשרת לנו לעבוד על כל פיקסל בנפרד מבלי לחשוש שאנו מוותרים על יתירות מרחבית.
2. ערכי הפיקסלים יוצרים פילוג נמוך אנטרופיה (לפני הקוונטיזר ובוודאי שגם אחריו). לכן, האות דחיס בצורה יעילה.

8.5 דחיסה עם שגיאות - גישת ההתמרה

כתחליף לגישת החיזוי מוצעת גישה אחרת - גישת ההתמרה. גם גישה זו מנסה להשיג את שתי התכונות הנ"ל אך בדרך שונה. ניקח את התמונה הנדרשת - נניח כי היא בגודל M שורות על N עמודות - ונבצע עליה התמרה. נניח כי אנו משתמשים בהתמרה ריבועית יוניטרית ומתקבלת לפיכך תמונת התמרה בגודל זהה. אם ערכי ההתמרה יהיו חסרי קשר סטטיסטי ביניהם, ובנוסף בעלי אנטרופיה נמוכה, נוכל להעביר כל אלמנט כסקלר לקידוד אנטרופיה, ולהשיג כתוצאה מכך דחיסה. על מנת להכניס שגיאות לתהליך ולשפר את יחס הדחיסה, נוכל לקחת את ערכי ההתמרה ולהעבירם דרך קוונטיזר בטרם העברתם לקידוד אנטרופיה. ציור 8.8 מראה סכימת בלוקים של מקודד ומפענח הבנויים על עקרון זה.



נותרה השאלה המרכזית - מיהי ההתמרה אשר תגשים יעד זה של ביטול הקשר בין הפיקסלים והקטנת האנטרופיה שלהם. שאלה זו מחזירה אותנו לפרק 7 ולתכונות שיצרו את התמרת ה-KLT. ראינו כי התמרת ה-KLT היא ההתמרה היוניטרית הטובה ביותר במובן שאם נזרוק חלק ממקדמי ההתמרה, השגיאה בין תמונת המקור לתמונה המשוחזרת לאחר התמרה הפוכה תהיה מינימלית. זריקת חלק מהאיברים כמוה כמו ביצוע של קוונטיזציה גסה מאוד עבור האיברים הנזרקים. יתרה מזו - הזכרנו כי ה-KLT מלכסנת את מטריצת האוטו-קורלציה של המידע בתמונה, ומסתבר (ללא הוכחה) כי ליכסון זה כמוהו כמו ביטול הקשר הסטטיסטי בין הפיקסלים השונים.

ברור כי אסור לשכוח את החסרון המרכזי של התמרת ה-KLT - היותה תלויה בתמונה בה אנו משתמשים. בשל תכונה זו לא ניתן להשתמש ב-KLT לדחיסה. אבל ראינו כי התמרת ה-DCT מהווה קירוב טוב מאוד ל-KLT, והדבר נכון במיוחד לתמונות קטנות (16 על 16 ומטה פיקסלים). לכן, כתחליף ל-KLT ניתן לקחת את התמונה, לחלק אותה לבלוקים קטנים (גודל אופייני לשימוש הוא 8 על 8), לבצע על כל בלוק כזה התמרת DCT דו-מימדית. התמרה זו יכולה להיעשות ע"י אלגוריתם מהיר המבוסס על ה-FFT.

לאחר ההתמרה, כל מקדם התמרה מועבר דרך קוונטיזר. מתוך 64 הערכים המתקבלים (בשל ההתמרה על 8 על 8), הקרובים לראשית הם החשובים בהיותם מייצגים מידע נמוך תדר. עבורם תיעשה קוונטיזציה עדינה (שקול לאמירה כי להם תינתן שגיאה קטנה יחסית). לערכים הרחוקים מהראשית תבוצע קוונטיזציה גסה יותר, ויתכן שאף ערכים אלו ייזרקו. גישה מקובלת היא לבצע קוונטיזציה יוניפורמית לערכי ההתמרה עם גודל מדרגה התלוי במיקום המקדם - קטן עבור מקדם הקרוב לראשית, וגדול עבור מקדם הרחוק מהראשית. גם כאן ניתן להשתמש בקוונטיזר מקס-לוייד או קוונטיזציה הנבנית מתוך שיקולי מערכת הראיה.

האלגוריתם שתואר הוא אלגוריתם ה-JPEG הידוע ביותר הסטנדרט לדחיסת תמונות. אלגוריתם זה מבצע דחיסה ע"י התמרת DCT דו-מימדית וקוונטיזציה אחידה תלויית מקדם.

דוגמה 8.12: נדחוס את התמונה Lena ע"י אלגוריתם בסיסי מבוסס התמרה. בשלב ראשון כל בלוק של 8 על 8 מותמר ע"י DCT דו-מימדי. כל בלוק של 8 על 8 עובר קוונטיזציה ל-8 סיביות באיבר ה-(0,0), ל-3 סיביות לאיברים (0,*), ו-(*,0), (סה"כ 14 איברים כאלה המייצגים את השורה הראשונה והעמודה הראשונה), ול-2 סיביות לאיברים (1,1), (1,2), (2,1), ו-(2,2). לכן, לכל בלוק שבמקור יוצג ע"י $8 \times 64 = 512$ סיביות אנו משתמשים ב-68 סיביות, כלומר יחס דחיסה של 1 ל-8 בקירוב. גודל המדרגה נקבע ע"י חישוב השונות של כל מקדם התמרה ע"י מיצוע הבלוקים בכל התמונה. התקבלה התמונה הבאה, המופיעה בציור 8.9. השגיאה בין שתי התמונות (מחושבת ע"י סכום ריבועי הפרשים מחולק במספר הפיקסלים ולאחר שורש) היא 13.14 (ביחידות של רמות אפור).

מעיון בתמונה המתקבלת ניכרת תופעה האופיינית לאלגוריתמי דחיסה הפועלים בבלוקים - אפקט הבלוקיות. גבולות הבלוקים ניכרים וצורמים - תופעה שנובעת מכך שהפעולות שבוצעו זרקו מידע

גדול מהבלוק, ולא נעשה כל ניסיון לשמור על רצף בין הבלוקים. קיימים אלגוריתמי דחיסה מוצלחים יותר אשר מטפלים יפה בבעיה ומצליחים להשיג יחס דחיסה של 1 ל-10 בלי שעין אנושית תוכל להבחין בהבדל. קיימים גם אלגוריתמים הנובעים מתוך תורת השחזור שנדונה בפרק 6 אשר יודעים לקחת תמונה עם אפקט קלקול זה ולתקנה תוך כפיית חלקות בין הבלוקים ושימוש שפות באופן כללי.



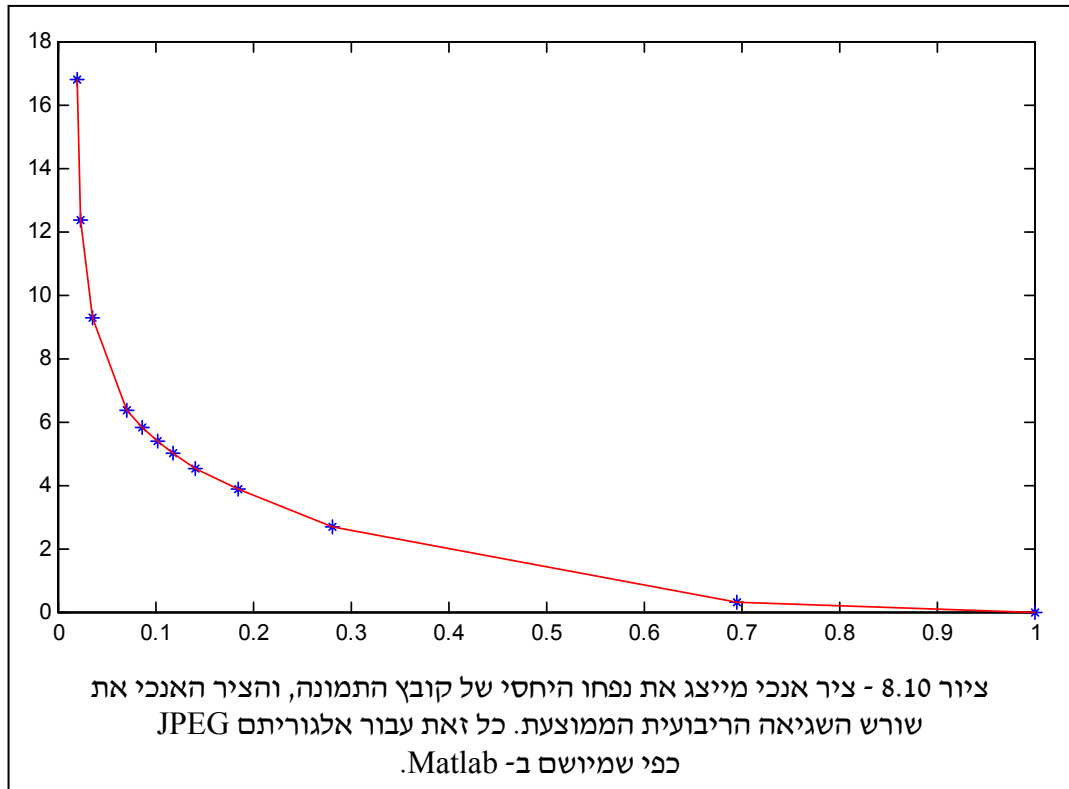
ציור 8.9 - תמונת מקור (משמאל) וגרסה דחוסה ופרוסה שלה ביחס 1 ל-8 ע"י אלגוריתם דחיסה פשטני המבוסס על התמרה, וללא קידוד אנטרופיה במוצאו

לשם השוואה, נפעיל את אלגוריתם ה-JPEG כפי שמיושם ב-Matlab. הפעלת אלגוריתם זה נעשית ע"י פקודת שמירת התמונה (imwrite) באופן הבא - בהינתן תמונה כמערך ב-Matlab הקרוי A (ונתון כ- uint8), הפקודה

```
imwrite(A,'temp.jpg','quality',70)
```

שומרת תמונה זו לקובץ בשם temp.jpg (השם מסגיר איזה פורמט התמונה מקבלת - במקרה זה דחיסת JPEG). באלגוריתם ה-JPEG, לא קובעים מראש את יחס הדחיסה אלא את מידת אובדן המידע. עבור איכות של 100% מקבלים דחיסה ללא אובדן, ועבור ערך נמוך יותר מתירים אובדן. כאשר ערך זה מאוד נמוך יתקבל יחס דחיסה מצוין על חשבון שגיאה גדולה בין התמונה המקורית לתמונה לאחר פריסה.

ציורים 8.10 ו-8.11 מראים תוצאות של דחיסת ה-JPEG. ציור 8.10 מראה גרף שהתקבל עבור מגוון ערכי האיכות (quality), וכנגד כל ערך כזה מוצג שורש השגיאה הריבועית הממוצעת כפונקציה של נפחו היחסי של הקובץ ביחס למקור. רואים כי עבור התמונה המקורית בנפח יחידה מקבלת שגיאה 0, ירידה לנפח של 0.7 מהמקור נותנת שגיאה זעומה של כ-0.3 רמות אפור. תוצאה זו מעניינת כיוון שהיא חושפת תכונה שלילית של קידוד מבוסס התמרה - יכולנו לקבל נפח של כ-0.5 ללא שגיאה כלל, לו עשינו שימוש בדחיסת אנטרופיה. לעומת זאת, בהתמרה, מתקבלת שגיאה גם לנפחים גדולים מזה. עבור נפח של כ-0.1 מתקבלת שגיאה של 5.4 - הרבה יותר טוב מהשגיאה אותה קיבלנו בדוגמה לעיל.



ציור 8.11 - תוצאות ה-JPEG: שמאל למעלה - מקור, ימין למעלה - יחס דחיסה 7.12 ושגיאה 11.65, למטה משמאל - יחס דחיסה 5.84 ושגיאה 4.5, למטה בצד ימין - יחס דחיסה 52.27 ושגיאה 16.8

פרק 9 – עיבוד סדרות של תמונות

195	פרק 9 – עיבוד סדרות של תמונות
196	9.1 תנועה בין תמונות
201	9.2 שערך תנועה בין תמונות - גישת בלוקים
205	9.3 שערך תנועה בין תמונות - גישה מלאה
210	9.4 ניקוי רעש בסדרות של תמונות
215	9.5 דחיסת סדרות של תמונות ומבוא ל-MPEG

בפרק זה אנו נכנסים לתחום חדש של טיפול בסדרות של תמונות, כלומר, סרטי וידאו. כבר אמרנו בפרק ראשון כי סרט אינו אלא סידרה ארוכה של תמונות בודדות. בדרך כלל מדובר בכ- 30 תמונות בשניה לשם קבלת אשליה של רצף כשמסתכלים על הסדרה בקצב זה. כשמסתכלים על צמד תמונות עוקבות אופייני בסרט כלשהו, ההתרשמות הראשונה היא שמדובר בתמונות זהות. כאשר מידת הפעילות (כמות התנועה) בסרט מועטה, התמונות תהיינה באמת כמעט זהות. השוני בין התמונות בד"כ מיוחס לתנועה ביניהן. תכונה זו מהווה את המוטיב העיקרי באלגוריתמים הדנים בסדרות של תמונות.

מהם הטיפולים הדרושים לסדרות של תמונות? ובכן, כל שאמרנו עד כה לתמונות בודדות רלוונטי גם לסדרות - נרצה לדגום ולבצע קוונטיזציה, נרצה לשפר במגוון אופנים, נרצה לשחזר מקלקולים ידועים, ונרצה לדחוס. במרבית המקרים ניתן לקחת אלגוריתמים שנועדו לתמונה בודדה ולאמצם לסדרה של תמונות ע"י הפעלתם על כל תמונה בנפרד. בד"כ גישה זו לא מוצלחת בשל שתי סיבות עיקריות:

1. משימה המבוצעת על סידרה של תמונות צריכה בד"כ לתת תוצאה טובה יותר בהשוואה לגישה בה כל תמונה מטופלת בנפרד. זאת כיוון שבסדרה קיים יותר מידע עליו ניתן להתבסס. לכן למשל, ניתן לנקות רעש בסדרה טוב יותר מאשר מתמונה בודדה, וניתן לדחוס סדרה עמוק יותר עבור אותה שגיאה מאותה סיבה בדיוק.

2. משימה המבוצעת על סידרה צריכה בד"כ להיות קלה יותר בהשוואה למשימה המבוצעת על תמונה בודדה. זאת כיוון שאם כבר סיימנו את המשימה עבור התמונה הקודמת, הרי שהטיפול לתמונה הנוכחית כמעט עשוי עוד בטרם התחלנו בשל הדמיון הרב בין התמונות. ברור כי לכל אלגוריתם יש למצוא את המכניזם לניצול תכונה זו.

במסגרת פרק זה אנו נתרכז בשלוש משימות עיקריות - הראשונה מציאת ושערוך תנועה בין תמונות. משימה זו מהווה מרכיב מרכזי בבואנו לטפל בסדרות של תמונות, ולרוב אינה מהווה מטרה בפני עצמה. המשימה השנייה בה נדון היא ניקוי רעש, והמשימה השלישית היא דחיסת סרטים. אנו נראה כי שני היישומים האחרונים משתמשים בשערוך תנועה (בדרכים שונות) על מנת למצות את הפוטנציאל הטמון במידע הנתון.

9.1 תנועה בין תמונות

נציג את הבעיה באופן הבא: נתונות שתי תמונות עוקבות בסרט וידאו - $X[k,j,t]$ ו- $X[k,j,t-1]$. השאלה הראשונה שנשאל היא - מהו הקשר בין תמונות אלה? נציג מספר צמידים מתוך סדרות של תמונות, ועל-פיהם ננסה להציע כלל כוללני. ציור 9.1 מראה שלושה צמידים עוקבים של תמונות משלושה סרטים שונים. לכל צמד, מוצגת גם תמונת ההפרש בערך מוחלט, וכפי שניתן לראות, תמונה זו בעיקרה שחורה - דבר המעיד על הפרש נמוך. האזורים הנעים הם החלקים עם השגיאה הגדולה. על בסיס זה מוצע המודל הבא לתאר את הקשר שבין שתי התמונות:

$$X[k,j,t] = X[k - dx(k,j,t), j - dy(k,j,t), t - 1]$$

כאשר $dx(k,j,t)$ ו- $dy(k,j,t)$ הם מרכיבי התנועה בציר האופקי והאנכי של הפיקסל ה- (k,j) בזמן t . משוואה זו טוענת בפועל כי כל פיקסל בתמונה בזמן t מקורו בפיקסל כלשהו מהתמונה הקודמת, כאשר יתכן שפיקסל זה גם עבר תמונה. לכן, רמות אפור נותרות קבועות בזמן, אך זזות ימנהלשמאלה ומעלהלמטה. ציור 9.2 ממחיש רעיון זה לוקטורים חד-מימדיים. הרעיון המרכזי הוא שאותות שונים קשורים זה לזה דרך עיוות של הקואורדינטות.

שאלה ראשונית בבואנו לטפל בסדרות של תמונות היא כיצד מוצאים את שדה התנועה בזמן t ? שדה תנועה זה יוכל לשרת אותנו לשם ייחוס נכון של הפיקסלים בין התמונות ובשל כך לייעל לאין ערוך פעולות כמו שיפור או דחיסה. סוגיה זו קרויה בשם הכולל "שערוך תנועה". כמות הנעלמים בבעיה כללית זו עצומה - פעמיים מגודל התמונה. במסגרת פרק זה נציג שתי אסכולות מרכזיות בנושא זה של שערוך תנועה - שערוך התנועה לכל פיקסל בנפרד, ושערוך לבלוקים של פיקסלים. לשתי גישות אלו יש יישומים מגוונים ושונים, והן משלימות זו את זו.

לעיתים קרובות התנועה בין התמונות ניתנת לתיאור פרמטרי פשוט. נסקור בקצרה מספר מודלים כאלה, ונראה את תכונותיהם. אנו נניח כי הקואורדינטות נתונות במישור ברצף:

1. הזזה גלובלית - במקרה זה הפיקסל המצוי במישור בנקודה (x,y) עובר למיקום החדש באופן הבא:



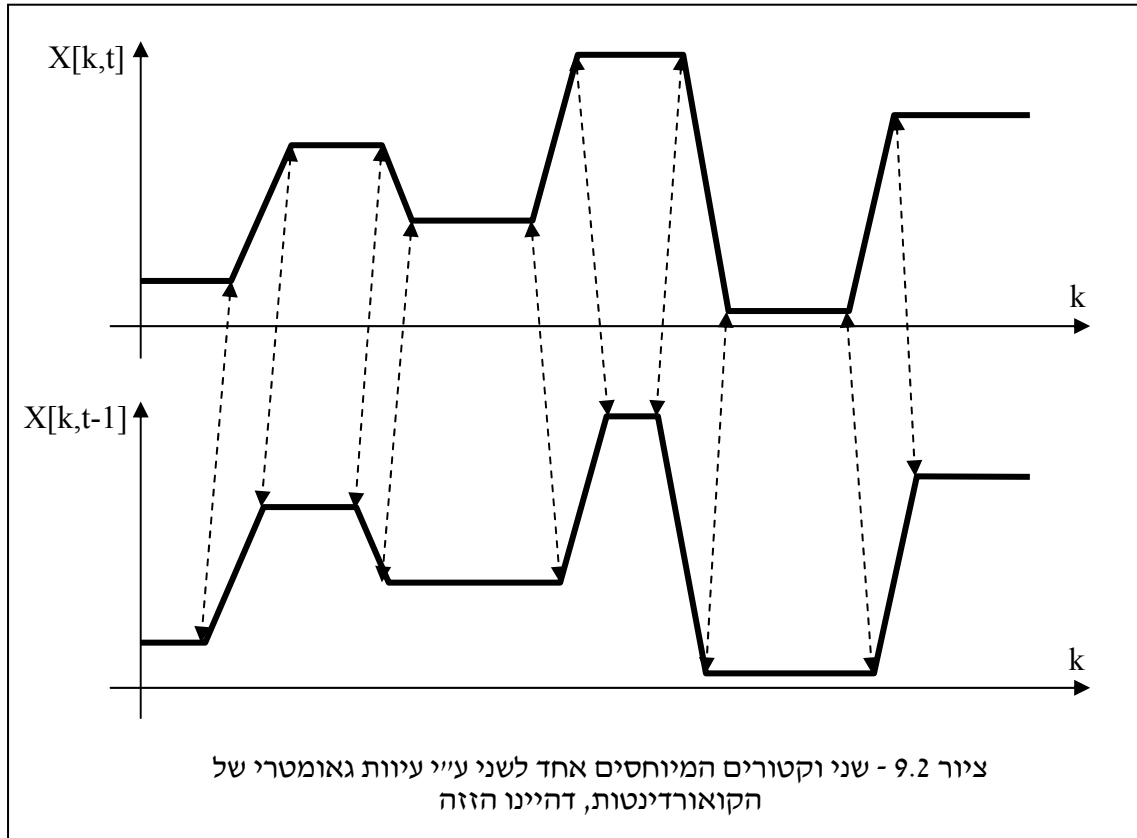
$$\begin{bmatrix} x \\ y \end{bmatrix}_{t-1} = \begin{bmatrix} x \\ y \end{bmatrix}_t + \begin{bmatrix} e \\ f \end{bmatrix} \Rightarrow \begin{bmatrix} x \\ y \end{bmatrix}_t - \begin{bmatrix} x \\ y \end{bmatrix}_{t-1} = -\begin{bmatrix} e \\ f \end{bmatrix}$$

במקרה זה, שדה התנועה בנוי מחיצים שגודלם קבוע ושווה $[-e, -f]$.

2. הזזה וסיבוב גלובליים (עיוות אויקלידי) - במקרה זה המיקום החדש נתון ע"י הקשר:

$$\begin{bmatrix} x \\ y \end{bmatrix}_{t-1} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}_t + \begin{bmatrix} e \\ f \end{bmatrix} \Rightarrow$$

$$\begin{bmatrix} dx(x, y, t) \\ dy(x, y, t) \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix}_t - \begin{bmatrix} x \\ y \end{bmatrix}_{t-1} = \begin{bmatrix} 1 - \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & 1 - \cos(\theta) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}_t - \begin{bmatrix} e \\ f \end{bmatrix}$$



המטריצה מסובבת בזווית θ ואז מתבצעת פעולת הזזה כמקודם. הפעם שלושה פרמטרים מאפיינים את מלוא התנועה. שדה התנועה הפעם הם חיצים תלויי מקום, כפי שניכר מהמשוואה

3. עיוות דמיון - במקרה זה המיקום החדש נתון עייו הקשר :

$$\begin{bmatrix} x \\ y \end{bmatrix}_{t-1} = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}_t + \begin{bmatrix} e \\ f \end{bmatrix}$$

הסיבוב וההזזה מתבצעים כמקודם. בנוסף, יש הגדלה/הקטנה שונה בכל ציר, ובסה"כ 5 פרמטרים. ברור כי גם הפעם יתקבל שדה חיצים משתנה במקום.

4. עיוות אפיני - הקשר הפעם נתון עייו :

$$\begin{bmatrix} x \\ y \end{bmatrix}_{t-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}_t + \begin{bmatrix} e \\ f \end{bmatrix} \Rightarrow$$

$$\begin{bmatrix} dx(x, y, t) \\ dy(x, y, t) \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix}_t - \begin{bmatrix} x \\ y \end{bmatrix}_{t-1} = \begin{bmatrix} 1-a & -b \\ -c & 1-d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}_t + \begin{bmatrix} e \\ f \end{bmatrix}$$

כך שישנם שישה פרמטרים המאפיינים את התנועה. תנועה אפינית מכילה בתוכה את עיוות הדמיון עם אפשרות של מרכיב הקרוי shear - גזירה - המבצע מתיחה אלכסונית של הצירים.

5. עיוות פרויקטיבי - הקשר כעת מסובך יותר ונתון ע"י:

$$\begin{bmatrix} x \\ y \end{bmatrix}_{t-1} = \frac{\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}_t + \begin{bmatrix} e \\ f \end{bmatrix}}{\begin{bmatrix} g & h \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}_t + 1}$$

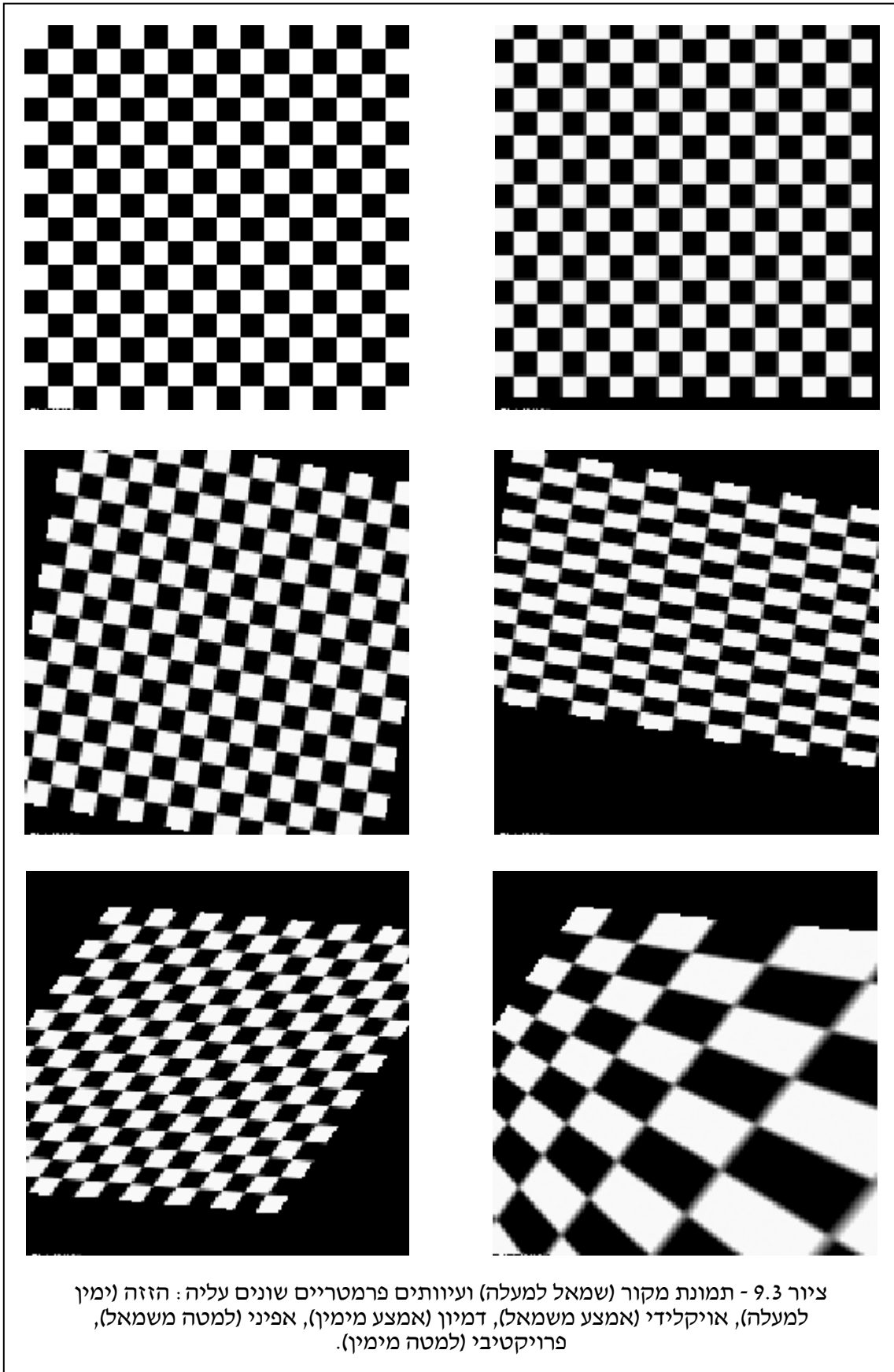
וכעת שמונה פרמטרים נותנים את הקשר בין הקואורדינטות. בעיוות זה ובכל קודמיו, קווים ישרים מומרים לקווים ישרים.

6. עיוות פולינומיאלי - בניגוד לעיוותים הקודמים אשר להם קשר פיזיקלי ישיר לדרך היווצרן של תמונות, בהיותן הטלות מגוונות מתלת מימד לדו-מימד, ניתן להציע קשרים שרירותיים תלוי פרמטרים בין הקואורדינטות. קשר אפשרי מסוג זה הוא הקשר הבא:

$$\begin{aligned} x(t-1) &= a_1 + b_1 \cdot x(t) + c_1 \cdot y(t) + d_1 \cdot x(t)y(t) + e_1 \cdot x(t)^2 + \dots \\ y(t-1) &= a_2 + b_2 \cdot x(t) + c_2 \cdot y(t) + d_2 \cdot x(t)y(t) + e_2 \cdot x(t)^2 + \dots \end{aligned}$$

בציור 9.3 מובאת תמונת לוח שחמט ועיוותים עליה על מנת להמחיש עיוותים אלו. ההזזה (שורה עליונה מימין) בוצעה עם וקטור תנועה של [2.1, 4.2], העיוות האויקלידי (אמצע בצד שמאל) בוצע עם הזזה של [-15, 15] וזווית של 10°. בעיוות הדמיון נעשה שימוש באותה זווית, עם הזזה של [-5, 5] ועם הגדלות של 0.7 ו-1.5 בהתאמה לשני הצירים. העיוות האפיני התקבל עם הזזה של [-45, -15] וארבעת איברי המטריצה הם [1, 0.7, -0.1, 1.3]. לקבלת העיוות הפרויקטיבי עשינו שימוש באותם ערכי התמרה כבמקרה האפיני עם שימוש בשני הפרמטרים הנוספים [0.01, 0.0062]. בדוגמאות אלו, כאשר מתקבלת נקודה חדשה במיקום שאינו שלם, בחרנו לבצע אינטרפולציה ביליניארית ע"י שימוש בארבעת השכנים של הנקודה. כמו כן, נקודות שפולשות אל מחוץ לשטח התמונה נחשבות כאפסים.

אם נחזור לצמד תמונות עוקבות בסדרה אופיינית, תנועה גלובלית הניתנת לתיאור ע"י מספר פרמטרים מצומצם ניתנת לייחוס לתנועת מצלמה כאשר העצם המצולם ניח ורחוק יחסית מהמצלמה, או אם העצם מישורי. כאשר לעצם תבנית תלת-מימדית עם עומק משמעותי ביחס למרחק מהמצלמה, התנועה מורכבת יותר כיוון שנקודות קרובות נעות מהר יותר מנקודות רחוקות. תנועה גלובלית ניתנת גם לקבלה כאשר המצלמה ניחת ואילו עצם קשיח הממלא את כל שטח התמונה נע בתלת-מימד. בכל מקרה אחר של עצמים תלת-מימדיים קרובים ומצלמה נעה, או תנועה מקומית של עצמים (כגון מכונית בכביש, כשהתמונה ניחת והמכונית מהווה חלק משדה הראיה), מתקבל שדה תנועה מורכב יותר בו בעקרון לכל פיקסל צפוי להיות חץ שונה המעיד על כיוון תנועתו. במסגרת הסעיפים הבאים נדון בשאלה כיצד משערכים את שדה התנועה הזה. אנו נגביל את הדיון לשערוך תנועה כללי מבלי להניח מודל פרמטרי - ישנם אלגוריתמים המסוגלים לפעול הרבה יותר מדויק ומהר אם מניחים מראש מבנה מוגבל של תנועה.



נקודה אחרונה בה נדון במסגרת סעיף זה היא ייצוג פעולת עיוות גיאומטרי כלשהו כמכפלה במטריצה. נסתכל על המשוואה

$$X[k, j, t-1] = X[k + dx(k, j, t), j + dy(k, j, t), t]$$

משוואה זו קובעת כי אופרטור העיוות הגיאומטרי המסוים על התמונה $X[k, j, t]$ מניב את התמונה $X[k, j, t-1]$. אופרטור זה ליניארי כיוון שאם נפעיל אותו על צירוף ליניארי של תמונות נקבל צירוף ליניארי של תוצאות העיוות. לכן ניתן לייצגו כמטריצה אשר תסומן $F[t, 1]$, כאשר סימון זה קובע כי אופרטור זה מעביר תמונה מזמן t לנקודת הזמן $t-1$. במקרה הכללי, $F[t, k]$ מעבירה את התמונה בזמן t לזמן $t-k$. ברישום לקסיקוגרפי נקבל את הקשר הבא:

$$\underline{X}(t-1) = F(t, 1)\underline{X}(t)$$

כאשר נניח כי אורך הוקטורים המעורבים הוא N . אם נסתכל על שורה אחת במערכת משוואות זו נוכל להבין כיצד בנויה המטריצה F . שורה מסוימת תיראה כך:

$$X_{n_0}(t-1) = X[k_0, j_0, t-1] = \sum_{n=1}^N F_{[n_0, n]}[t, 1]X_n(t)$$

משוואה זו אומרת כי הפיקסל ה- n_0 בתמונה הקודמת (אינדקס זה הוא האינדקס בסידור לקסיקוגרפי, והוא שקול למיקום ה- $[k_0, j_0]$ בדו-מימד) מתקבל כצירוף ליניארי של פיקסלים מהתמונה הנוכחית. מקדמי הצירוף הליניארי הזה יהיו יסתכמו ל-1. למשל, אם הפיקסל ה- n_0 מתקבל על ידי הזזה של הפיקסל ה- n_0+3 , הרי שכל השורה תהיה אפסים ורק האיבר $F_{[n_0, n_0+3]}[t, 1]$ יהיה 1.

בעיקרון, כאשר אנו דנים בייצוג של עיוות גיאומטרי כמטריצה, אנו מניחים ייצוג דיסקרטי של התמונה. לכן, בבניית המטריצה ליישום התנועה הנדרשת מסתתרת גם תפיסת האינטרפולציה שתיושם. למשל, אם נרצה אינטרפולציה השכן הקרוב (Nearest Neighbor), כל שורה ב- F תכיל 1 במקום כלשהו ואפס בותר. בשימוש באינטרפולציה בי-ליניארית נקבל כי כל שורה מכילה 4 איברים אי-שליליים שסכומם 1, ואלו מייצגים את המשקלות של רביעיית הפיקסלים הסמוכים לנקודה אליה עברנו. בכל מקרה, המטריצה המתקבלת דלילה מאוד, דהיינו, מרבית איבריה הם אפסים, ולכן קל לאחסנה ולתמרן איתה אם נחוץ.

9.2 שערך תנועה בין תמונות - גישת בלוקים

נניח כי אנו עוסקים בתמונות בגודל של 512 על 512 פיקסלים. מציאת וקטור התנועה לכל פיקסל בנפרד מציבה בפנינו מספר קשיים אובייקטיביים:

1. כמות הנעלמים גדולה ושווה לפעמיים מספר הפיקסלים. גם אם יוצע תהליך מהיר לחישוב הוקטור לכל פיקסל, ביצוע תהליך זה כל כך הרבה פעמים הופך למשימה כבדה מבחינה חישובית.
2. גם אם איננו חוששים מכמות החישובים, כמות הזיכרון הנדרשת מהווה גורם מפריע. לא די בכך שעלינו לשמור את שתי התמונות הנדונות, גם עלינו לשמור את התוצאה כשני מערכים בגודל התמונה שבמקרה הכללי צפויים להחזיק ערכים ממשיים. במקביל, אם עלינו לשדר לגורם כלשהו שדה וקטורים זה, הרי שכמות המידע הנדרש לשידור עצומה - למעשה - יותר מהתמונה עצמה. לכן, באלגוריתמי דחיסה אין מקום לבניית שדה תנועה מלא המייחס וקטור לכל פיקסל.
3. אם איננו חוששים מחישובים או זיכרון עדיין קיימת בעיה. אם נתבונן בפיקסל $X[k,j,t]$ ונצא לחפש את מקורו בתמונה הקודמת, קרוב לודאי שיתקבלו מספר פתרונות אפשריים. כמו כן, רעש בתמונה ישפיע באופן קיצוני על איכות ההתאמה. איך נברור מתוכם את הפתרון הנכון? בחירה טבעית היא להניח סוג של חלקות על שדה התנועה המתקבל. אם לפיקסל כלשהו יש וקטור תנועה נתון, צפוי כי שכניו ינועו באופן דומה, וזאת מתוך הנחה שהם שייכים לאותו עצם בצילום.

בשל שלושת הסיבות, נוכל להציע את התפיסה הבאה - במקום ליחס וקטור תנועה שונה לכל פיקסל בתמונה, נחלק את התמונה לבלוקים (לדוגמה בגודל של 8 על 8 פיקסלים), ולכל בלוק כזה נייחס וקטור תנועה אחיד. באופן זה נצטרך לחשב ולאגור הרבה פחות וקטורים (בדוגמה שלנו - פי 64 פחות). מעבר לכך, גם נפתור גם את הבעיה השלישית, כיוון שאם נחפש את מיקומו של הבלוק כמקשה אחת בתמונה הקודמת, אנו צפויים לקבל התאמה חסינה יותר לרעש. החסרון של תפיסה זו ברור - אם בבלוק נתון קיימות שתי אוכלוסיות של פיקסלים שנעו באופן שונה - אנו לא נדע מכך וקרוב לודאי שנייחס לכל הבלוק וקטור לא מוצלח.

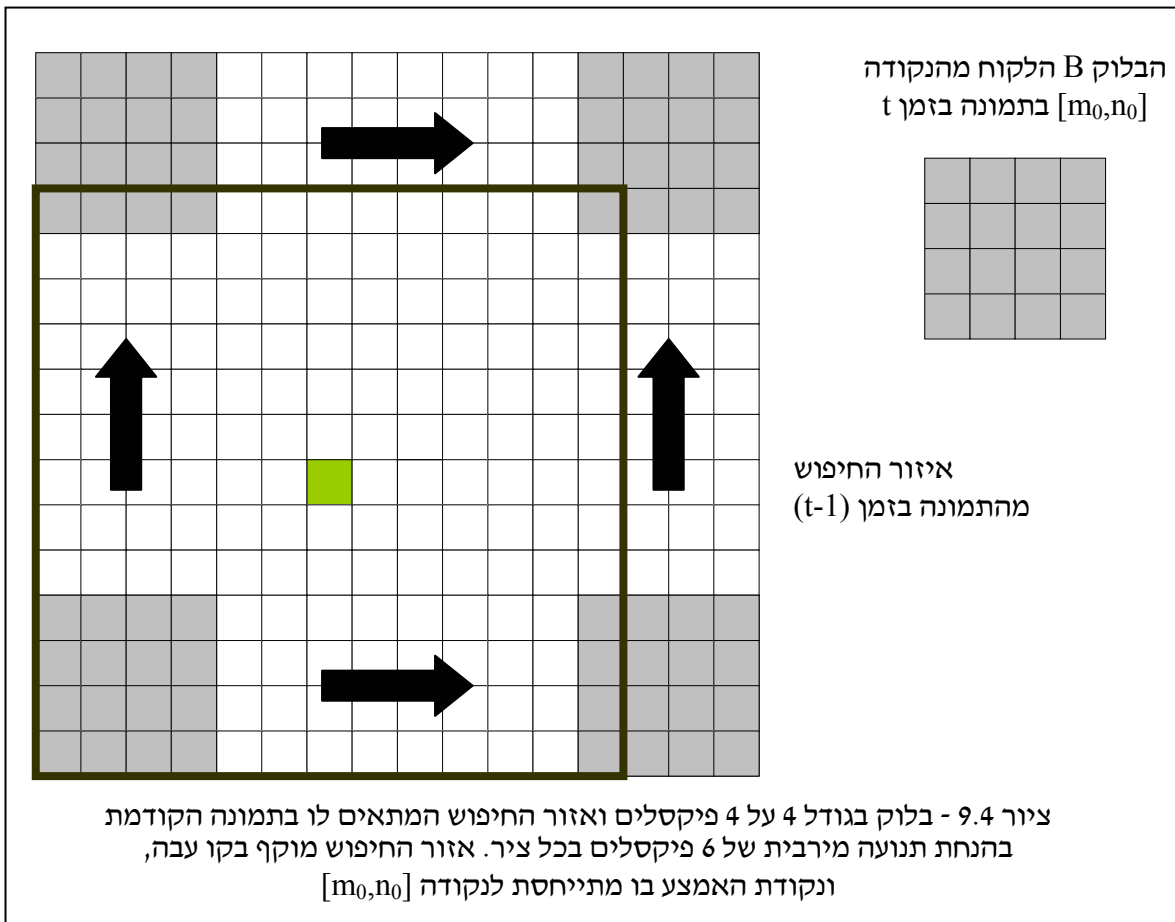
נניח כי בחרנו לעבוד עם בגודל של 512 על 512, וחילקנו אותה ל-64 על 64 בלוקים סמוכים כשכל אחד בגודל של 8 על 8 פיקסלים. לכל בלוק כזה עלינו למצוא וקטור תנועה, והשאלה בה נטפל כעת היא כיצד נעשה זאת. השיטה בה ננקוט היא הבאה - לבלוק נתון מהתמונה $X[k,j,t]$ נבדוק עבור כל מיקום אפשרי בתמונה הקודמת $X[j,k,t-1]$ את מידת התאמתו. מידת ההתאמה תיקבע ע"י החסרתו מהפיקסלים במיקום המוצע, העלאת הפרשים בריבוע וסיכומם. ערך נמוך יעיד על התאמה טובה וערך גבוה על התאמה גרועה. אם נסמן את הבלוק המטופל ב-B, ונניח כי בלוק זה נלקח מהמיקום $[m_0, n_0]$ בתמונה $X[k,j,t]$, כלומר:

$$B[k, j] = X[m_0 + k, n_0 + j, t] \quad 1 \leq k, j \leq 8$$

אנו ננסה להביא למינימום את הפונקציה הבאה:

$$E[m,n] = \sum_{k=1}^8 \sum_{j=1}^8 (B[k,j] - X[m+m_0+k, n+n_0+j, t-1])^2$$

הערכים $[m+m_0, n+n_0]$ מתארים את המיקום ממנו עקרנו את הבלוק המועמד מתוך התמונה הקודמת, ופונקציה זו מודדת בעצם את המרחק בין B לבלוק זה. בפועל, אין צורך לבדוק את שיעור ההתאמה לכל $[m,n]$ בשטח התמונה - יש להתמקד באזור חיפוש סביר, תוך התייחסות להסטה המקסימלית האפשרית. למשל, אם ידוע לנו כי בלוק לא ינוע יותר מ-6 פיקסלים לכל כיוון, הדבר מגדיר אזור חיפוש סביב הנקודה $[m_0, n_0]$ שגודלו 13 על 13 פיקסלים. לכן, נתעניין בערכי $E[m,n]$ עבור ערכי m ו- n בתחום 6- ועד 6+. ציור 9.4 ממחיש רעיון זה עבור בלוקים בגודל 4 על 4.



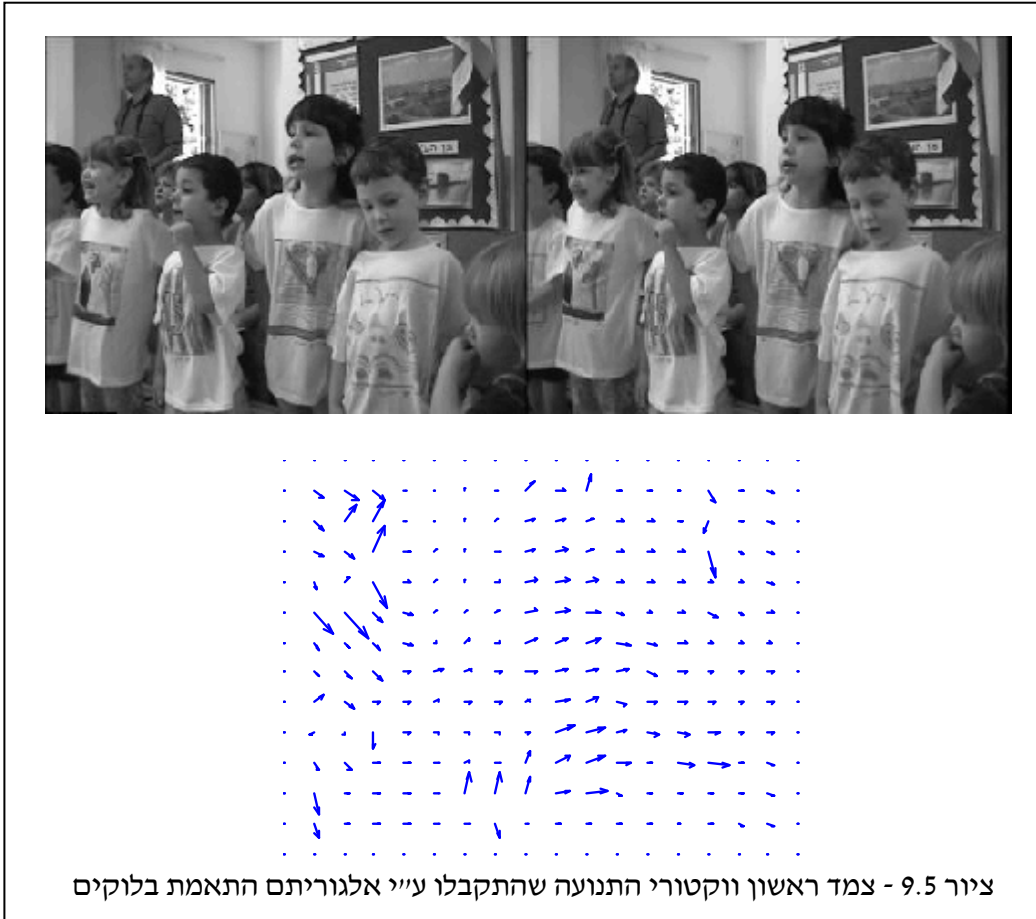
תהליך המינימיזציה דורש חיפוש מייגע - מעבר על כל המיקומים האפשריים (עבור תנועה מקסימלית p יש $(2p+1)^2$ מיקומים כאלה), חישוב הערך E , ומציאת הנמוך שביניהם. מסתבר כי ליישומים רבים גם תהליך זה אינו נחשב כזול מבחינה חישובית, ונדרש אלגוריתם מהיר ממנו. דרך אחת להקל היא לשנות את E לפונקציה הבאה:


$$E[m,n] = \sum_{k=1}^8 \sum_{j=1}^8 |B[k,j] - X[m+m_0+k, n+n_0+j, t-1]|$$

באופן זה המרנו העלאה בריבוע לערך מוחלט - פעולה קלה יותר.

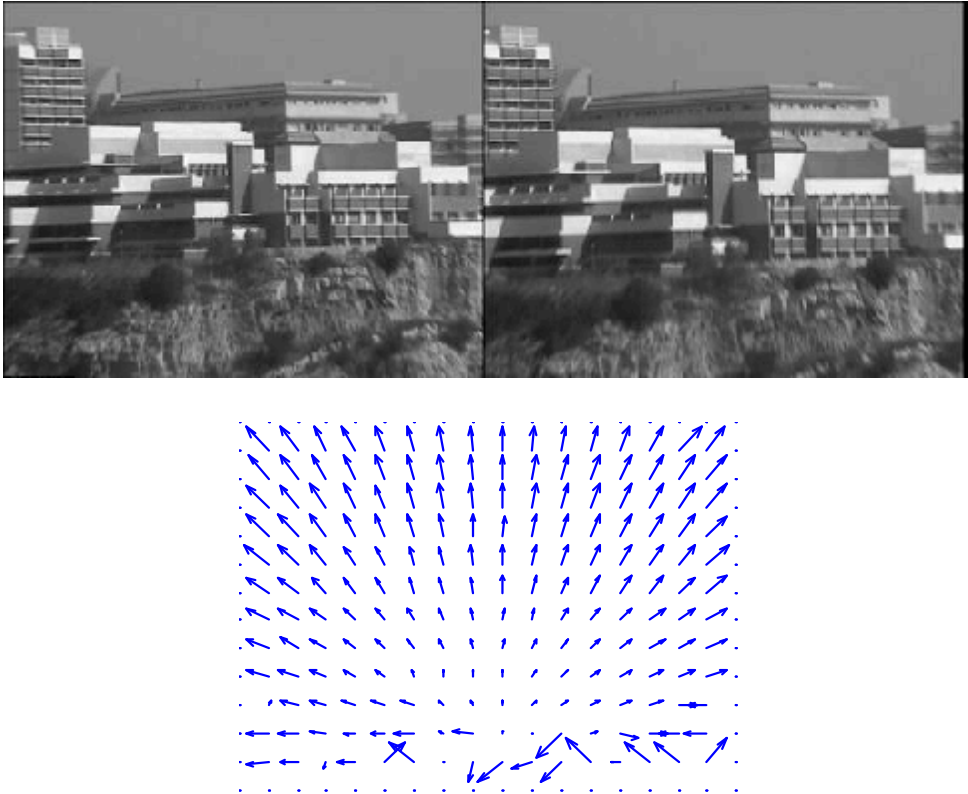
בנוסף, בהנחה שהפונקציה $E[m,n]$ הינה פונקציה קמורה עם מינימה יחידה (כפי שמסתבר שבדרך-כלל מתקבל), ניתן להציע מגוון אלגוריתמים מהירים למציאת נקודת המינימום. אנו לא נציג אלגוריתמים אלו במסגרת זו, אך נאמר כי ברמה עקרונית הם מציעים לחשב את $E[m,n]$ באופן סלקטיבי ולפי התוצאות הקודמות לקבוע את מיקומן של נקודות הבדיקה הבאות.

בציורים 9.5, 9.6 ו-9.7 אנו מראים שוב את צמדי התמונות מציור 9.1 והפעם עם תוצאת שערך התנועה בגישה זו לבלוקים של 16 על 16 ועם הנחת תנועה מירבית של 10 פיקסלים לכל צד. נשים לב כי בצמדים השני והשלישי הייתה תנועה גלובלית של המצלמה, וחיצי התנועה נראים כהזזה בצמד השני וכזום פנימה בצמד השלישי. בצמד הראשון התנועה היא מקומית וכך יוצא שמרבית וקטורי התנועה קרובים לאפס.





ציור 9.6 - צמד שני ווקטורי התנועה שהתקבלו ע"י אלגוריתם התאמת בלוקים



ציור 9.7 - צמד שלישי ווקטורי התנועה שהתקבלו ע"י אלגוריתם התאמת בלוקים

9.3 שערך תנועה בין תמונות - גישה מלאה

החסרון המרכזי בגישה שערך התנועה שתוארה בסעיף קודם היא רמת הדיוק שניתן לצפות ממנה. ברור כי חלוקה לבלוקים מקילה על החישובים ועל צריכת הזיכרון - אך הדבר בא על חשבון דיוק, כיוון שההנחה שפיקסלים בבלוק נעים כמקשה אחת אינה נכונה. התחליף הוא למצוא וקטור תנועה לכל פיקסל, ובוה נעסוק בסעיף זה. המשוואה אותה צריכים וקטורי התנועה לקיים היא המשוואה הבאה:

$$I(x + dx(x, y, t), y + dy(x, y, t), t) = I(x, y, t - 1)$$

אותה כבר ראינו בתחילת פרק זה בגרסה שונה מעט (דיסקרטית, ועם תנועה הפוכה). שתי בעיות מתעוררות בבואנו לעשות שימוש במשוואה זו:

1. הביטוי המתקבל הינו פונקציה מורכבת של הנעלמים.
2. הביטוי כופה אילוץ אחד על שני נעלמים - דבר המעיד כי חסר מידע בבואנו לקבוע את וקטורי התנועה.

על מנת לפתור את הבעיה הראשונה מוצעת הגישה הדיפרנציאלית, אשר מהווה כיום את הרעיון המרכזי לעשרות אלגוריתמי שערך תנועה. הרעיון פשוט בעיקרו - נניח כי התנועה $[dx(x,y,t), dy(x,y,t)]$ קטנה יחסית (כלומר, מצויה בתחום $[-1, +1]$ פיקסלים). כיוון שכך, נוכל לפתח את אגף שמאל של המשוואה לטור טיילור סביב הנקודה (x, y, t) ולהזניח איברים מסדר גדול מ-1. מתקבל:

$$\begin{aligned} I(x, y, t - 1) &= I(x + dx(x, y, t), y + dy(x, y, t), t) = \\ &= I(x, y, t) + \frac{\partial I(x, y, t)}{\partial x} dx(x, y, t) + \frac{\partial I(x, y, t)}{\partial y} dy(x, y, t) = \\ &= I(x, y, t) + I_x(x, y, t) dx(x, y, t) + I_y(x, y, t) dy(x, y, t) \end{aligned}$$

קיבלנו משוואה ליניארית בנעלמים שלנו $[dx(x,y,t), dy(x,y,t)]$, ומקדמי משוואה זו ידועים לנו. אנו מקבלים את שתי התמונות ויכולים לחשב גם נגזרת אופקית ואנכית של התמונה $I(x,y,t)$. נשים לב כי הגורם החופשי הוא ההפרש בין התמונות העוקבות, וניתן להתייחס אלי כנגזרת זמנית. פתרנו אם כך את הבעיה הראשונה אותה הצגנו. המשוואה שהתקבלה קוריה "משוואת אילוץ הבהירות" (Brightness - Constrained Equation - BCE).

במקרה הדיסקרטי עלינו להשתמש בקירוב הנגזרות כפי שהוצג בפרק 5. עבור המקרה הדיסקרטי רצוי מאוד לבצע החלקה מקדימה על שתי תמונות הכניסה, על מנת לדחות רעש אשר עלול להזיק בחישובי הנגזרות, ועל מנת להפוך את ה-BCE למדויקת יותר. כאשר האזור חלק יותר, הנגזרות הגבוהות של התמונה חלשות יותר ואז ההזנחה שלנו בטור טיילור הקצוץ נכונה יותר.

נציג כעת שני אלגוריתמים לשערוך תנועה המשתמשים ב- BCE בבואם לשערך את התנועה. האלגוריתם הראשון הוצע ע"י Lucas ו- Kanade, והוא מציע מוטיב הדומה לתהליך השערוך בבלוקים אותו ראינו בסעיף קודם. נניח כי רצוננו לשערך את התנועה לפיקסל $[k_0, j_0]$ (והתהליך שיתואר יבוצע לכל פיקסל באופן דומה). נפתח חלון של 5 על 5 פיקסלים סביב פיקסל זה ונניח כי לכל הפיקסלים יש אותה תנועה, אותה נסמן $[dx[k_0, j_0, t], dy[k_0, j_0, t]]$. בכל פיקסל יש לנו משוואה ליניארית אחת בה שני ערכים אלו הם הנעלמים - נרכז את 25 המשוואות שמתקבלות ונמצא את הפתרון הטוב ביותר במובן LS, דהיינו, אם נרכז את כל המשוואות באופן הבא:

$$A = \begin{bmatrix} I_x[k_0 - 2, j_0 - 2, t] & I_y[k_0 - 2, j_0 - 2, t] \\ \vdots & \vdots \\ I_x[k_0, j_0, t] & I_y[k_0, j_0, t] \\ \vdots & \vdots \\ I_x[k_0 + 2, j_0 + 2, t] & I_y[k_0 + 2, j_0 + 2, t] \end{bmatrix} \quad V = \begin{bmatrix} dx[k_0, j_0, t] \\ dy[k_0, j_0, t] \end{bmatrix}$$

$$B = \begin{bmatrix} I[k_0 - 2, j_0 - 2, t-1] - I[k_0 - 2, j_0 - 2, t] \\ \vdots \\ I[k_0, j_0, t-1] - I[k_0, j_0, t] \\ \vdots \\ I[k_0 + 2, j_0 + 2, t-1] - I[k_0 + 2, j_0 + 2, t] \end{bmatrix}$$

$$\Rightarrow AV = B$$

אנו נגדיר את וקטור התנועה V להיות:

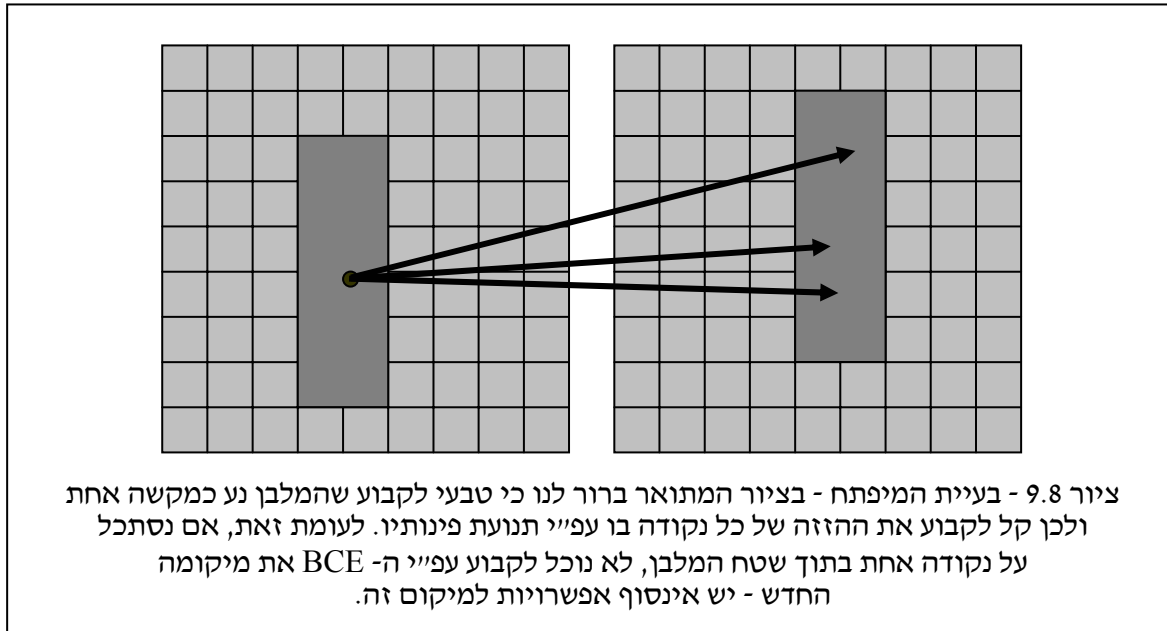
$$V_{\text{opt}} = \underset{V}{\text{ArgMin}} \|AV - B\|^2 = (A^T A)^{-1} A^T B$$

וכך, על-ידי ערכי התמונות ונגזרותיהן נקבל את וקטור התנועה. המטריצה אותה יש להפוך כאן היא בגודל של 2 על 2, ולכן מורכבות חישובית אינה מטרידה כאן. שיפור מסוים לאלגוריתם יכול להתקבל אם נכניס משקל לכל משוואה לפי מרחקה מהפיקסל המרכזי.

מסתבר כי אלגוריתם זה מניב תוצאות טובות מאוד בדרך כלל, כל עוד התנועה קטנה כפי שהנחנו. כאשר היא אינה קטנה, יש להפעיל את האלגוריתם, לקבל שדה תנועה, לעוות את התמונה $I(x, y, t-1)$ אל $I(x, y, t)$ לפי התנועה שהתקבלה ולחזור על התהליך. באופן זה התמונות מתקרבות זו אל זו בכל איטרציה והתנועה היא התנועה המצטברת.

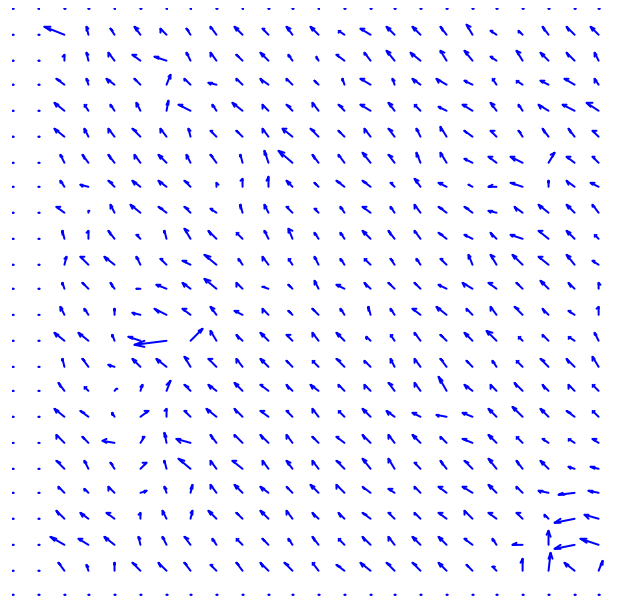
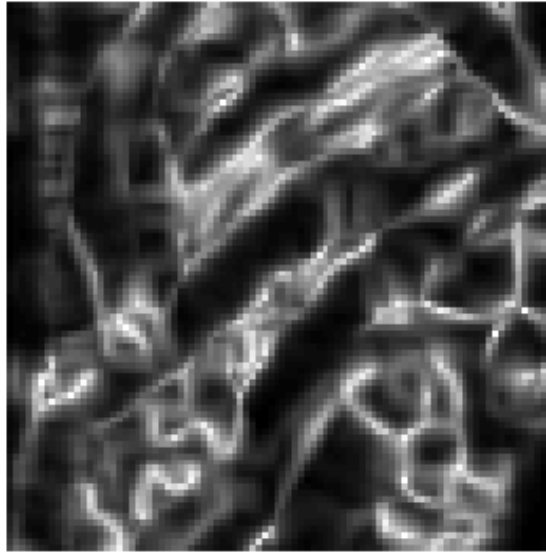
תופעה מעניינת המתקבלת באלגוריתם זה כמוצר-לוואי היא שלכל פיקסל אנו יכולים לקבל מדד לאמינות הוקטור ששוערך עבורו. לצורך חישוב הוקטור עלינו להפוך את המטריצה $A^T A$, ואם מטריצה זו לא הפיכה ברור כי אין פתרון. מצב זה מתקבל כאשר כל השורות ב- A תלויות ליניארית זו בזו - למשל - אם כל השורות זהות. מצב זה מוכר בשם "תופעת המיפתח" (Aperture Problem). ציור 9.8 ממחיש את

התופעה, ורואים כי כשמסתכלים דרך "חור המנעול" על נתח תמונה, קשה לקבוע את מיקומה החדש אם איננו עושים שימוש בקונטקסט הסביבתי שלה. באופן זה תיארו מצב קיצוני בו וקטור התנועה לא ניתן לחישוב. במצבים פחות קיצוניים, המטריצה הנהפכת אמנם אינה סינגולרית, אך קרובה לכך. מדד יפה למידת בריאותה הוא ה- Condition Number הקובע את יחס הערכים העצמיים - הקטן חלקי הגדול. אם יחס זה קטן מסף מסוים נוכל לקבוע כי הוקטור המתקבל אינו כל כך אמין.



ציור 9.10 מראה צמד תמונות עם הזזה מלאכותית מדויקת של פיקסל אחד שמאלה ולמעלה. ציור זה גם מראה כיצד נראות תמונות הנגזרות האופקית, האנכית והזמנית. ציור 9.11 מראה את תוצאת השערוך של התנועה לפי אלגוריתם Lukas & Kanade. בתמונת האמינויות רואים כי אזורי שפה מניבים שערוך טוב יותר.





ציור 9.11 - תוצאות שערך תנועה עפ"י אלגוריתם Lukas & Kanade. למעלה - מפת האמינויות של הוקטורים בה ערכים בהירים יותר מעידים על אמינות גדולה יותר. למטה - שדה התנועה המשוער

בדוגמה זו בה התנועה היא $[-1,+1]$, הוקטור הממוצע בשדה הוקטורים המשוער הוא $[-0.919,0.937]$. ע"י שינויים קלים כגון הגדלת חלון העבודה, הכנסת משקלות לאיברי החלון, ושינוי דרך ביצוע הנגזרת לגישה מדויקת יותר, ניתן להגיע לדיוק טוב יותר.

אלגוריתם שני לביצוע שרורך התנועה מורכב יותר אך יפה יותר, והוא פותח על ידי Horn ו-Schunck. אלגוריתם זה מציע להתייחס לשדה התנועה כולו כנעלם בבעיית אופטימיזציה ולמוצא את כולו במקביל. נסמן בוקטור \underline{U} את כל קואורדינטות ה- x של שדה התנועה שאינו ידוע, וב- \underline{V} את קואורדינטות ה- y . אלה וקטורים המתקבלים מסידור לקסיקוגרפי של תמונות התנועה האופקית והאנכית, ממש כשם שאנו עושים זאת לתמונות רגילות. משוואות ה-BCE מציעות את הקשר הבא:

$$\begin{bmatrix} I_x[1,1,t] & & 0 & I_y[1,1,t] & & 0 \\ & \ddots & & & \ddots & \\ 0 & & I_x[M,N,t] & 0 & & I_y[M,N,t] \end{bmatrix} \begin{bmatrix} \underline{U} \\ \underline{V} \end{bmatrix} = \begin{bmatrix} I_t[1,1,t] \\ \vdots \\ I_x[M,N,t] \end{bmatrix}$$

$$\Rightarrow [H_x, H_y] \begin{bmatrix} \underline{U} \\ \underline{V} \end{bmatrix} = \underline{Y}$$

כאשר הנחנו כי התמונות הן בגודל N על M . למעשה הקשר הנ"ל הוא פשוט ריכוז של כל משוואות אילוץ הבהירות לכל הפיקסלים לביטוי אחד. יש כאן MN משוואות עם $2MN$ נעלמים. נשתמש כעת בתכונת החלקות הצפויה לשדה התנועה. אם נפעיל על "התמונה" U גזירה כגון לפלסיאן, נרצה כי נורמת התוצאה תהיה נמוכה. באופן זה נכפה חלקות על U ועל V , ממש כשם שעשינו זאת בדיון בשחזור תמונות. לכן, פונקציית המחיר אותה נמוזער היא:

$$\{\hat{\underline{U}}_{opt}, \hat{\underline{V}}_{opt}\} = \underset{\underline{U}, \underline{V}}{\text{ArgMin}} \left[\|H_x \underline{U} + H_y \underline{V} - \underline{Y}\|^2 + \lambda [\|D\underline{U}\|^2 + \|D\underline{V}\|^2] \right] =$$

$$\Rightarrow \begin{bmatrix} \hat{\underline{U}}_{opt} \\ \hat{\underline{V}}_{opt} \end{bmatrix} = \begin{bmatrix} H_x H_x + \lambda D^T D & H_x H_y \\ H_y H_x & H_y H_y + \lambda D^T D \end{bmatrix}^{-1} \begin{bmatrix} H_x \underline{Y} \\ H_y \underline{Y} \end{bmatrix}$$

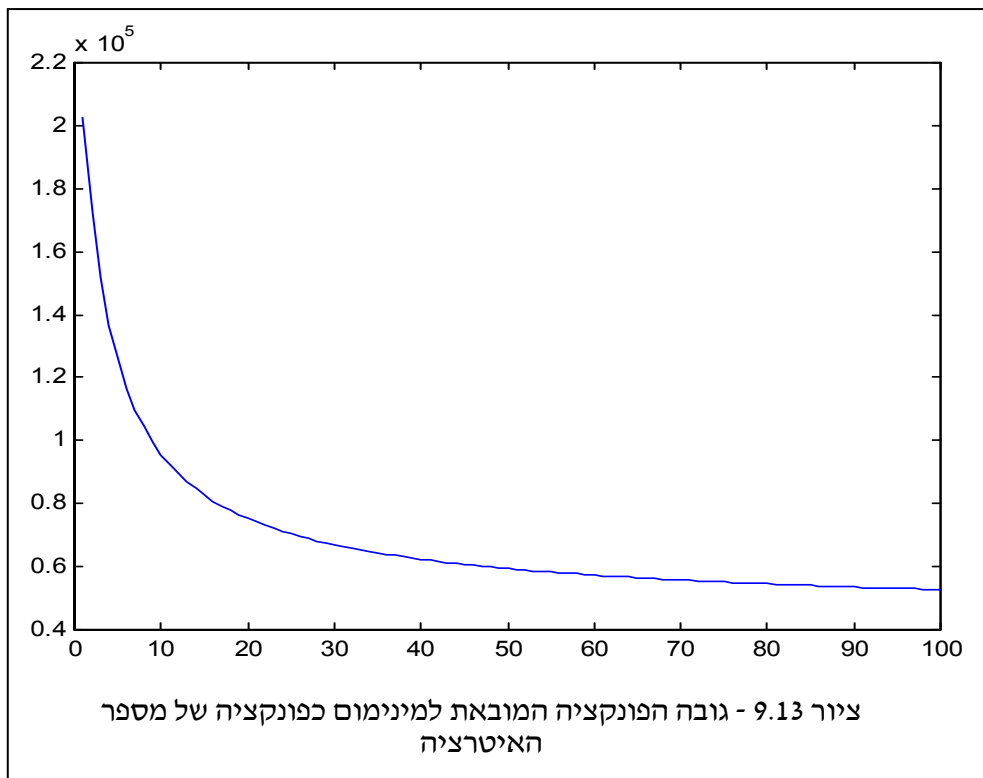
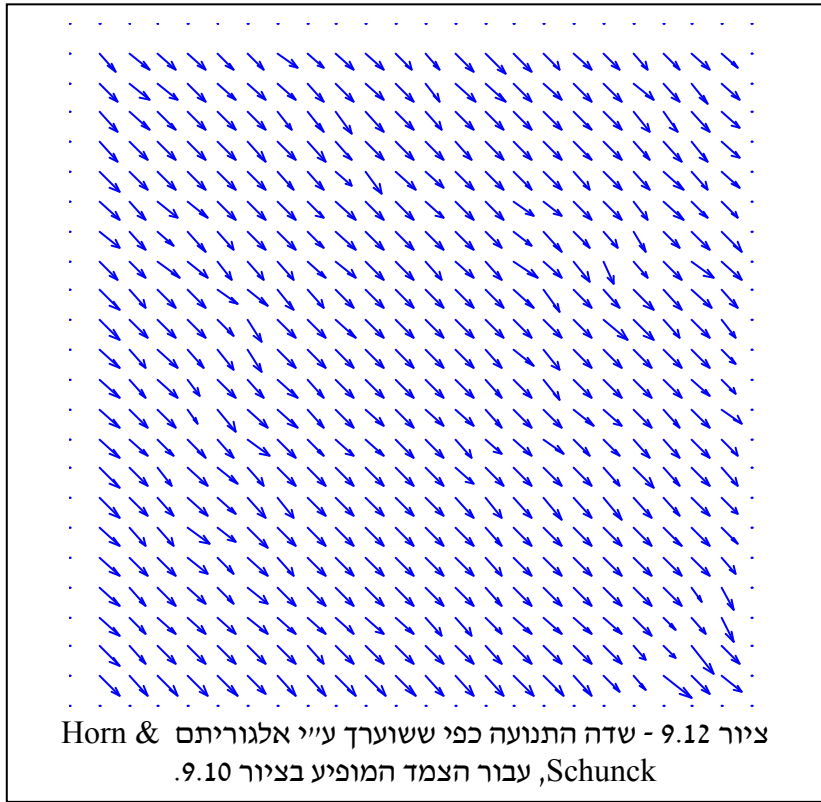
הגודל λ שולט במידת החלקות הנכפית על הפתרון - עבור ערך גבוה מאוד נקבל שדה תנועה מאוד חלק, ועבור ערך נמוך נרשה שינויים מרחביים בוקטורים. כמו בשחזור תמונות, היפוך המטריצה המתקבל אינו בא בחשבון, וכתחליף מוצעת גישה איטרטיבית לפי ה-SD, בה מתחילים בניחוש (נניח אפסים), ומעדכנים הפתרון לפי:

$$\begin{bmatrix} \hat{\underline{U}} \\ \hat{\underline{V}} \end{bmatrix}_{k+1} = \begin{bmatrix} \hat{\underline{U}} \\ \hat{\underline{V}} \end{bmatrix}_k - \mu \left\{ \begin{bmatrix} H_x H_x + \lambda D^T D & H_x H_y \\ H_y H_x & H_y H_y + \lambda D^T D \end{bmatrix} \begin{bmatrix} \hat{\underline{U}} \\ \hat{\underline{V}} \end{bmatrix}_k - \begin{bmatrix} H_x \underline{Y} \\ H_y \underline{Y} \end{bmatrix} \right\}$$

גם באלגוריתם זה, אין צורך לבנות מטריצות או לייצג את הנעלמים בפועל כוקטורים. הפעולות המתבצעות במשוואה זו על תמונות הנעלמים U ו- V פשוטות - הפעלת לפלסיאן כקונבולוציה, והכפלה במטריצות אלכסוניות כפירושו הכפלת התמונות U ו- V איבר-איבר באיברי האלכסון הראשי.

פרק 9 - עיבוד סדרות של תמונות -

בציור 9.12 מובאות תוצאות של אלגוריתם זה לאותו צמד כבתמונות 9.10 ו- 9.11. בחרנו $\lambda = 100$ וכך נכפתה חלקות חזה על הוקטורים. הוקטור הממוצע הוא $[-0.9952, 0.9921]$. בציור 9.13 מובא הגרף המראה את גובה הפונקציה שהובאה למינימום כפונקציה של מספר האיטרציה.



9.4 ניקוי רעש בסדרות של תמונות

נעבור כעת לדון באפליקציות העושות שימוש בשערוך תנועה. במסגרת סעיף זה נראה כיצד מנקים סידרת תמונות מרעש אדיטיבי גאוס וכן לבן. לצורך יישום זה נדרש להשתמש בשערוך תנועה מדויק ככל האפשר, ולכן אנו נניח כי אלגוריתם כדוגמת האלגוריתמים הדיפרנציאליים משמש אותנו.

נתחיל את הדיון בקביעת סימונים. סידרת התמונות המקורית (ולא רעש מסומנת ב- $\underline{X}(t)$), בה כל איבר הוא וקטור המתאר תמונה בגודל M על N פיקסלים. הסרט הפגום הוא $\underline{Y}(t)$, אשר גם הוא בנוי מתמונות בגודל M על N , אך קיים הקשר הבא:

$$\underline{Y}(t) = \underline{X}(t) + \underline{N}(t)$$

כלומר, התמונה $\underline{Y}(t)$ היא גרסה רועשת של $\underline{X}(t)$. הרעש כאמור לבן, וזה אומר כי ערכיו במקום ובזמן מפולגים באופן זהה (פילוג גאוס) וחסר תלות. בהגיע התמונה $\underline{Y}(t)$, רצוננו להפיק שערוך מוצלח ונקי מרעש שיסומן $\hat{\underline{X}}(t)$. אנו נניח מערכת סיבתית, כלומר מערכת שבה $\hat{\underline{X}}(t)$ נבנה על-פי מידע שכבר התקבל ולא עפ"י תמונות עתידיות. הנחה אחרונה היא שאנו יכולים לשערך את התנועה בין כל צמד תמונות (ואפילו רועשות) בדיוק מעולה. לכן, כשנשערך את התנועה בין הצמד $\{\underline{X}(t), \underline{X}(t-k)\}$ נקבל אותה תוצאה כמו עבור הצמד $\{\underline{Y}(t), \underline{Y}(t-k)\}$. פעולת העיוות הגיאומטרי הזו תסומן ב- $G(t, k)$ - זוהי מטריצה שמשמעותה התנועה שיש להעביר את $\underline{X}(t-k)$ על מנת לקבל את $\underline{X}(t)$, וכבר פגשנו את הפעולה ההפוכה לה F בתחילת פרק זה. לכן קיים הקשר:

$$\underline{X}(t) = G(t, k)\underline{X}(t-k)$$

השוויון במשוואה זו אינו מדויק כיוון שככל ש- k גדול יותר כך יותר מידע חדש מוזן לתמונה עם הזמן ולכן זהו שוויון מקורב, ואיכותו נופלת עם הגדלת k . לו עבדנו עם תמונה בודדה, ורצינו לנקות רעש לפי המתודולוגיה של פרק 6 - שחזור תמונות - היינו מציעים את המינימיזציה של הפונקציה הבאה:

$$\hat{\underline{X}}_{\text{Intra}}(t) = \underset{\underline{X}(t)}{\text{ArgMin}} \|\underline{Y}(t) - \underline{X}(t)\|^2 + \lambda \|\underline{DX}(t)\|^2$$

כאשר האיבר הראשון בפונקציה המחיר כופה קירבה ל- \underline{Y} , והאיבר השני כופה חלקות מרחבית ע"י פעולת גזירה כלשהי. הפתרון קרוי Intra כיוון שהוא אינו פועל בין תמונות (Inter-frame) אלא בתוכן (Intra-frame). בשל כך גם אין כל שימוש בתנועה בינתיים. נניח כי אנו רוצים לצרף למשחק את התמונה הקודמת. נוכל להציע את הפונקציה הבאה:

$$\hat{\underline{X}}_{In}(t) = \underset{\underline{X}(t)}{\text{ArgMin}} \|\underline{Y}(t) - \underline{X}(t)\|^2 + \alpha \|G(t,1)\underline{Y}(t-1) - \underline{X}(t)\|^2 + \lambda \|D\underline{X}(t)\|^2$$

כעת כפינו על הפתרון $\underline{X}(t)$, לאחר עיוות נאות, להיות קרוב לתמונה הקודמת $\underline{Y}(t-1)$. המקדם α יהיה מספר כלשהו קטן במקצת מ-1 הקובע את חשיבות הקרבה החדשה ביחס לקרבה לתמונה הנוכחית. נוכל באופן זה להציע ביטוי מורכב יותר אשר יביא בחשבון את כל התמונות בסדרה מזמן $t=0$. ביטוי זה הוא:

$$\hat{\underline{X}}(t) = \underset{\underline{X}(t)}{\text{ArgMin}} \sum_{k=0}^t \alpha^k \|G(t,k)\underline{Y}(t-k) - \underline{X}(t)\|^2 + \lambda \|D\underline{X}(t)\|^2$$

נשים לב כי כל איבר הוכפל במשקל α^k כך שהתמונות הקרובות תקבלנה משקל גדול, והרחוקות משקל הדועך לאפס. נפתור בעיית אופטימיזציה זו ונראה כי פתרונה אינו מורכב. גזירה והשוואה לאפס נותנים:

$$2 \sum_{k=0}^t \alpha^k [\underline{X}(t) - G(t,k)\underline{Y}(t-k)] + 2\lambda D^T D \underline{X}(t) = 0$$

$$\left[\sum_{k=0}^t \alpha^k I + \lambda D^T D \right] \hat{\underline{X}}(t) = \left[\frac{1 - \alpha^{t+1}}{1 - \alpha} I + \lambda D^T D \right] \hat{\underline{X}}(t) = \sum_{k=0}^t \alpha^k G(t,k)\underline{Y}(t-k)$$

לכאורה נראה כי עלינו לזכור את כל התמונות בסדרה, מנקודת הזמן 0 ועד לזמן t , לשם מציאת התמונה הנוכחית במוצא. בפועל נוכל להשתמש בקשר

$$G(t,k) = G(t,1)G(t-1,1) \cdots G(t-k+2,1)G(t-k+1,1)$$

ולקבל

$$\begin{aligned} \underline{P}(t) &= \sum_{k=0}^t \alpha^k G(t,k)\underline{Y}(t-k) = \sum_{k=0}^t \alpha^k G(t,1) \cdots G(t-k+1,1)\underline{Y}(t-k) = \\ &= \underline{Y}(t) + \alpha G(t,1) \sum_{k=1}^t \alpha^k G(t-1,1) \cdots G^T(t-k+1,1)\underline{Y}(t-k) = \\ &= \underline{Y}(t) + \alpha G(t,1) \sum_{j=0}^{t-1} \alpha^j G(t-1,1) \cdots G^T(t-j,1)\underline{Y}(t-j) = \\ &= \underline{Y}(t) + \alpha G(t,1)\underline{P}(t-1) \end{aligned}$$

כלומר, על מנת לחשב את $\hat{\underline{X}}(t)$ עלינו לקחת את התמונה $\underline{P}(t-1)$ (שנניח שמצויה בידינו), לעדכנה לפי המשוואה הרקורסיבית הנ"ל, ולפתור את מערכת המשוואות

$$\left[\frac{1}{1-\alpha} I + \lambda D^T D \right] \hat{\underline{X}}(t) \cong \underline{P}(t)$$

השוויון בערך נובע מכך שהנחנו כי t גדול דיו ולכן $\alpha^t \cong 0$. מערכת משוואות זו יכולה להיפתר באופן איטרטיבי ע"י אלגוריתם ה-SD הצועד בכיוון הגרדיאנט:

$$\hat{X}_{k+1}(t) = \hat{X}_k(t) - \frac{\mu}{1-\alpha} \hat{X}_k(t) - \lambda D^T D \hat{X}_k(t) + \mu \underline{P}(t)$$

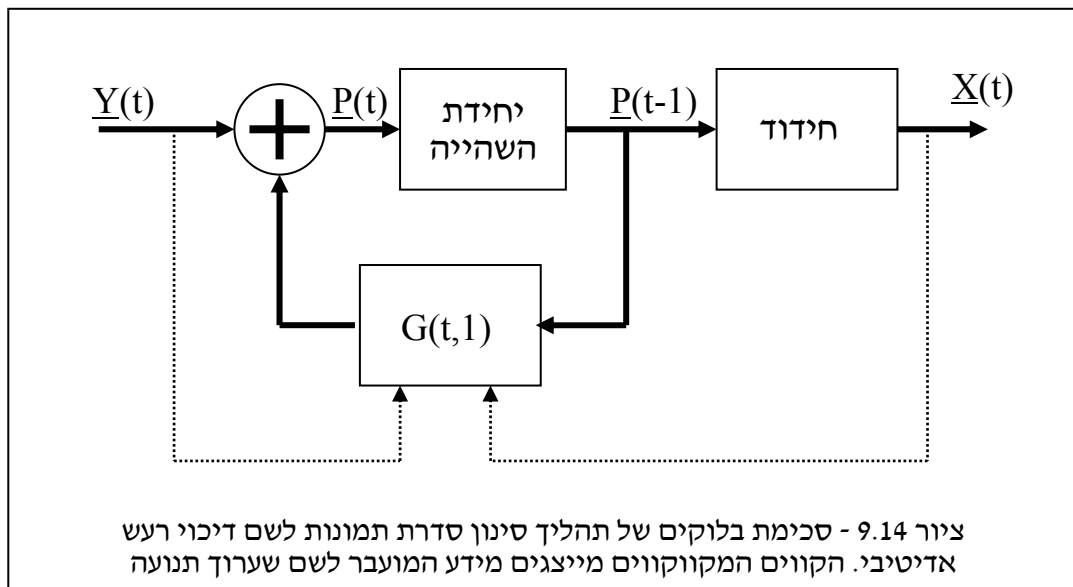
כתחליף, כיוון שהמטריצה הנהפכת הינה מטריצה בלוק-סיבובי, ניתן להפעילה בהיפוך כמסנן ליניארי קבוע במקום.

הוקטור $\underline{P}(t)$ אינו אלא ממוצע זמני של התמונות בסדרה, כאשר בממוצע זה ישנן שתי תופעות חשובות - (I) קיים משקל דועך אקספוננציאלית לתמונות רחוקות, ו-(II) התמונות מקוזזות תנועה בטרם תיכנסנה לממוצע. התהליך האיטרטיבי נועד לבצע חידוד על תמונת ממוצע זה. אם נקבע $\lambda = 0$, כלומר ללא מרכיב של חלקות מרחבית בפתרון, יתקבל כי:

$$\begin{aligned} \hat{X}(t) &= (1-\alpha)\underline{P}(t) = (1-\alpha)\underline{Y}(t) + (1-\alpha)\alpha G(t,1)\underline{P}(t-1) \\ \Rightarrow \hat{X}(t) &\cong (1-\alpha)\underline{Y}(t) + \alpha G(t,1)\hat{X}(t-1) \end{aligned}$$

קיבלנו כי $\hat{X}(t)$ מתקבל ע"י משוואה רקורסיבית מסדר ראשון - מסנן בעל קוטב יחיד בזמן.

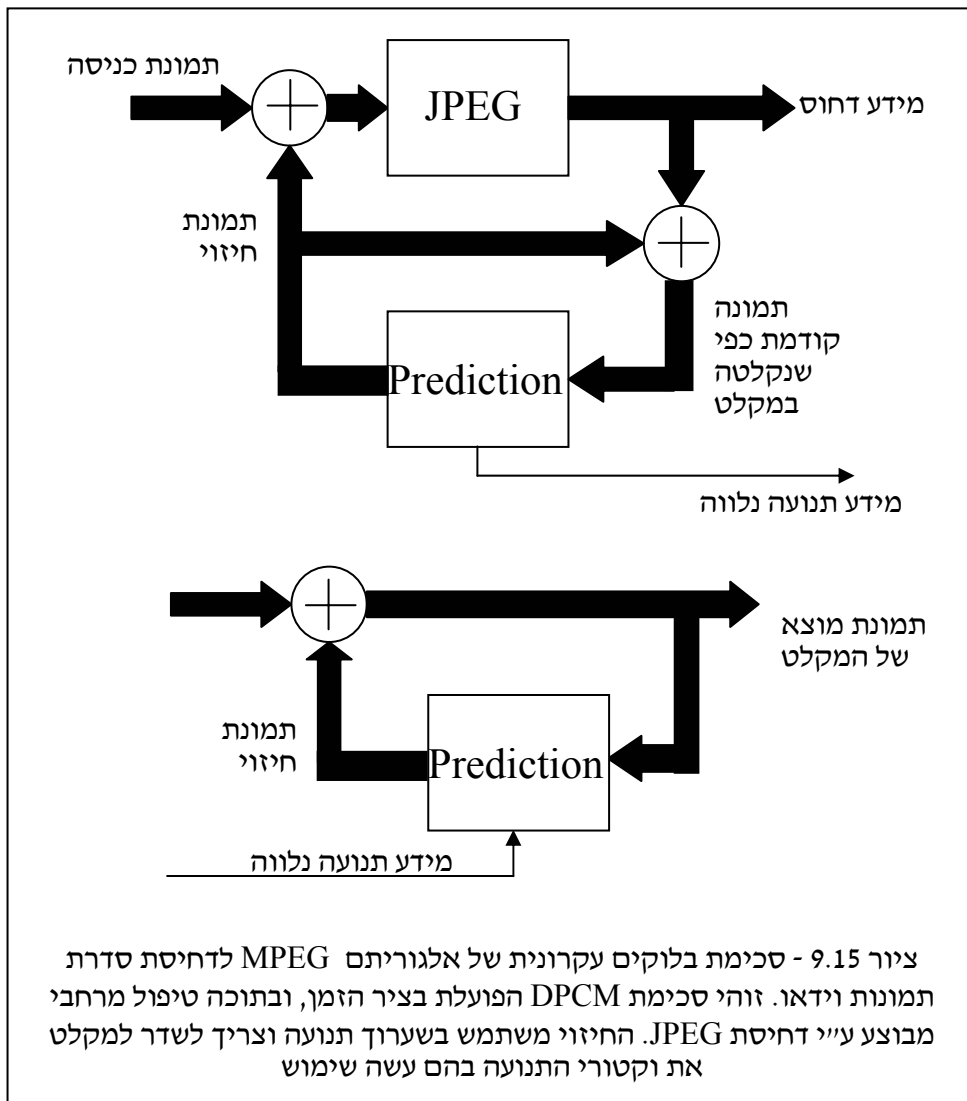
נקודה חשובה בתהליך זה היא דיוק התנועה בה נעשה שימוש. על המטריצה G לייצג תנועה מדויקת בדיוק תת-פיקסלי, אם נרצה תוצאה באיכות טובה. לכן רצוי לבצע את שערך התנועה בין $\hat{X}(t-1)$ ו- $\underline{Y}(t)$, ולא בין $\underline{Y}(t-1)$ ל- $\underline{Y}(t)$. ציור 9.14 מראה סכימת בלוקים של פעולת הניקוי לסדרות של תמונות.



9.5 דחיסת סדרות של תמונות ומבוא ל-MPEG

בסעיף זה נדון בדחיסה של סדרות של תמונות כאשר הדחיסה נעשית On-Line, דהיינו, על כל תמונה בהגיעה. זאת בניגוד לדחיסה של סדרת תמונות כמידע תלת-מימדי השמור על הדיסק כמקשה אחת. כזכור, דחיסת מידע מצריכה ניצול של היתריות באות הנדחס. יתירות בסדרת תמונות הינה יתרות הסתברותי, יתרות מרחבית, ולראשונה גן - יתרות זמנית, שהרי כבר אמרנו שתמונות עוקבות כמעט זהות למעט תנועה.

במסגרת סעיף זה נציג אלגוריתם דחיסה אחד מיני רבים אפשריים. ייחודו של אלגוריתם דחיסה זה בפשטותו היחסית וביצועיו הטובים. למעשה, אלגוריתם זה הוא הגרעין המרכזי באלגוריתם ה-MPEG המהווה כיום את הסטנדרט לדחיסת וידאו. הרעיון המרכזי באלגוריתם זה הוא ניצול התהליך שהוגדר באלגוריתם ה-JPEG, והכללתו לכך שיעבוד על סדרות של תמונות. הרעיון הוא להשתמש בחיזוי לצורך מיצוי היתרות הזמנית, והתמרה לצורך מיצוי היתרות המרחבית. סכימת בלוקים של אלגוריתם זה מתוארת בציור 9.15.



כפי שניתן לראות, הסכימה היא סכימת DPCM קלאסית בה הערוצים מעבירים תמונות שלמות. בהגיע התמונה העכשווית $X(t)$, נבנית עבורה תמונת חיזוי אותה נסמן ב- $\tilde{X}(t)$. תמונה זו נבנית ע"י לקיחת התמונה הקודמת כפי שנקלטה במקלט, $\hat{X}(t)$, וביצוע קיזוז תנועה עליה לפי תנועה ששוערכה בין שתי התמונות האחרונות. וקטורי התנועה לא מצויים במקלט, ועל מנת שהמקלט ישתמש באותו חיזוי בדיוק עלינו לשדר את וקטורי התנועה. לשם כך רצוי לעשות שימוש בשערוך תנועה אחד לבלוק של פיקסלים, בו יש מעט וקטורים לשדר. בשל כך, תמונת החיזוי נבנית באופן הבא:

1. חלק את התמונה הנוכחית $X(t)$ לבלוקים בגודל של 16 על 16 פיקסלים הממלאים את שטח התמונה.
 2. בנה תמונת חיזוי שתסומן $\tilde{X}(t)$ - נאתחל תמונה זו כתמונת אפסים בגודל של $X(t)$.
 3. לכל בלוק מהבלוקים הנ"ל חפש את מיקומו בתמונה הקודמת $\hat{X}(t-1)$. בהנחה כי נמצא המיקום, קח את הבלוק המתאים מהתמונה $\hat{X}(t-1)$ והעתק אותו למיקום הבלוק ב- $\tilde{X}(t)$. הסיבה לכך שאנו עובדים עם $\hat{X}(t-1)$ ולא עם $X(t-1)$ (התמונה המקורית) היא שרצוננו לקבל אותה תמונת חיזוי במקלט.
 4. באופן זה אנו מטפלים בכל בלוק וממלאים את כל שטח התמונה $\tilde{X}(t)$. התוצאה היא תמונת חיזוי המנסה להיות קרובה ככל האפשר לתמונה $X(t)$ - זוהי בדיוק מהותה של תמונת חיזוי.
- לאחר בניית החיזוי $\tilde{X}(t)$, מחוסרת תמונה זו מהתמונה המקורית ומקבלים את תמונת שגיאת החיזוי:

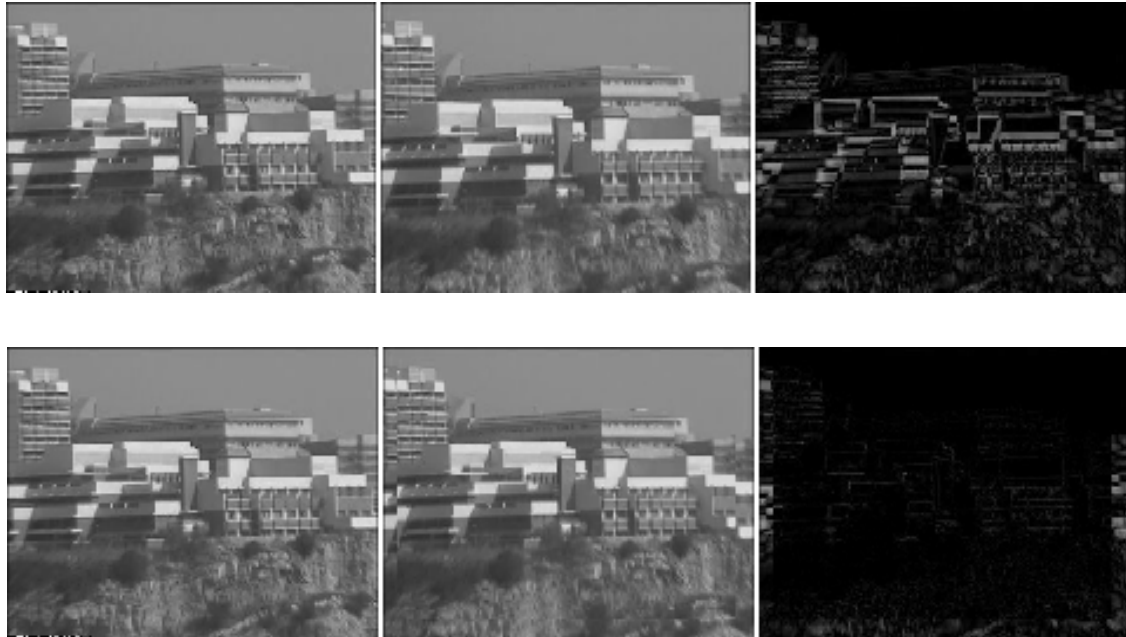
$$E(t) = X(t) - \tilde{X}(t)$$

תמונת הפרש זו אמורה להיות חלשה ומרוכזת סביב האפס. אם בסכימת ה- DPCM המקורית שגיאת החיזוי הייתה סקלאר ועברה קוונטיזציה, הרי שגם כאן אנו נזקקים לקוונטיזציה, אך כיוון ש- $E(t)$ תמונה ויש בה עדיין יתרות מרחבית, רצוי לנצלה. זאת עושים ע"י דחיסה המזכירה את אלגוריתם ה- JPEG - התמונה $E(t)$ מחולקת לבלוקים, כל בלוק עובר התמרת DCT דו-מימדית, ומקדמי ההתמרה עוברים קוונטיזציה לפי חשיבותם. התוצאה בשלב זה היא התמונה $\hat{E}(t)$ - אשר אמורה להיות דומה לתמונה $E(t)$ למעט המידע שנזרק.

במקלט מתקבלת התמונה $\hat{E}(t)$ ואליה מחוברת תמונת החיזוי $\tilde{X}(t)$ - זוהי אותה תמונת חיזוי כבמשדר, כיוון שהיא נבנית עפ"י אותם וקטורי תנועה ועפ"י התמונה $\hat{X}(t-1)$. התוצאה היא:

$$\hat{X}(t) = \hat{E}(t) + \tilde{X}(t) = E(t) + \tilde{X}(t) - [E(t) - \hat{E}(t)] = X(t) - [E(t) - \hat{E}(t)]$$

ושוב קיבלנו כי שגיאת סכימת ה- DPCM אינה אלא זו שמוכנסת ע"י הקוונטיזציה בבלוק ה- JPEG. ציור 9.16 מראה צמד תמונות, ההפרש ביניהן, תמונת החיזוי הנבנית ושגיאת החיזוי.

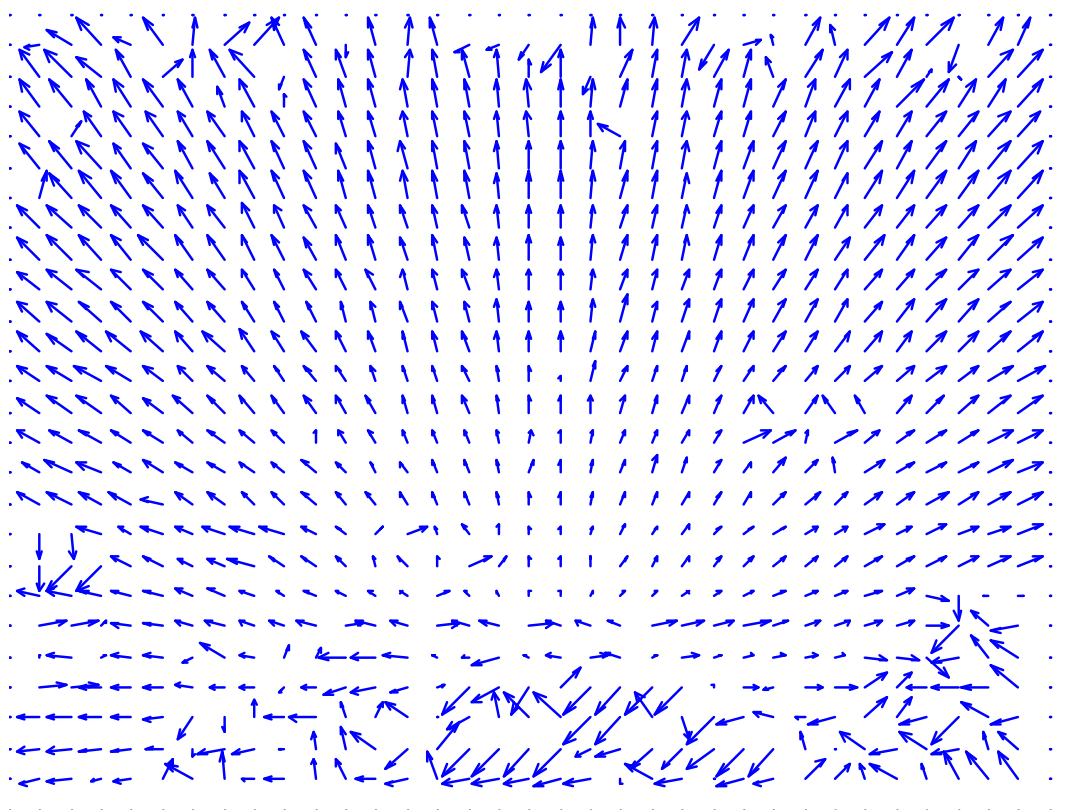


ציור 9.16 - למעלה: צמד תמונות עוקבות בסדרת וידאו (משמאל) ותמונת ההפרש ביניהן בערך מוחלט (מימין). למטה: התמונה העכשווית (שמאל), תמונת החיזוי הנבנית מתוך התמונה הקודמת ע"י קיזוז תנועה בבלוקים של 8 על 8 (אמצע), ותמונת שגיאת החיזוי (ימין).

בתמונה זו, אנרגיית השגיאה בין שתי התמונות (הפרש הפיקסלים בריבוע המחולק במספר הפיקסלים הוא 1172, בעוד שאנרגיית שגיאת החיזוי היא 235 - יחס של כ-4. באלגוריתם שיושם בחרנו לא לטפל בבלוקים המצויים בקצוות התמונה על מנת להימנע מתופעות קצה המצריכות טיפול ייחודי. בשל כך השגיאה בקצוות גדולה יותר משמעותית. כאשר מסירים את השוליים, האנרגיה של השגיאה המקורית היא 1265 ואילו זו של החיזוי היא 65 - יחס של קרוב ל-20.

התמונה הנבנית כתמונת חיזוי היא תמונה הבנויה כקולאז' של בלוקים המועתקים ממיקומים שונים של התמונה הקודמת בסרט הנקלט. בשל כך, לא מפתיע לגלות כל מיני תופעות של גבולות צורמים של בלוקים בתמונה זו. למרות שבצמד שבתמונה 9.16 קיבלנו שגיאה נמוכה, גם בה קיימת אותה תופעה. תמונה 9.17 מראה את תמונת החיזוי, וניכר כי בתחתית בצד ימין מורגשת התופעה ביתר שאת. ביתר התמונה לא רואים את הבעיה, בשל אופיו של היעד המצולם הבנוי בבלוקים כמעט אחידים עם שפות אנכיות ואופקיות, ובשל התנועה הפשוטה בין שתי התמונות. בציור זה גם מובאת מפת התנועה שעל-פיה נבנה החיזוי. ממפה זו ניכר כי ברוב האזורים התנועה חלקה ולכן לא צפוי לקבל קווי תפר צורמים בין הבלוקים. בתחתית התמונה החיצים מבולבלים ולכן שם יש סיכוי סביר לגלות את התופעה בחריפות יתר.

אם סרט וידאו דורש זיכרון של 30 תמונות בגודל של 500 על 500 פיקסלים בשניה, הרי שמדובר בקצב של 7.5 MB לשניה. אלגוריתם ה-MPEG יכול להוריד קצב זה ביחס של 10-200 (תלוי באיכות הנחוצה). הפוטנציאל הטמון ביתרות הזמנית גדול והוא מקנה לטיפול בסדרות יחס דחיסה גדול יותר עבור אותה שגיאה ריבועית לתמונה נתונה.



ציור 9.17 - תמונת החיזוי ומפת התנועה לפי בלוקים של 8 על 8

פרק 10 – שחזור תמונה מהיטלים

219	<u>פרק 10 – שחזור תמונה מהיטלים</u>
220	10.1 בעיות שחזור מהיטלים וניסוחן המתמטי
221	10.2 התמרת ראדון
223	10.3 משפט החתך המרכזי והיפוך התמרת ראדון
226	10.4 שחזור מהיטלים ע"י גישות תדר
228	10.5 שחזור מהיטלים ע"י קונבולוציה
229	10.6 שחזור מהיטלים ע"י גישות אלגבריות

בפרק 6 דנו בשחזור (Restoration) תמונה מקלקול בעל מודל ידוע, וראינו כלים מבוססי שערך סטטיסטי נרתמים לשם פתרון בעיות מסוג זה. למעשה, בפרק 6 הנחנו את התשתית המתמטית לפרק מרכזי בעיבוד אות הקרוי בעיות היפוך (Inverse Problems). בעיית היפוך היא בעיה בה נתון סט מדידות, נתון קשר מתמטי מוגדר היטב (ולאו דווקא דטרמיניסטי – ראה עניין הרעש בשחזור תמונה) בין מדידות אלו ובין איזשהו אות אידיאלי רצוי, ועלינו לבצע היפוך לשם הפקת האות הרצוי מהמדידות המצויות. בשחזור תמונה המדידות הן התמונה המקולקלת, האות האידיאלי היא התמונה המקורית והנקייה, והקשר בין השתיים ידוע כנובע מאופרטור ליניארי (טשטוש) ותוספת של רעש בעל מבנה סטטיסטי ידוע.

בפרק זה נדון בשחזור (Reconstruction) תמונה מתוך סט מדידות שלה אשר מתקבל כהטלות. מסתבר כי מבנה מעין זה נפוץ בצילום רפואי, מדידות עקיפות של קרקע, חישה מרחוק ע"י מכ"ם, ועוד. גישת צילום עקיפה זו ידועה בשם טומוגרפיה ממוחשבת (CT - Computerized Tomography). אנו נראה כי מבנה בעיה זו נופל אף הוא לקטגורית בעיות ההיפוך, וככזה, כלים שכבר פגשנו הם הכלים הנאותים לטיפול בבעיית השחזור, אלא שנדרשות התאמות בשל השוני הקיים. גם כאן נראה כי הקשר בין המדידות לאות האידיאלי ניתן לביטוי כמערכת ליניארית והוספת רעש, אלא שהפעם למדידות אין הקשר ויזואלי – אלה הן הטלות חד-ממדיות שיחד מכילות את המידע המאפשר חזרה לתמונה המוטלת. זוהי הסיבה לאבחנה בין המושג Restoration ובין Reconstruction.

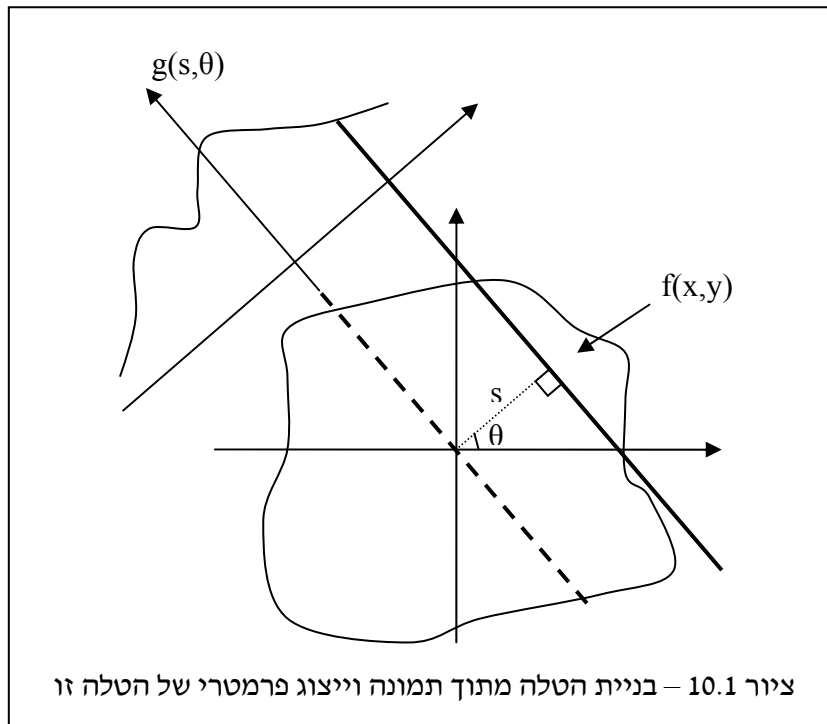
לצד הצגת נושא שחזור מהטלות כבעיית היפוך ושימוש בכלים שכבר הוצגו לשם בניית התמונה מהטלותיה, נציג גם יסודות מתמטיים של אופרטור ההטלה, התמרת ראדון הנובעת ממנה, וייצוג תדרי של כלל הבעיה. לייצוג זה פוטנציאל חשוב בביצוע שחזור מהיר של תמונה, אך, אם להודות על האמת, שחזור זה סובל מאותן חולשות מוכרות שגרמו לנו לתחושת החמצה בדיון בשחזור תמונה מטשטוש. הצגת נושא זה מובאת לשם השלמות של ההצגה, והיכרות עם מושגי היסוד בתחום.

10.1 בעיות שחזור מהיטלים וניסוח המתמטי

נתחיל את הדיון בפונקציה דו-מימדית המוגדרת מעל הרצף $f(x,y)$. פונקציה זו מייצגת את התמונה אותה רצוננו לשחזר. אנו מניחים שתמונה זו מייצגת צפיפות חומר בדו-מימד כפונקציה של המקום (x,y) . נניח כי קרן בעוצמה התחלית I_0 עוברת דרך משטח החומר $f(x,y)$, נעה לאורך הישר L , ומקלט מתעד את עוצמת הקרן במוצא החומר. בכל אינטרוול אינפיניטסימלי באורך dL בו עוברת הקרן לאורכו של הישר היא מונחתת בעוצמה יחסית $\exp\{-f(x,y)dL\}$ (בהנחה שביחידת אורך קצרה זו $f(x,y)$ קבוע בערכו). לכן, לאחר מעבר לכל אורכו של הישר נקבל כי העוצמה הנקלטת במוצא היא מכפלת הניחותים. מכפלה זו מומרת לסכום המעריכים באקספוננט,

$$I = I_0 \cdot \exp\left\{-\int_L f(x,y)dL\right\}$$

אם נניח כעת כי למעשה מסרק מקבילי של קרניים חלף דרך החומר, נקבל פונקציה של ערכי מוצא במקום ערך בודד כמקודם. מבנה זה מתואר בציור 10.1.



אנו נייצג כל קרן ישרה L ע"י שני פרמטרים – זוויתו ביחס לאופק θ ומרחקו מהראשית s כמו שמתואר בצירוף. תחת ייצוג זה, הפונקציה לאחר ההטלה היא

$$I(s, \theta) = I_0 \cdot \exp \left\{ - \int \int_{x, y} f(x, y) \delta(x \cos \theta + y \sin \theta - s) dx dy \right\}$$

בהינתן משפחת פונקציות $I(s, \theta)$ המתייחסת לזוויות הטלה שונות בתחום $0 \leq \theta < \pi$ ולהסטה מהראשית בתחום $-\infty < s < \infty$, רצוננו לבנות את $f(x, y)$. נגדיר את הפונקציה $g(s, \theta)$ כ:

$$g(s, \theta) = -\log \left\{ \frac{I(s, \theta)}{I_0} \right\} = \int \int_{x, y} f(x, y) \delta(x \cos \theta + y \sin \theta - s) dx dy$$

אנו רואים כי בין ההטלות המדודות ובין פונקצית היעד לשחזור יש קשר ליניארי פשוט. קשר זה מתאר את המודל הנחוי בבעיית ההיפוך - אנו ננצל קשר זה בבואנו לפתור את בעיית ההיפוך זו.

נתייחס בקצרה לסוגי ההדמיה הניתנים לתיאור הני"ל: כאשר התמונה $f(x, y)$ הינה פונקצית צפיפות הריקמה בחתך גוף כלשהו, והקרן הנורית הינה קרן רנטגן, אנו מקבלים סידרת הטלות רנטגן והשחזור הנוצר ידוע בשם X-Ray CT. הטלות יכולות גם להתקבל בדרכים אחרות כגון החזרה מתווך לא ידוע, או ממקורות המצויים בתוך התווך. למשל, בחישה סיסמית בה המטרה היא לקבל מפת צפיפות הקרקע מתחת לפני השטח, נוצר הלם קול (בד"כ ע"י מוט חומר נפץ שהוכנס לקדח בקרקע), והחישה נעשית ע"י סידרת חיישני רעד או קול הפרוסים בשטח. כדוגמה אחרת, בצילום PET (Positron Emission Tomography) חומר רדיואקטיבי המצוי בתוך הרקמות בשיעורים שונים מתפרק ומשחרר אגב כך קרינת גאמה. במקרה זה המקורות נמצאים בתוך הגוף, ומיקומן וכמותם ביחידת שטח היא הנעלם אותו נשחזר. המידע הנתון הוא תיעוד אנרגיית הקרינה מחוץ לגוף. כדוגמה שלישית נזכיר את צילום ה-MRI בו שדה מגנטי מבוקר משנה את הספינים (כיוון המגנט) של גרעיני אטומי החנקן, ועם מדידה של שינויים אלו מתקבלת תמונת הטלה של ריכוזי הנוזלים בריקמה. דוגמאות אלו, וישנן גם אחרות, ממחישות דרכים שונות להשיג מידע על גודל בלתי ניתן למדידה ע"י קבלת הטלות או מידע המתייחס להטלות בדרך כלשהי.

10.2 התמרת ראדון

למעשה הוגדרה התמרת ראדון למעלה בהגדרת משפחת ההטלות. בהינתן פונקציה המוגדרת מעל

הקואורדינטות הקרטזיות, $f(x, y)$, התמרת ראדון שלה שתסומן $\mathfrak{R}\{f(x, y)\}$ נתונה ע"י

$$g(s, \theta) = \mathfrak{R}\{f(x, y)\} = \int \int_{x, y} f(x, y) \delta(x \cos \theta + y \sin \theta - s) dx dy$$

כאשר תחום הגדרת הפונקציה g הוא $0 \leq \theta < \pi$ ו- $-\infty < s < \infty$. כל ערך $g(s, \theta)$ מהווה אינטגרציה על פני ישר אחד על הפונקציה $f(x, y)$. מעניין לציין כי אם נתחיל את ההתמרה עם תמונה f המוגדרת מעל קואורדינטות פולריות $f(r, \varphi)$ כך ש- $x = r \cos \varphi$ $y = r \sin \varphi$, נקבל כי התמרת ראדון תהיה

$$\begin{aligned} g(s, \theta) = \mathfrak{R}\{f(x, y)\} &= \iint_{x, y} f(x, y) \delta(x \cos \theta + y \sin \theta - s) dx dy = \\ &= \int_{r=0}^{\infty} \int_{\varphi=0}^{2\pi} r f(r, \varphi) \delta(r \cos \varphi \cos \theta + r \sin \varphi \sin \theta - s) dx dy = \\ &= \int_{r=0}^{\infty} \int_{\varphi=0}^{2\pi} r f(r, \varphi) \delta(r \cos(\varphi - \theta) - s) dx dy \end{aligned}$$

משמעות ביטוי זה היא שבאינטגרציה על-פני הפונקציה $f(r, \varphi)$ ליצירת ערך ההתמרה $f(s, \theta)$ נאספות כל הנקודות $\{r, \varphi\}$ על-פני הסינוסואידה $s = r \cos(\varphi - \theta)$. זאת בניגוד לישר בקואורדינטות (x, y) .

נדון כעת במספר תכונות בסיסיות של התמרת ראדון $g(s, \theta) = \mathfrak{R}\{f(x, y)\}$: מההגדרה ניכר כי

התמרת ראדון היא התמרה ליניארית, דהיינו, עבור שתי תמונות $f_1(x, y)$ ו- $f_2(x, y)$ מתקבל כי

$$\mathfrak{R}\{\alpha f_1(x, y) + \beta f_2(x, y)\} = \alpha \mathfrak{R}\{f_1(x, y)\} + \beta \mathfrak{R}\{f_2(x, y)\}$$

כמו כן ברור כי אם התמונה $f(x, y)$ בעלת תמך סופי, דהיינו, $f(x, y) \equiv 0$, $\forall (x, y) | x^2 + y^2 \geq D$, אזי גם התמרת ראדון בעלת תמך סופי $-D \leq s \leq D$. התמרת ראדון מחזורית בזווית - $g(s, \theta) = g(s, \theta + 2\pi k)$, וסימטרית - $g(s, \theta) = g(-s, \theta + \pi)$. סימטריות זו היא הסיבה להגבלת ההגדרה לאינטרוול $[0, \pi]$.

כאשר מוזזת התמונה $f(x, y)$ בשיעור (x_0, y_0) נקבל התמרה חדשה המוזזת רק בציר s . קל לראות

זאת ע"י:

$$\begin{aligned} \mathfrak{R}\{f(x - x_0, y - y_0)\} &= \iint_{x, y} f(x - x_0, y - y_0) \delta(x \cos \theta + y \sin \theta - s) dx dy = \\ &= \iint_{x, y} f(x, y) \delta(x \cos \theta + y \sin \theta + x_0 \cos \theta + y_0 \sin \theta - s) dx dy = \\ &= g(s - x_0 \cos \theta - y_0 \sin \theta, \theta). \end{aligned}$$

באופן דומה ניתן להראות תכונת הגדלה - $\mathfrak{R}\{f(ax, ay)\} = g(as, \theta) / |a|$. תכונה אחרונה שנוכיר היא שימור הממוצע אשר קלה להוכחה ישירות מההגדרה -

$$\forall \theta, \iint_{x, y} f(x, y) dx dy = \int_s g(s, \theta) ds$$

תכונה זו טריוויאלית אם נחשוב על משמעותה הגיאומטרית. בבחירת זווית הטלה שרירותית, אם נתייחס למתואר בציר 10.1, כל ערך בהטלה הוא אינטגרציה לאורך ישר כלשהו (המוכתב ע"י s). ע"י אינטגרציה לאורכה של כל ההטלה, אנו אוספים את כל הישרים המקבילים ובכך מקבלים את סך נפח הפונקציה $f(x, y)$.

כיוון שאנו חושבים במונחים של התמרה, שאלה מיידית שנשאלת היא האם ההתמרה המוגדרת

הפיכה, דהיינו, האם בהינתן $g(s, \theta)$ בתחום $0 \leq \theta < \pi$ ו- $-\infty < s < \infty$ ניתן לחזור לתמונה $f(x, y)$. מיד נראה כי היפוך ההתמרה אפשרי, וכיצד לעשותו.

10.3 משפט החתך המרכזי והיפוך התמרת ראדון

דבר מעניין ובעל חשיבות קורה כאשר מתבצעת התמרת פוריה על וקטורי ההטלה. בהינתן הפונקציה הדו-מימדית $g(s, \theta)$ נבצע התמרת פוריה ביחס למשתנה s ונקבל פונקציה חדשה

$$\mathfrak{F}_1\{g(s, \theta)\} = \mathfrak{F}_1\{\mathfrak{R}\{f(x, y)\}\} = G(u, \theta)$$

נפתח פעולה זו ונקבל

$$\begin{aligned} G(u, \theta) &= \mathfrak{F}_1\{g(s, \theta)\} = \int_s g(s, \theta) \exp\{-j2\pi su\} ds = \\ &= \mathfrak{F}_1\{\mathfrak{R}\{f(x, y)\}\} = \iint_{x, y} \int_s f(x, y) \delta(x \cos \theta + y \sin \theta - s) \exp\{-j2\pi su\} ds dx dy = \\ &= \iint_{x, y} f(x, y) \left[\int_s \delta(x \cos \theta + y \sin \theta - s) \exp\{-j2\pi su\} ds \right] dx dy = \\ &= \iint_{x, y} f(x, y) \exp\{-j2\pi(u x \cos \theta + u y \sin \theta)\} dx dy \end{aligned}$$

ביטוי זה נראה שתום. אך נזכור כי התמרת פוריה דו-מימדית על הפונקציה $f(x, y)$ נתונה ע"י

$$F(\omega_x, \omega_y) = \iint_{x, y} f(x, y) \exp\{-j2\pi(\omega_x x + \omega_y y)\} dx dy$$

אם נתייחס לנקודות התדר בקואורדינטות פולריות $\omega_x = u \cos \theta$, $\omega_y = u \sin \theta$ נקבל כי התמרת הפוריה

העוברת מ- $f(x, y)$ ל- $F(u \cos \theta, u \sin \theta) = \tilde{F}(u, \theta)$ ניתנת לרישום כ:

$$\tilde{F}(u, \theta) = \iint_{x, y} f(x, y) \exp\{-j2\pi(xu \cos \theta + yu \sin \theta)\} dx dy$$

וזוהי בדיוק הפונקציה $G(u, \theta)$ שקיבלנו בעקבות התמרת פוריה חד-מימדית על התמרת ראדון. תוצאה זו ידועה בשם "משפט החתך המרכזי" (Projection Slice Theorem). פירוש הדבר הוא שבהינתן התמרת ראדון של $f(x, y)$ אותה סימנו כ- $g(s, \theta)$, ביצוע ההתמרה הפוכה ייעשה ע"י השלבים הבאים המתוארים גם בסכימת הבלוקים בצויר 10.2:

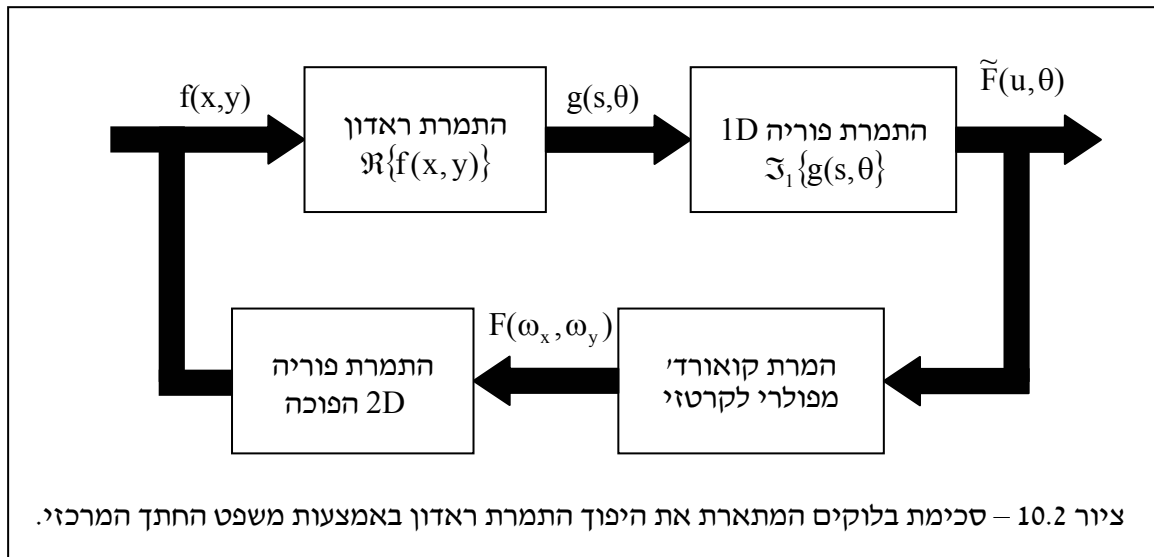
1. בצעי התמרת פוריה חד-מימדית על הפונקציה $g(s, \theta)$ ביחס למשתנה s . התוצאה היא

הפונקציה הדו-מימדית החדשה $G(u, \theta)$. אנו ראינו כי זוהי גם התמרת הפוריה הפולרית $\tilde{F}(u, \theta)$.

2. המרי קואורדינטות $\{u, \theta\}$ המהוות קואורדינטות פולריות לקואורדינטות קרטזיות מהצורה $\omega_x = u \cos \theta$, $\omega_y = u \sin \theta$. ברצף פעולה זו היא פעולה באנאלית – רישום שונה ולא יותר

(כשנעבור לתיאור דיסקרטי נגלה שהדברים מסתבכים). כך תתקבל הפונקציה $F(\omega_x, \omega_y)$.

3. בצעי התמרת פוריה הפוכה דו-מימדית על $F(\omega_x, \omega_y)$ לקבלת $f(x, y)$.



למעשה, דרך משפט החתך המרכזי הצגנו דרך אפשרית אחת לביצוע היפוך להתמרת ראדון. כל צעדי התהליך מוגדרים היטב לפיכך ההתמרה ההפוכה נראית כתהליך יציב ואמין המוביל לקשר חד-חד-ערכי בין תמונה והתמרתה.

ישנן גם דרכים אחרות לביצוע ההתמרה ההפוכה – נציג שתיים כאלה – גישה ישירה וגישת ההטלות לאחור. נתחיל עם הגישה הישירה אותה נציג ללא הוכחה. בהינתן $g(s, \theta)$ – התמרת ראדון של הפונקציה קיים הקשר הבא:

$$f(x, y) = \frac{1}{2\pi^2} \int_0^\pi \int_{-\infty}^\infty \frac{\frac{\partial}{\partial s} g(s, \theta)}{x \cos \theta + y \sin \theta - s} ds d\theta$$

וקשר זה אינו אלא ההתמרה ההפוכה בדרך ישירה. מהביטוי הנ"ל ניכר כי הפונקציה $g(s, \theta)$ עוברת שלב ראשון בו היא נגזרת לפי s – פעולה הניתנת להתייחסות כסינון מגביר גבוהים בציר אחד. לאחר מכן מתבצעת פעולה המוכרת בשם "התמרת הילברט". אנו לא ניכנס למשמעות פעולות אלו משום שחשיבותן תיאורטית גרידא.

קיימת דרך אחרת לבניית התמרת ראדון הפוכה. ברור כי דרך זו קשורה בעבותות לשתי הגישות הקודמות. עם זאת, חשיבותה בזוית ראייה אחרת שהיא מספקת לעניין היפוך ההתמרה – זווית שתוביל לאלגוריתמים מעשיים בשחזור תמונה מהיטליה. נתחיל בהגדרה של אופרטור מעניין הקרוי "אופרטור ההטלה לאחור". אופרטור זה המוגדר ע"י

$$b(x, y) = \int_0^\pi g(x \cos \theta + y \sin \theta, \theta) d\theta$$

מתחיל בפונקציית ההטלות $g(s, \theta)$ ובפעולת אינטגרציה על-פניה מקבל פונקציה במקום (x, y) . אם נתייחס לנקודת יעד (x, y) קבועה, הרי שהאינטגרציה נעה על-פני $g(s, \theta)$ לאורך קו קבוע ב- θ וסינוסואידה ב- s . מה כה מיוחד באופרטור זה? ובכן – מספר דברים אותם נציין ללא הוכחה.

אופרטור ההטלה לאחור הוא למעשה ה-Adjoint של התמרת ראדון. במילים אחרות ופשוטות, לו היינו חושבים על התמרה דיסקרטית ונותנים להתמרת ראדון ייצוג כמטריצה R המכפילה את התמונה f (בסידור לקסיקוגרפי), הרי שהאופרטור שהגדרנו כאן אינו אלא פעולת R^T (Transpose של המטריצה). כשנגיע לדיון בדרכים אלגבריות לשחזור מהיטלים נראה כי אופרטור זה חיוני בהיפוך ההתמרה.

מסתבר כי קיים הקשר הבא המציע כי תהליך ההטלה לאחור עושה כיברת דרך לקראת היפוך ההתמרה. קשר זה אומר שאם נתחיל בפונקציה $f(x,y)$, נבצע התמרת ראדון לקבלת $g(s,\theta)$, נפעיל את אופרטור ההטלה לאחור לקבלת $b(x,y)$, נקבל כי:

$$b(x,y) = f(x,y) \otimes \frac{1}{\sqrt{x^2 + y^2}}$$

כלומר, בפעולה נוספת של דה-קונבולוציה ע"י מסנן שייצוגו התידרי הוא כגוזר רדיאלי נגיע חזרה לתמונת המקור. וכך אנו רואים דרך שלישית לביצוע ההתמרה ההפוכה.

אנו נעצור את הדיון כאן ונעבור לגישות מעשיות לשחזור מהטלות. פירוש הדבר שאנו מפסיקים לחשוב במונחי הרצף שליוו אותנו עד כה, ונתחיל לחשוב במונחים דיסקרטיים - ידוע לנו סט הטלות באוסף הזוויות ובערכי הסטת הישר

$$\theta_n = \pi n / N, n = 0, 1, \dots, N-1$$

$$s_m = mS/M, -M/2 \leq m < M/2$$

יש אם כך N הטלות הפרוסות ברווחי זווית קבועים, וכל הטלה כוללת M דגימות בתחום $[-S, +S]$.

גודל צעד הדגימה S/M יוכתב ישירות מהתדר הרדיאלי המירבי אותו רצוננו לשחזר בתמונה $f[k,j]$. בהינתן תמונה דו-מימדית בעלת התמרת פוריה המתאפשת מחוץ לדיסקה ברדיוס $2\pi\omega_0$, כל הטלה שלה אף היא חסומת תדר עם אותו סף. לכן, כל הטלה צריכה להידגם במרווח שיהיה קטן מ- $1/2\omega_0$. בדרך זו נוכל לקבוע את M .

כמות ההטלות N קשה יותר לקביעה ובחירתו צריכה להיות קשורה ל- $2\pi\omega_0$ אף היא, אם כי בדרך עקיפה. במונחים גסים של חשבונאות המידע, אם יש לנו MN ערכי מדידה, סביר כי התמונה המשוחזרת $f[k,j]$ תכיל כמות דומה או נמוכה של דגימות, שאם לא כן – בעיית השחזור אינה מוגדרת היטב.

בידיעה שהתמונה חסומת תדר לרדיוס $2\pi\omega_0$, פירוש הדבר שמרווח הדגימה הקרטזי צריך להיות לכל היותר $D = 1/2\omega_0$. אנו גם מניחים שהתמונה f בעלת תמך מרחבי בצורת דיסקה שרדיוסה S (כך הרי קבענו כשהגבלנו את s לאינטרוול $[-S, +S]$). לכן, בניתוח גס, בריבוע בגודל $2S$ על $2S$ ייכנסו $(2S/2\omega_0)^2$ דגימות. כל הטלה מכילה לפחות $(2S/2\omega_0)$ דגימות, ולכן תידרשנה $(2S/2\omega_0)$ הטלות

לפחות על מנת לספק את הצורך לאזן בין המידע הנתון והמידע המבוקש. בפועל, רצוי לתת יותר מדידות מנעלמים לשם ייצוב השחזור.

עניין אחר עליו יש לתת את הדעת במעבר לאותו דגומים במקום ובמרחב ראדון הוא עניין דרך היווצרן של ההטלות מהמידע הדגום. דרך אחת לחשוב על הקשר היא במונחי אותות חסומי תדר שנדגמו. אפשרות אחרת היא לחשוב על התמונה המשוחזרת כתמונה הבנויה מפיקסלים בעלי תמך ריבועי אשר על-פניהם ערך קבוע של $f(x,y)$.

10.4 שחזור מהיטלים ע"י גישות תדר

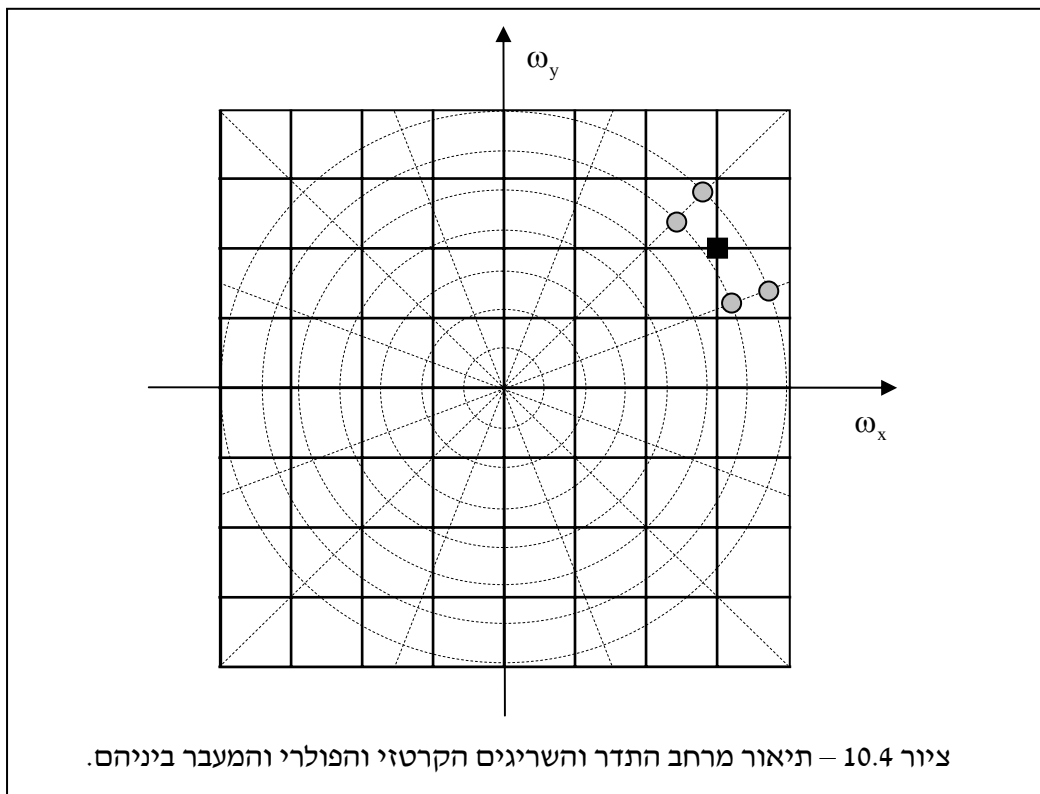
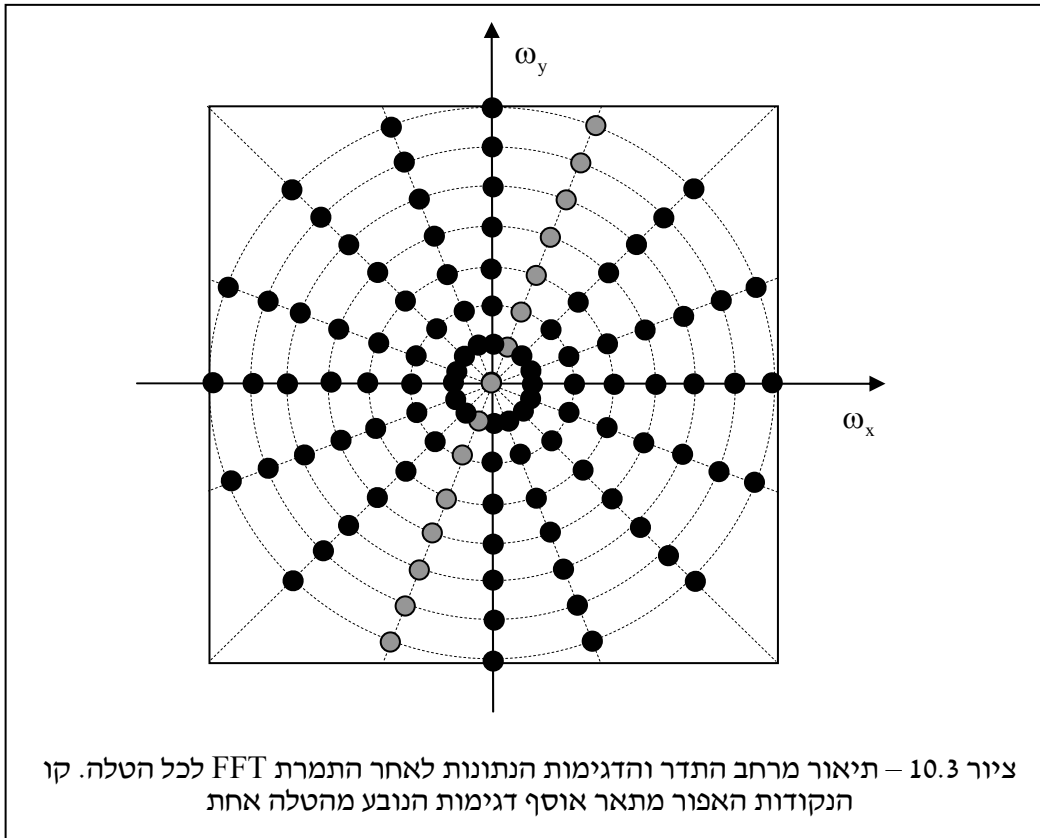
אחת הדרכים הפופולריות לשחזור תמונה מהטלותיה היא גישת התדר המתבססת על משפט החתך המרכזי. משפט זה הציע כי ביצוע התמרת פוריה חד-מימדית על כל הטלה מביא אותנו לפונקציה חדשה אשר ניתנת להתייחסות כתמונה הרצויה בתדר כשהיא דגומה בקואורדינטות פולריות. כעת נתאר תהליך מעשי שנועד לשחזור תמונה בגישה זו, תוך התייחסות לעובדה שכל הגדלים המעורבים דגומים.

נקודת המוצא היא הפונקציה $g[m,n]$ המוגדרת בתמך $-M/2 \leq m < M/2, 0 \leq n < N$. כל שורה (m קבוע) - $g[m,:]$ (במונחי Matlab) של מערך זה מייצגת זווית קבועה, וכל עמודה (n קבוע) - $g[:,n]$ (במונחי Matlab) מייצגת הטלה אחת.

בצעד ראשון ניקח כל הטלה ונבצע עליה 1D-FFT. למרות שכל הטלה כזו מכילה M ערכים, אין סיבה להגביל את ה-FFT ל- M ערכי מוצא. ע"י ריפוד באפסים של וקטור הכניסה ניתן לייצר סידרת מוצא באורך גדול יותר. אלגוריתמית, ניקח את המערך הדו-מימדי בגודל M שורות על N עמודות, נוסף לכל עמודה N' אפסים לקבלת מערך חדש בגודל $M \times (N+N')$, ואז לכל עמודה נבצע FFT חד-מימדי. המוצא יהיה מערך בגודל זה $M \times (N+N')$, בו כל עמודה מהווה חתך מרכזי בתדר, כמתואר בציור 10.3. חתכים אלו פזורים במרווחי זווית קבועים ונותנים בפועל דגימה פולרית של מרחב התדר.

כל שעלינו לעשות עכשיו הוא התמרת פוריה הפוכה מהדגימות הפולריות כמתואר בציור 10.3 אל התמונה במקום $f[k,j]$ כשהיא בקואורדינטות פולריות. לרוע המזל, התמרה כזו לא ידועה לנו, ואם נבחר לעבוד לפי הגדרת התמרת פוריה נאלץ לבצע תהליך נומרי עתיר חישובים. ה אלטרנטיבה היא המרת אוסף הדגימות הפולריות באוסף אחר קרטזי. עבור אוסף כזה, ביצוע 2D-FFT ייתן בצורה מינימלית ומהירה את התוצאה הרצויה. המרת הקואורדינטות תיעשה ע"י אינטרפולציה – כל נקודת יעד בשריג הקרטזי תחושב כממוצע משוקלל של שכניותיה מהשריג הפולרי עבורן יש לנו ערכים. ציור 10.4 ממחיש המרה זו. זו גם הסיבה לריפוד באפסים שביצענו בצעד הראשון – ריפוד זה יוצר לכל קרן החולפת דרך הראשית יותר נקודות דגימה, ולכן האינטרפולציה שתיעשה תהיה מדוייקת יותר. ברור כי בנוסף יש לדאוג ליותר קרניים כאלו – לזה קשר ישיר למספר ההטלות הנתונות.

לאחר ביצוע האינטרפולציה וקבלת N על N נקודות תדר בשריג קרטזי, התמרת 2D-FFT הפוכה תוביל לתמונה הרצויה. בדרך-כלל, צעד האינטרפולציה מהווה אבן נגף בתהליך שתואר וביצועיו ירודים.



10.5 שחזור מהיטלים ע"י קונבולוציה

שחזור מהיטלים בגישת התדר הישירה נובעת ממשפט החתך המרכזי אשר הוכח להיות אחת הדרכים לביצוע התמרת ראדון הפוכה. באופן דומה, נוכל להציע אלגוריתם לשחזור מהיטלים המבוסס על אופרטור ההטלה לאחור.

שוב נקודת המוצא היא הפונקציה $g[m,n]$ המוגדרת בתוך $-M/2 \leq m < M/2, 0 \leq n < N$. כל שורה (m קבוע) של מערך זה מייצגת זווית קבועה, וכל עמודה (n קבוע) מייצגת הטלה אחת שלמה. לפי התכונה המרכזית של אופרטור ההטלה לאחור, אם ניקח את סט ההטלות $g[m,n]$ ונבצע אינטגרציה מהצורה (דיסקרטיזציה פשוטה של נוסחאת האופרטור)

$$b(x, y) = \sum_{n=0}^{N-1} g \left[x \cos\left(\frac{\pi n}{N}\right) + y \sin\left(\frac{\pi n}{N}\right), n \right]$$

תתקבל תמונה הקשורה לתמונה הרצויה עד-כדי פעולה ליניארית וקבועה במקום. נוכל לדגום את מישור המקום ע"י $x=kd, y=jd$ כש- d מרווח הדגימה ולכתוב

$$b(kd, jd) = b[k, j] = \sum_{n=0}^{N-1} g \left[kd \cos\left(\frac{\pi n}{N}\right) + jd \sin\left(\frac{\pi n}{N}\right), n \right]$$

ברור כי נצטרך לעגל את הגודל שמחליף את האינדקס m בפונקציה g בסכימה על מנת לקבל שימוש בפונקציה הדיסקרטית ממנה התחלנו.

הצעד השני והאחרון שנדרש לשם השלמת השחזור הוא ביצוע הסינון. הקשר עליו אנו מסתמכים

הוא:

$$\begin{aligned} b(x, y) &= f(x, y) \otimes \frac{1}{\sqrt{x^2 + y^2}} \\ \Rightarrow B(\omega_x, \omega_y) &= F(\omega_x, \omega_y) \cdot \mathfrak{F} \left\{ \frac{1}{\sqrt{x^2 + y^2}} \right\} \\ \Rightarrow F(\omega_x, \omega_y) &= B(\omega_x, \omega_y) \cdot \mathfrak{F}^{-1} \left\{ \frac{1}{\sqrt{x^2 + y^2}} \right\} \end{aligned}$$

המסנן הנדרש לשם קבלת F הוא גוזר רדיאלי אידיאלי, ומסנן זה אינו מעשי כיוון שתהיה לו נטייה להגביר רעשים. לכן, בספרות מציעים מגוון מסנני שחזור שנראים כגוזר סביב הראשית, ודועכים כשמתרחקים מהראשית על מנת לדכא הגבר רעש. רעיון היוריסטי זה לא זר לנו ואנו ראינו כמותו כשדיברנו על שחזור תמונות מטשטוש ורעש, שם ראינו כי רגולריזציה המדעיכה את המסנן המשחזר רצויה לשם קבלת תוצאות טובות. עם זאת, גם ראינו כי יש אלטרנטיבה בעלת רקע תיאורתי נאות לשם תכנון שחזור זה המבוסס על שערך סטטיסטי. אנו ניגש למשפחת שיטות אלו בסעיף הבא.

חשוב להבהיר כי בדומה לגישות התדר, גם גישה זו סובלת מצרות לרוב בשל הדגימות השונות, קירובי האינטגרלים לסכימות סופיות, ולבסוף, ביצוע סינון הפוך לא יציב.

10.6 שחזור מהיטלים ע"י גישות אלגבריות

אם נחזור להתחלה, נוכל להציע מודל הקושר את ההטלות הנתונות לתמונה הרצויה כשכל הגדלים המעורבים מיוצגים כדגומים. מודל זה יתייחס לתמונת היעד $f[k,j]$ כוקטור עמודה בסידור לקסיקוגרפי \underline{f} . מטריצה R_{θ_0} תייצג את תהליך ההטלה של \underline{f} לקבלת וקטור ההטלה \underline{g}_{θ_0} , ונקבל סידרת משוואות מהצורה

$$\left\{ \underline{g}_{\theta_n} = R_{\theta_n} \underline{f} + \underline{v}_n \right\}_{n=0}^{N-1}$$

לכל זווית הטלה יש משוואה כזו, המטריצה המתארת את ההטלה מוגדרת היטב, ועלינו להתייחס באופן נאות לרעש אדיטיבי שמופיע בכל הטלה – אנו נניח כי רעש זה גאוסי, בעל ממוצע אפס, לבן, וחסר קורלציה בין ההטלות השונות. עפ"י כלים שכבר ראינו בפרק 6, יישום ה- Maximum Likelihood יוביל לבעיית האופטימיזציה

$$\text{Min.}_{\underline{f}} \varepsilon\{\underline{f}\} = \sum_{n=0}^{N-1} \left\| \underline{g}_{\theta_n} - R_{\theta_n} \underline{f} \right\|_2^2$$

אשר פתרונה נתון ע"י

$$\underline{f}_{\text{opt}} = \left[\sum_{n=0}^{N-1} R_{\theta_n}^T R_{\theta_n} \right]^{-1} \left[\sum_{n=0}^{N-1} R_{\theta_n}^T \underline{g}_{\theta_n} \right]$$

נוכל גם לרכז את כל המשוואות למערכת אחת ע"י:

$$\underline{g} = \begin{bmatrix} \underline{g}_{\theta_0} \\ \underline{g}_{\theta_1} \\ \vdots \\ \underline{g}_{\theta_{N-1}} \end{bmatrix} = \begin{bmatrix} R_{\theta_0} \\ R_{\theta_1} \\ \vdots \\ R_{\theta_{N-1}} \end{bmatrix} \underline{f} + \begin{bmatrix} \underline{v}_0 \\ \underline{v}_1 \\ \vdots \\ \underline{v}_{N-1} \end{bmatrix} = R \underline{f} + \underline{v}$$

ואז הפתרון הנ"ל אינו אלא

$$\underline{f}_{\text{opt}} = (R^T R)^{-1} R^T \underline{g}$$

עיון קצר יגלה כי הביטוי $R^T \underline{g}$ אינו אלא הפעלת אופרטור ההטלה לאחור אותו הצגנו קודם. ההכפלה בהיפוך המטריצה $R^T R$ נותנת את הקונבולוציה לה נדרשנו בשיטה זו. המסקנה הנגזרת היא ששיטת השחזור מבוססת קונבולוציה אינה אלא שחזור מבוסס ML. ככזו, אנו מכירים את חסרונותיה היטב ויודעים כי זו אינה שיטה חזקה דיה להתמודד עם רעש ועם אופרטור קלוקל בעל Condition-Number גרוע (כלומר, מטריצת קלוקל קרובה לסינגולרית). ראינו כי הפתרון הוא רגולריזציה, ובפרט כזו מבוססת MAP בה אנו מניחים מבנה לתמונה הנעלמת. לשם דוגמה, עבור רגולריזציה פשטנית מהסוג

$$\text{Min.}_{\underline{f}} \varepsilon\{\underline{f}\} = \sum_{n=0}^{N-1} \left\| \underline{g}_{\theta_n} - R_{\theta_n} \underline{f} \right\|_2^2 + \lambda \|\underline{Df}\|_2^2$$

נקבל כי השחזור נתון ע"י

$$\underline{f}_{\text{opt}} = \left[\sum_{n=0}^{N-1} \mathbf{R}_{\theta_n}^T \mathbf{R}_{\theta_n} + \lambda \mathbf{D}^T \mathbf{D} \right]^{-1} \left[\sum_{n=0}^{N-1} \mathbf{R}_{\theta_n}^T \underline{g}_{\theta_n} \right] = \left[\mathbf{R}^T \mathbf{R} + \lambda \mathbf{D}^T \mathbf{D} \right]^{-1} \mathbf{R}^T \underline{g}$$

אם קודם סיננו ע"י מסנן גוזר תיאורתי המיוצג ע"י ההיפוך של $\mathbf{R}^T \mathbf{R}$, הרי שהמסנן שמוצע כאן אינו אלא גירסה מרוסנת שלו כפי שגישות היוסטיות בספרות בחרו להפעיל. ברור כי ניתן להציע דרכים מוצלחות יותר כרוח הדברים בהם נגענו בפרק 6, כגון שימוש בפונקציות רובסטיות ועוד.

מרכיב חדש ואשר מסבך את העניינים בשחזורים מבוססי ייצוג מטריצי כגון אלו הוא בעיית אי-דלילות של המטריצות המעורבות. בשחזור תמונות מטשטוש, מטריצת הקלקול דלילה מאוד וכך כמובן גם השחלוף שלה. כאן לעומת זאת כל שורה במטריצה \mathbf{R} מכילה N איברים (כל אינטגרציה על ישר חוצה את התמונה ואוספת N ערכי פיקסלים) שסוכמו. יתרה מכך, מבנה המטריצה אינו פשוט כמו במקרה של טשטוש מקומי – מיקום האיברים השונים מאפס דורש עבודת חישוב מורכבת יותר. אלו נושאים בעלי השפעה ישירה על מורכבות הפתרון הנומרי לשחזור. מסיבות אלו מחפשים מפתחי אלגוריתמים בתחום זה קיצורי דרך והאצות של שחזור איטרטיבי של \underline{f} . אנו נראה כאן שתי גישות כאלו – אחת פשוטה להבנה אך איטית להחריד, ואחרת בה ישנה האצה משמעותית אך אי-דיוקים שנובעים מהאצה זו. נקדים ונאמר כי יש שיטות רבות אחרות.

ראינו כי בבסיס הדברים הפעלת ML יחד עם רגולריזציה יוצרים את פונקציית המחיר הבאה

בבעיית האופטימיזציה אותה עלינו לפתור:

$$\varepsilon\{\underline{f}\} = \|\underline{g} - \mathbf{R}\underline{f}\|_2^2 + \lambda \|\mathbf{D}\underline{f}\|_2^2$$

גרדיאנט פונקציה זו נתון ע"י

$$\frac{\partial \varepsilon\{\underline{f}\}}{\partial \underline{f}} = 2\mathbf{R}^T (\mathbf{R}\underline{f} - \underline{g}) + 2\lambda \mathbf{D}^T \mathbf{D}\underline{f}$$

ולכן איטרציות של אלגוריתם השיפוע המירבי (Steepest Descent) ייעשו ע"י המשוואה האיטרטיבית

$$\underline{f}_{k+1} = \underline{f}_k - \mu (\mathbf{R}^T (\mathbf{R}\underline{f}_k - \underline{g}) + 2\lambda \mathbf{D}^T \mathbf{D}\underline{f}_k)$$

ברור כי אנו לא מעוניינים לבנות מטריצות לשם יישום האלגוריתם, ולכן ניתן למשוואה זו הסבר במונחים של פעולות על התמונה f – הכפלה ב- \mathbf{D} אינה אלא הפעלת מסנן לפלסיאן (או גוזר אחר שימדוד חלקות), הפעלת \mathbf{D}^T זהה בשל סימטריות המטריצה. עד כה זה החלק הקל. הפעלת \mathbf{R} על התמונה פירושה בניית ההטלות מהתמונה – זוהי פעולה תובענית בחישובים אך היא חיונית. הפעולה \mathbf{R}^T היא פעולת ההטלה לאחור אותה פגשנו כבר. שתי פעולות אלו (\mathbf{R} ו- \mathbf{R}^T) יכתיבו את סיבוכיות האלגוריתם באיטרציה אחת. יש לזכור כי לשם התכנסות קרוב לוודאי שנזדקק לעשרות איטרציות ופירוש הדבר אלגוריתם איטי למדי.

אם נצביע על הנקודה הכואבת מכל באלגוריתם שהוצע – יהיה זה חישוב ההטלה וההטלה לאחור

– אם אין מנוס מלחשבם, רצוי שנידרש לכך פעמים ספורות בלבד. זהו הרעיון מאחורי ההאצה הבאה

המוכרת לאנשי עיבוד אותות כאלגוריתם ה-LMS. נציג אלגוריתם זה ללא איזכור של נושא הרגולריזציה לשם פשטות. פונקצית היעד שלנו ניתנת לתיאור כ:

$$\varepsilon\{\underline{f}\} = \|\underline{g} - R\underline{f}\|_2^2 = \sum_{j=1}^L (g_j - \underline{r}_j^T \underline{f})^2$$

בתיאור זה פירקנו את נורמת השגיאה ב-L המשוואות לסכום של L ביטויים ריבועיים פשוטים, כאשר L הוא מספר המדידות הסקלאריות (מספר ההטלות כפול מספר הדגימות בכל הטלה). אלגוריתם ה-SD אותו הצגנו מציע את העדכון

$$\underline{f}_{k+1} = \underline{f}_k - \mu R^T (R\underline{f}_k - \underline{g}) = \underline{f}_k - \mu \sum_{j=1}^L \underline{r}_j^T (\underline{r}_j \underline{f}_k - g_j)$$

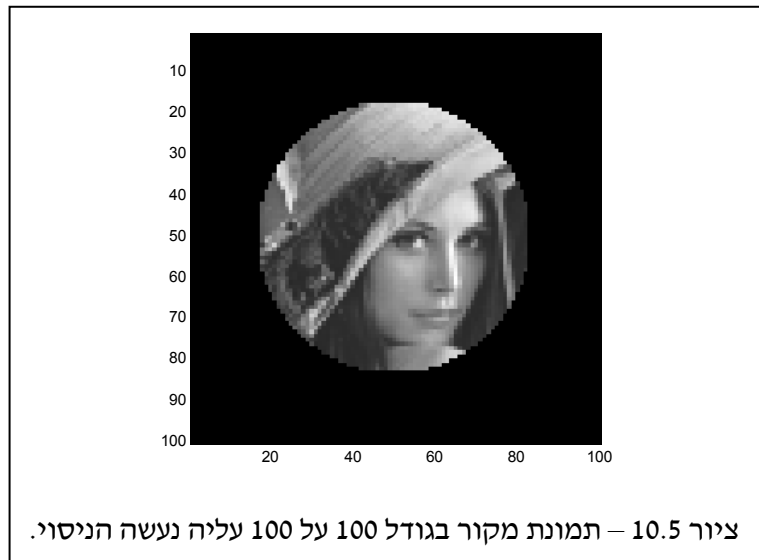
אנו רואים כי העדכון ניתן אף הוא להצגה כפירוק של L עדכונים. באלגוריתם הנ"ל, L עדכונים מחושבים עפ"י תוצאת האיטרציה הקודמת, ורק אז נעשה עדכון מאסיבי לפתרון. אלגוריתם ה-LMS מציע כתחליף לעדכן את f אחרי כל אחד מאיברי הגרדיאנט. לכן האלגוריתם יהיה הבא

$$\left\{ \begin{array}{l} \underline{f}_1^k = \underline{f}_1^{k-1} - \mu \underline{r}_1^T (\underline{r}_1 \underline{f}_1^{k-1} - g_1) \\ \underline{f}_2^k = \underline{f}_2^{k-1} - \mu \underline{r}_2^T (\underline{r}_2 \underline{f}_1^{k-1} - g_2) \\ \vdots \\ \underline{f}_{j+1}^k = \underline{f}_{j+1}^{k-1} - \mu \underline{r}_{j+1}^T (\underline{r}_{j+1} \underline{f}_j^{k-1} - g_{j+1}) \\ \vdots \\ \underline{f}_L^k = \underline{f}_L^{k-1} - \mu \underline{r}_L^T (\underline{r}_L \underline{f}_{L-1}^{k-1} - g_L) \end{array} \right.$$

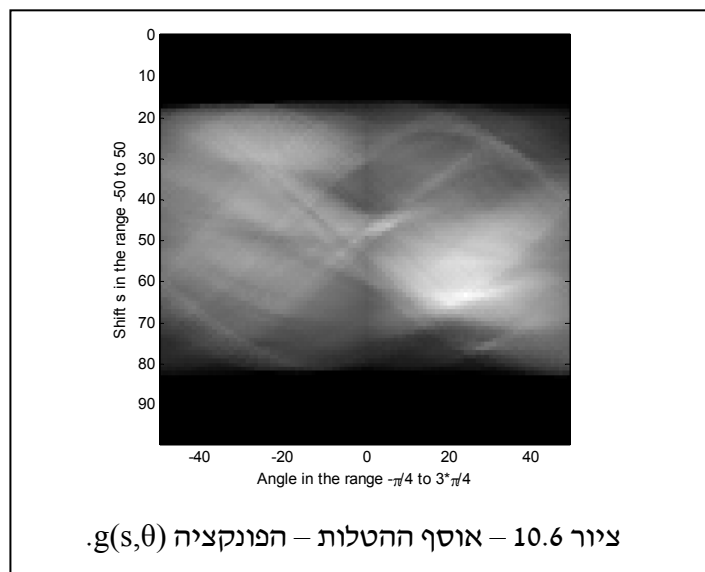
באופן זה, כל מדידה סקלארית שמגיעה יוצרת עדכון לפתרון – דבר נוח למדי במצבים בהם חישוב יכול להיעשות במהלך רכישת מדידות. האלגוריתם הנ"ל מציע L עדכונים השחזור במהלך של כמות חישובים ששקולה לאיטרציה בודדה של ה-SD. עם זאת, ניתן להמשיך ולעדכן ע"י סיבוב נוסף של עדכונים על המדידות.

מסתבר כי אלגוריתם זה בעל תכונות התכנסות טובות למדי – אנו נימנע מלהיכנס לפאן התיאורטי של ניתוח זה ונציין כי בכמות חישובים השקולה לאיטרציה אחת של ה-SD (דהיינו – פעם אחת חישוב הטלות ופעם שימוש באופרטור ההטלה לאחור) מגיע אלגוריתם ה-LMS לתמונה בעלת תוכן רב יותר ומשמעותי. יתרון משמעותי של אלגוריתם ה-LMS עליו לא דיברנו הוא יכולת העקיבה שלו – במקרה בו האובייקט המצולם נע תוך כדי רכישת ההטלות, יתכן כי נרצה לשחזר סרט ולא תמונה בודדה. אלגוריתם ה-LMS מייצר סידרת תמונות בהן להטלות האחרונות יותר משקל בקביעת התמונה וכך נקבל סידרת תמונות נאמנה יותר למתרחש.

לשם שלמות תיאור נושא השחזור מהיטלים, נציג כאן תוצאות ניסוי קצר ומלאכותי במקצת. אנו מתחילים בתמונת מקור - תמונה בגודל של 100 על 100 המתוארת בצירוף 10.5.



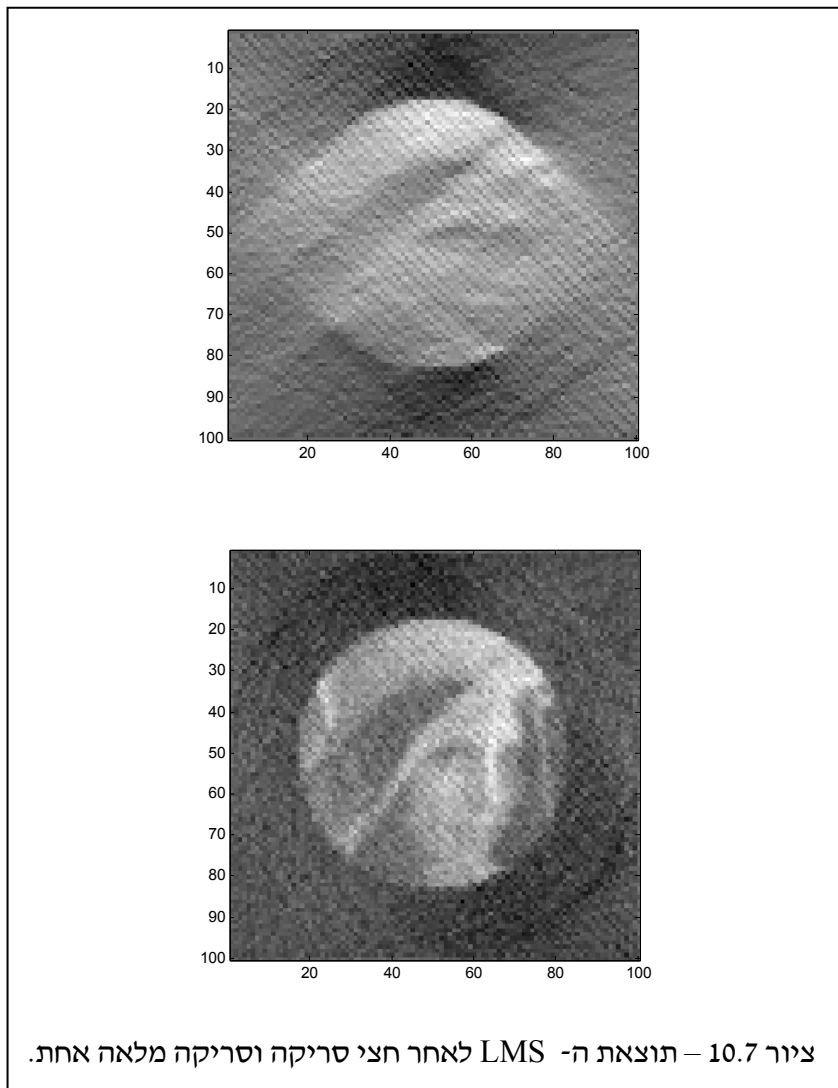
מתמונה זו אנו יוצרים סידרת היטלים הכוללת 100 הטלות בפיזור זוויות אחיד באינטרוול $[0, \pi]$, כשבכל הטלה יש 100 דגימות המתייחסות להסטה s בתחום -50 ועד 49 . לצורך יצירת ההטלות מהמידע הדיסקרטי נעשית הפרדה בטיפול בין ישרים בזוויות אופקיות $[-\pi/4, -\pi/4]$ וישרים בזוויות אנכיות $[\pi/4, 3\pi/4]$. אם נתייחס לישרים בזוויות אופקיות, עבור הטלה בזווית מסויימת θ והסטה s , וערכי הקואורדינטה x בתחום $[-50, 49]$ (אנו מניחים כי מרכז התמונה הוא הראשית) נקבע את הקואורדינטות המתאימות y להיות אלו המקיימות את משוואת הישר $x \cos \theta + y \sin \theta - s = 0$. בהינתן אוסף של הקואורדינטות (x, y) הכוללת 100 צמדים, חלקם ייפול מחוץ לתמונה ואלו צריכים להיזרק. הנותרים מתארים מקום גאומטרי בו נופל הישר על שטח התמונה. נעגל את קואורדינטות y ונסכום (שקול לאינטגרציה) את הפיקסלים בנקודות המתקבלות לשם קבלת $g(s, \theta)$. תהליך זה חוזר על עצמו 100 פעמים כנגד כל ערך של s ו-100 פעמים כנגד כל ערך של θ . תמונת ההטלות מתוארת בציור 10.6.



הטיפול הישירים בעלי זווית אנכית דומה כשהפעם y נקבע להיות 100 ערכים בתחום $[-50, 49]$ ו- x נקבע מתוך המשוואה $x \cos \theta + y \sin \theta - s = 0$ ועובר העגלה. כתחליף להעגלה (שפירושה שיטת אינטרפולציה Nearest-Neighbor), ניתן לבצע הטלות מדויקות יותר ע"י אינטרפולציה נאמנה יותר לישר, אך כפי שנראה, בשל מלאכותיות הניסוי, אין עניין זה יוצר הפרעה משמעותית.

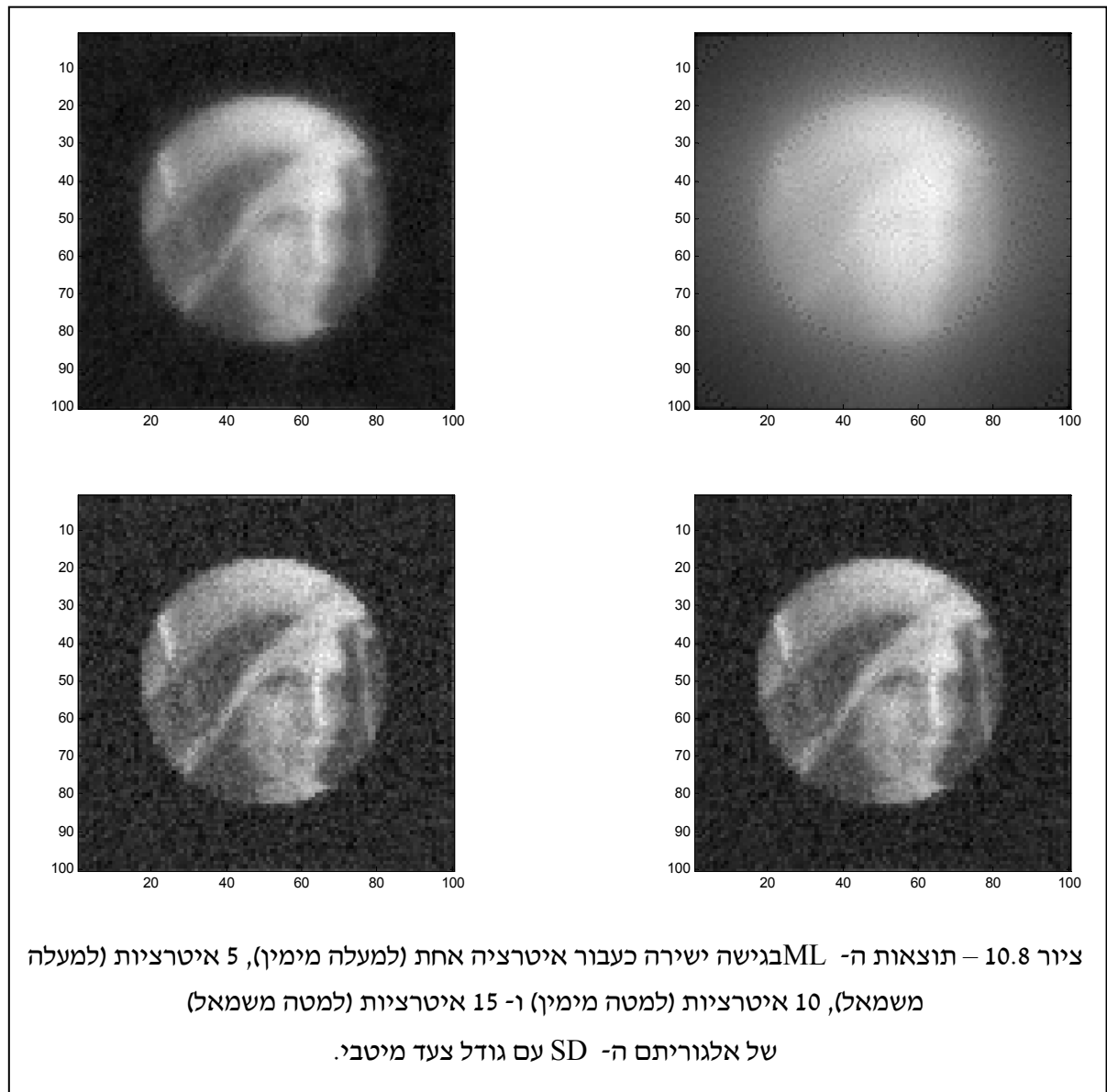
כל ערכי ההטלה זוהמו ברעש גאוסי ולבן בעל ממוצע אפס ושונות $\sigma = 200$ (יש לזכור כי ההטלות צוברות כ- 100 ערכי רמות אפור בתחום $[0, 255]$, ולכן עוצמת רעש זו מקבילה לרעש גאוסי בעל שונות $\sigma = 20$ בכל פיקסל).

ציור 10.7 מתאר את תוצאות אלגוריתם ה-LMS כעבור חצי סריקה מלאה וסריקה מלאה של מערכת המשוואות. אנו רואים כי כבר לאחר סריקה אחת, שמשמעה ביצוע R ו- R^T פעם אחת כל אחד, התוצאה כבר "טובה". בסימולציה זו בחרנו שלא לערב את הרגולריזציה. לכן התוצאה סובלת מתופעת "צילצולים" האופיינית כל-כך לפתרון ML ישיר.



בביצוע ה-LMS יצרנו את הוקטורים \underline{x}_j ממש כשם שאלו נבנו בעת הפקת ההטלות. לכן, גם אם יצרנו פגיעה בתהליך הדגימה של הישרים, פגיעה זו הופיעה גם בעת השחזור ואינה מהווה בעיה. ביישומים מעשיים בהם אין זה כך, יכול הפער בין המודל המייצר את ההטלות, והמודל בו משתמשים לשחזור להוות מקור לשגיאות גדולות. לעומת זאת נוכל לטעון כי סטיות בין הייצור האמיתי של ההטלות לבין המודל שישמש לשחזור נספג לתוך הרעש האדיטיבי שבמודל.

כיוון שהוקטורים \underline{x}_j נאספו, וכיוון שמימדי מערכת המשוואות מאפשרים זאת, צברנו גם את המטריצה R במלואה – זוהי מטריצה בגודל 10000 על 10000 שבה כל שורה כוללת עד 100 איברים שהם '1' (שפירושו – אלו נצטברו בהטלה). בהינתן מטריצה זו אנו יכולים להפעיל את הגישה הישירה כפי שתואר קודם. ציור 10.8 מראה תוצאת גישה זו לאחר $\{1,5,10,15\}$ איטרציות ה-SD עם גודל צעד אופטימלי לקבלת מהירות התכנסות מירבית.



פרק 11 – מבנים פירמידליים ושימושיהם בתמונות

235 פרק 11 – מבנים פירמידליים ושימושיהם בתמונות
236 11.1 הפירמידה הגאוסית
240 11.2 הפירמידה הלפלסיאנית
249 11.3 התמרת Wavelet ע"י ה-Lifting Scheme

מושג ששב ועולה במהלך הדיון בתמונות הוא מושג ה-"רזולוציה". רזולוציה של תמונה הינה שם נרדף למידת חדותה, אן לגודל הפרט הקטן ביותר הניתן בה לאבחנה. לא תמיד יותר פיקסלים משמעם יותר רזולוציה. בהקשר זה, הדיון במשפט הדגימה מציע מדד ברור הקושר בין מרווח הדגימה ובין הרזולוציה של התמונה, כשזו נמדדת במונחים תדריים. החולשה של תפיסה זו בהיותה גלובלית וחד-משמעית. באשר לגלובליות, משפט הדגימה מציע מרווח דגימה אחד התקף לכל התמונה – אין אנו יודעים לאפיין אותות בעלי רזולוציה תלוית מקום ומרווח הדגימה הנאות לייצוגם בכל מקום. באשר לחד-המשמעותיות, מדוע שניח כי במקום מסוים בתמונה יש רק רזולוציה אחת? אם נפעל כך, בהכרח נציע את הרזולוציה הגבוהה ביותר במקום, ונתעלם ממגוון תופעות הקורות באותה נקודה ואשר להן רזולוציה טבעית אחרת. שתי חולשות מהוות את התשתית לצורך במבנים פירמידליים בהם רזולוציה יכולה להשתנות ממקום למקום, וכל מקום יכול לקבל ייצוג נאות במגוון רזולוציות.

תפיסה מולטי-רזולוציונית כבר הוצגה קודם כשדנו בהתמרת Wavelet, אלא ששם בחנו את ההתמרה כהתמרה ולא כייצוג החושב "רזולוציה". בפרק זה נחזור על עיקרי הדברים העומדים מאחורי התמרה זו ודומות לה, אלא שההדגש יהיה הפעם על היות שיטות אלה שיטות ייצוג מולטי-רזולוציוניות המקנות גמישות בעיבודים שונים. אין גישה מאוחדת לתחום זה והדרך בה נבחר להציג את הדברים היא הצגת האפשרויות השונות ודרךן את היישומים הנובעים מהן.

11.1 הפירמידה הגאוסיאנית

ככלל, פירמידה הינה אוסף תמונות המאורגנות בסדר כלשהו (נציע סידור אנכי המתחיל למטה ועולה למעלה) ואשר כולן נובעות מתמונה אחת שעברה עיבוד. נניח כי נתונה לנו תמונה $X(m,n)$ בגודל N על N . נגדיר תמונה זו כתחתיתה של הפירמידה P_G ונסמך אותה כ- $P_G(0)$. נפעיל על תמונה זו מסנן LPF מסוג כלשהו ונקבל תמונה בה הפרטים מטושטשים קמעה. התמונה החדשה, $P_G(1)$, תהיה השכבה השנייה בפירמידה. נמשיך ונטשטש את התמונה $P_G(1)$ ונקבל את $P_G(2)$ וכך הלאה. אוסף התמונות שתקבלנה יהיה הפירמידה הרצויה, כאשר כל תמונה מייצגת רמת רזולוציה מסוימת.

נניח כי מסנן ה- LPF הוא כזה שמדכה בצורה נאותה תדרים גבוהים מעבר ל- $\pi/2$ (מחצית טווח התדרים). פירוש הדבר הוא שהתמונה לאחר סינון יכולה לעבור דילול ביחס 2:1 ללא אובדן. לכן, יחד עם פעולות הסינון תבואנה פעולות דילול והתמונות תקטנה. ציור 11.1 ממחיש את המבנה המתקבל.



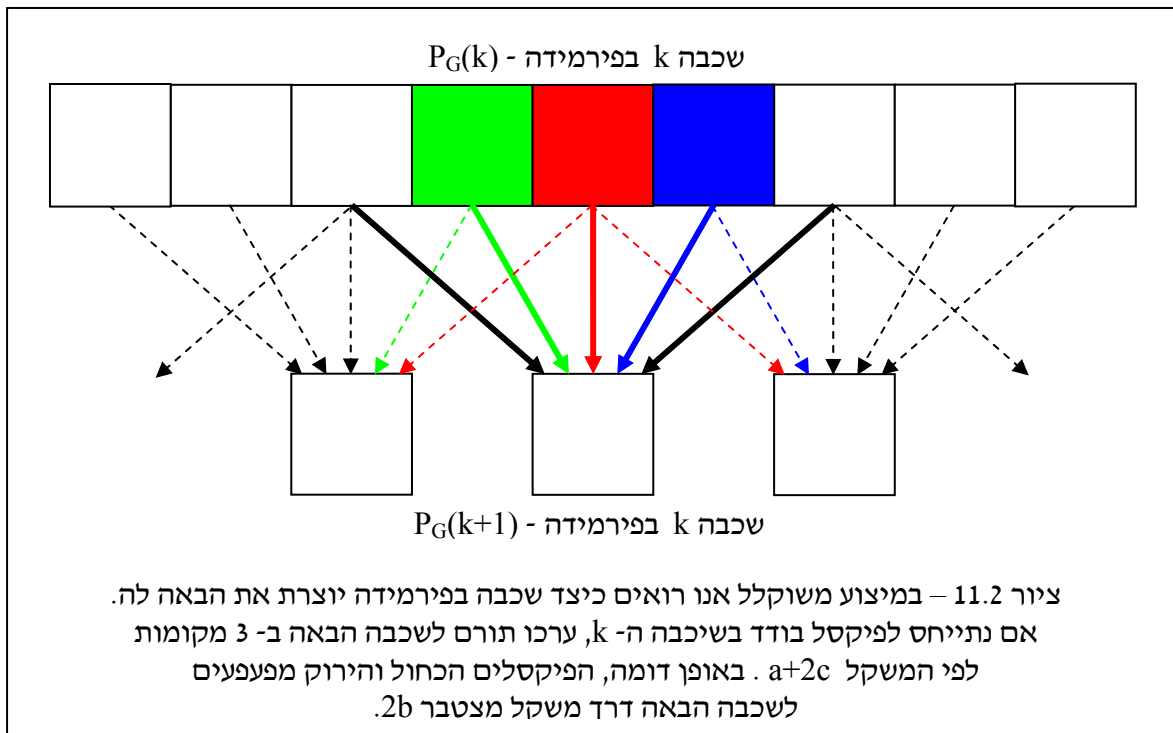
מהו המסנן בו יש לעשות שימוש? ישנן מגוון גישות לבחירת המסנן. לדוגמה, נוכל לחפש בחירה שתבטיח דיכוי תדרים גבוהים כיאות על-מנת שהדילול יהיה מותר ללא אובדן בקירוב טוב. בעבודתם הציגו Burt ו- Adelson גישה שונה – הם הציגו את הכללים הבאים אותם מסנן זה יקיים, בהנחה שזהו

מסנן ספרבילי באורך $2N+1$ שתגובת ההלם שלו היא $\{w_k\}_k$

$$\begin{aligned}
 1. \sum_{k=-N}^N w_k &= 1, & 2. \forall k > 0, w_k &= w_{-k} \\
 3. \forall k > 0, w_k &\leq w_{k-1}, & 4. \forall k, \sum_{j=-\infty}^{\infty} w_{k+2j} &= \text{const}
 \end{aligned}$$

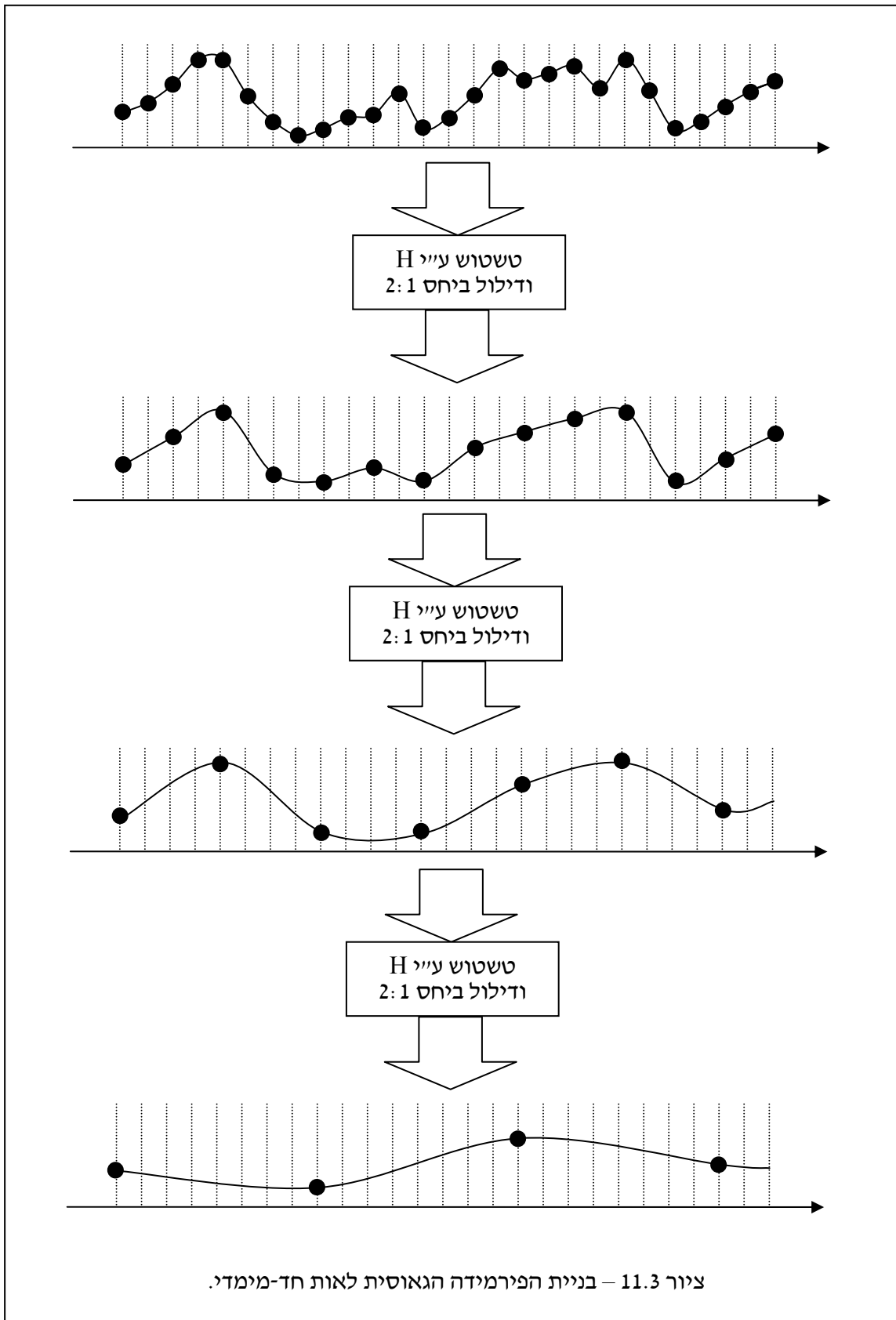
הדרישה הראשונה מבטיחה כי הפעלת מסנן זה על התמונה תשמר את הממוצע המקומי. הדרישה השנייה יוצרת מסנן סימטרי בו אין הטיה לצד אחד על פני האחר. הדרישה השלישית באה לוודא דעיכה מונוטונית של השפעת השכנים על הפיקסל המרכזי בסינון. הדרישה הרביעית והמורכבת מכולן פירושה שכל פיקסל תורם בצורה זהה ביצירת שכבת הפירמידה הבאה המדוללת. ציור 11.2 ממחיש עניין זה ומצדיק את ניסוח הדרישה כפי שנקבע במשוואה.

לשם דוגמה, אם נניח כי $N=2$ ונניח כי המסנן בנוי כ- $\{c \ b \ a \ b \ c\}$, הדרישה הראשונה תהיה $2c + 2b + a = 1$. הדרישה השנייה מוטמעת כבר בייצוג הפרמטרי שבחרנו, והדרישה השלישית תהיה $c \leq b \leq a$. הדרישה הרביעית תתורגם ל- $2c + a = 2b$. לכן ברור כי $b=0.25$ (צירוף דרישה ראשונה ורביעית), a צריך בהכרח להיות גדול מ- 0.25 (דרישה שלישית), ו- c יהיה $c=0.5(a-0.25)$. עבור הבחירה $a=0.4$ מתקבל מסנן דומה לפעמון גאומטרי ומכאן השם "הפירמידה הגאומטרית". ציור 11.3 ממחיש את דרך בניית הפירמידה הגאומטרית לאותות חד-מימדיים, והרי הדבר זהה בדו-מימד.



מה אפשר לעשות בעזרת פירמידה גאומטרית? לא הרבה. זהו בראש ובראשונה אמצעי לייצוג התמונה במגוון רזולוציות. בעוד השכבה התחתונה עושה מאמץ להראות את כל הפרטים, שכבות עליונות יותר זורקות אט-אט את הפרטים הקטנים ומתרכזות בסיפור הכללי של התמונה. לדוגמה, בציור 11.1, נוצת הכובע המפורטת בשכבה הראשונה הופכת לחסרת פרטים מהשכבה השלישית ואילך.

נזכיר מספר יישומים בהם עושים שימוש בפירמידה גאומטרית. מבנה פירמידלי זה יכול לעזור בחיפוש תבניות בתמונה. אחת הדרכים הפשוטות ביותר והישירות ביותר לחפש תבנית בתמונה היא שיטת הקורלציה שדומות לה כבר פגשנו בשערוך תנועה של בלוקים (ראה פרק 9). בהינתן תמונה X בגודל



N על N , ובהינתן תבנית Y בגודל L על L ($L \ll N$) אותה אנו מחפשים (נניח פרצוף), החיפוש יכול להיעשות ע"י החישוב

$$C[n, m] = \frac{\left(\sum_{k=1}^L \sum_{j=1}^L X[n+k, m+j] Y[k, j] \right)}{\sqrt{\left(\sum_{k=1}^L \sum_{j=1}^L X[n+k, m+j]^2 \right) \left(\sum_{k=1}^L \sum_{j=1}^L Y[k, j]^2 \right)}}$$

בכל מקום $[n, m]$ בתמונה, וקביעת מציאה במקום בו ערך זה גדול מסף מסוים. זוהי כמות עצומה של חישובים וניתן להקטינה בצורה משמעותית אם נבנה פירמידה גאוסית ל- X ול- Y . במקרה כזה, חיפוש מקדים בשכבות גבוהות יותר של הפירמידה יכול לגלות איזורים חשודים בהם ייעשה חיפוש גם ברזולוציות גבוהות (דהיינו, שכבות תחתונות של הפירמידה). שימוש אחר הוא כאשר הפירמידה נעשית רק על X והחיפוש נעשה בכל השכבות עם Y המקורי. גישה זו נחוצה כאשר מחפשים את התבנית בידיעה כי היא צפויה להופיע במגוון גדלים אפשריים.

למעשה, ניתן להשתמש בפירמידה הגאוסית לצורך שערך תנועה בבלוקים והדבר יכול להביא להאצה ולהגדלת הדיוק בתהליך השערך. לדוגמה, נניח שערך התנועה צריך להיעשות בתמונות בגודל N על N בבלוקים של 32 על 32, וגדלי וקטורי התנועה הם 16 פיקסלים לכל כיוון. ניתן להמיר תהליך זה בשערך תנועה על צמד תמונות שמהוות שיכבה שנייה בפירמידה הגאוסית ואשר גודלן הוא $N/2$ על $N/2$. הפעם גדלי הבלוקים הם 16 על 16, ולכן מספר וקטורי התנועה לא קטן. אבל עם השינוי שבוצע איזור החיפוש של הוקטור קטן בפקטור 4 (כי וקטורי התנועה המירביים הם 8 לכל כיוון כעת) וכך גם התמך לחישוב ה- MSE (בחישוב ההתאמה נצבור 256 איברים ולא 1024). פירוש הדבר חסכון בפקטור 16 רק במעבר לשכבת הפירמידה השנייה. למעשה, אין סיבה לעצור כאן, וניתן לעלות עוד ועוד כשבכל שכבה הבלוק ואיזור החיפוש קטנים. ברור כי יש לעצור לפני שהבלוק מגיע לגודל קטן מידי שיביא לשערך לא אמין, וכך ראוי לעצור בגודל 4 על 4 (שכבה רביעית במקרה זה).

בהינתן שערך משכבה $k+1$, יש לעדכנו ע"י ירידה חזרה לשכבה ה- k . עדכון זה צריך להיעשות בחיפוש מקומי סביב הוקטור שנמצא בשלב קודם (כשהוא מוכפל ב-2). באופן כזה הקטנו באופן משמעותי את כמות החישובים הדרישה למציאת וקטור התנועה. לכן, אם נדרשנו במקור לאיזור חיפוש בגודל 32 על 32 שמשמעו 1024 נקודות בדיקה של מיקום הבלוק, הרי שהפעם בשכבה הרביעית נבצע חיפוש ב-4 על 4 מקומות ועם בלוק קטן משמעותית, כך שכמות החישובים זניחה. עם הפעפוע לשכבה הבאה, נחפש בכל פעם רק ב-8 מקומות סביב הפתרון הקודם, וכך נחסוך באופן גס ביחס של כ-100 בכמות החישובים הכוללת, כיוון שרק החישובים של השיכבה התחתונה בפירמידה מהותיים בהשוואה.

בנוסף, הדיוק של השערך צפוי לגדול משתי סיבות - אנו נוכל בלבצע חיפוש באיזור חיפוש גדול יותר, ומכיוון שהוקטור הראשוני מחושב ע"פי מתווה כללי בתמונה תוך התעלמות מפרטים קטנים שיכולים לסחוף השערך לערך לא רלוונטי. עם הפעפוע מטה בפירמידה, אנו מגבילים את העדכון לסביבת הפתרון הקודם, וכך מכוונים את החיפוש למקום הראוי. הדבר דומה למינימיזציה של פונקציה מאוד לא חלקה וחשש לנפילה למינימה לוקאלית. המעבר לשכבות הרזולוציה הבאות בפירמידה משול להחלקת פונקציה זו ומחיקה של מינימות לוקליות כתוצאה מכך.

11.2 הפירמידה הלפלסיאנית

החסרון המרכזי של הפירמידה הגאוסית הוא ששכבות הפירמידה לא כוללות רזולוציות זרות המצטרפות זו לזו, אלא כל שכבה כוללת הפחתה מסך הרזולוציות. כך יוצא שהשכבה הראשונה בפירמידה כוללת את כל מרכיבי התמונה (זו התמונה ממנה אנו יוצאים לדרך!), וכל שכבה מעליה כוללת תוכן רזולוציוני המוכל בתמונות שמתחתיה. חסרון זה קל לתיקון ע"י החסרת תמונות – דבר המביא לבנייתה של הפירמידה הלפלסיאנית.

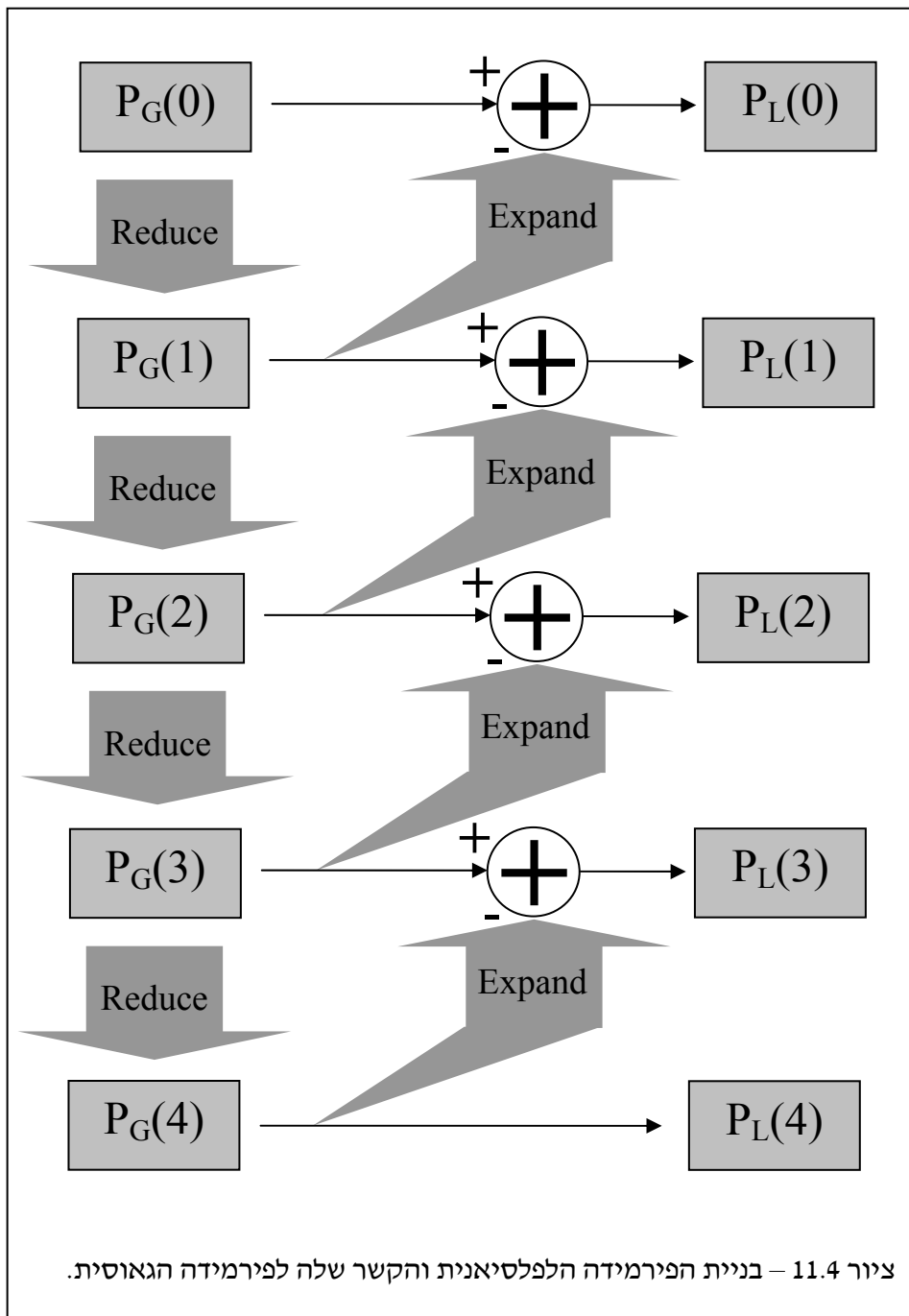
על רקע הנאמר לעיל, נציע לבנות את הפירמידה הלפלסיאנית P_L באופן הבא – את התמונה X בגודל N על N נטשטש ונקבל תמונה HX בגודל זהה. נחסר ונגדיר את השכבה הראשונה בפירמידה ע"י $P_L(0)=X-HX$. זוהי תמונת הפרשים בגודל המקורי N על N המכילה את הפרטים הקטנים כשהם מתוארים כהפרשים של הערך האמיתי במקום פחות ממוצע מקומי שלו – כבר ראינו בעבר כי זוהי פעולת הלפלסיאן על תמונה. את התמונה HX נוכל לדלל כשם שעשינו בפירמידה הגאוסית P_G . על התמונה המתקבלת שגודלה $N/2$ על $N/2$ נוכל לחזור על התהליך – טשטוש, החסרה שיוצרת את השכבה הבאה בפירמידה, ודילול. נקבל סידרת תמונות תמונות הפרש ובראש הפירמידה תמונה קטנה המתלכדת עם התמונה בראש הפירמידה הגאוסית.

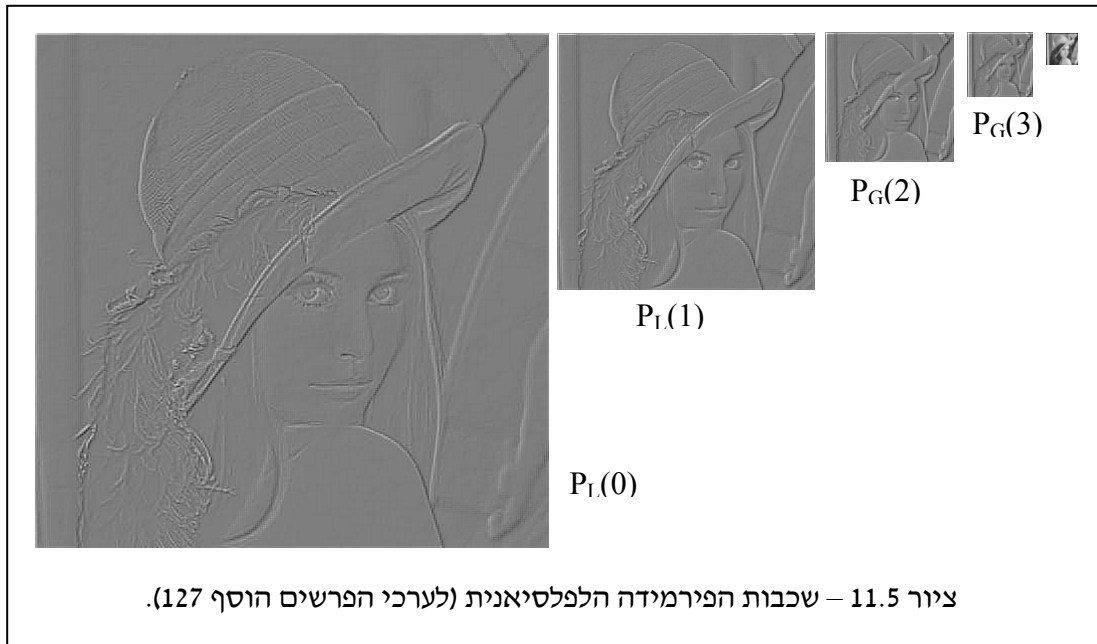
הגיוני ומפתה ככל שיהיה, תהליך פשוט זה של בניה סובל מחולשה מהותית והיא חוסר היכולת לחזור לתמונה המקורית. זאת כיוון שבפעולת הדילול אנו זורקים מידע שאינו מצוי בעת השחזור. אם נתייחס לבסיס הפירמידה, התמונה $P_L(0)$ מכילה בצורה נאותה את התדרים הגבוהים של התמונה. אבל, תדרים הנמוכים אשר אמורים להישמר בתמונה HX נפגמו בשל דילול תמונה זו לשם בניית צעדי הפירמידה הבאים.

תיקון הבעיה יכול להיעשות באופן הבא. במקום להגדיר את $P_L(0)$ כהפרש בין X ל- HX , נגדיר תמונה זו כהפרש בין X ל- HX לאחר דילול והרחבה. כלומר, התמונה HX תעבור דילול כמקודם, ואז תנופח חזרה לגודל המלא באינטרפולציה (כלשהי! כל אינטרפולציה תספק את הצורך בשחזור התמונה מהפירמידה!). מתמונה מוגדלת זו ייעשה החיסור. במונחים אלגבריים, נגדיר את האופרטור Reduce כאופרטור מטשטש ומדלל, והאופרטור Expand כאופרטור המבצע הגדלה חזרה. אזי הפירמידה תיבנה בסידרת הצעדים הבאים, כשאנו מסתייעים בבניית הפירמידה הגאוסית:

1. התחלוי מהתמונה X . זוהי גם $P_G(0)$ - שכבת הפירמידה הגאוסית התחתונה.
2. קבע את k ל-1.
3. בצעד ה- k , בצעו פעולת Reduce על $P_G(k-1)$ וקבלו את התמונה $P_G(k)$ - שכבת הפירמידה הגאוסית ה- k .
4. בצע פעולת Expand על $P_G(k)$ והגדר $P_L(k-1) = P_G(k) - \text{Expand}\{P_G(k)\}$.
5. הגדל את k . אם הגענו לשכבה $k=M$ קבע $P_L(k) = P_G(k)$ ועצור. אחרת חזור לשלב 3.

בבניה זו של הפירמידה, בניית התמונה המקורית פשוטה ותיעשה בצורה רקורסיבית ע"י הפעלת Expand על התמונה בראש הפירמידה $P_L(M)$, הוספת השכבה ה- $M-1$, שוב Expand וחוזר עד להגעה לתחתית הפירמידה הלפלסיאנית, כשהתמונה הנוצרת זהה במדויק לזו ממנה התחלנו. ציור 11.4 ממחיש את דרך בניית שתי הפירמות והקשר ביניהן, וציור 11.5 מראה את התמונות המתקבלות בפועל בהקבלה לציור 11.1. תהליך בניית התמונה מהפירמידה הלפלסיאנית מתואר בציור 11.6 - פעפוע מעלה בציור, כשמתחילים בשכבה $P_L(4)$ מוביל לתמונה הרצויה, וקל לראות את הדמיון לתהליך הבניה בציור 11.4.





ציור 11.5 – שכבות הפירמידה הלפלסיאנית (לערכי הפרשים הוסף 127).

כיצד נבחר את פעולת ה-Expand? אמרנו קודם כי לא חשוב כיצד תוגדר, בניית התמונה מהפירמידה תביא לשחזור מדויק כל עוד התהליך עקבי במובן שאותה פעולת Expand משמשת בבניית הפירמידה ובשחזור. בציור 11.5 בנינו את הפירמידה על ידי פעולת Pixel Replication בו בעליה ברזולוציה כל פיקסל מייצר רביעיית ערכים זהים לו. זוהי קרוב לוודאי הגישה הפשוטה ביותר האפשרית וגם הגרועה ביותר.

מה ההבדל בין בחירות שונות? נענה על שאלה זו דרך מתן אינטרפרטציה מעט שונה לפירמידה

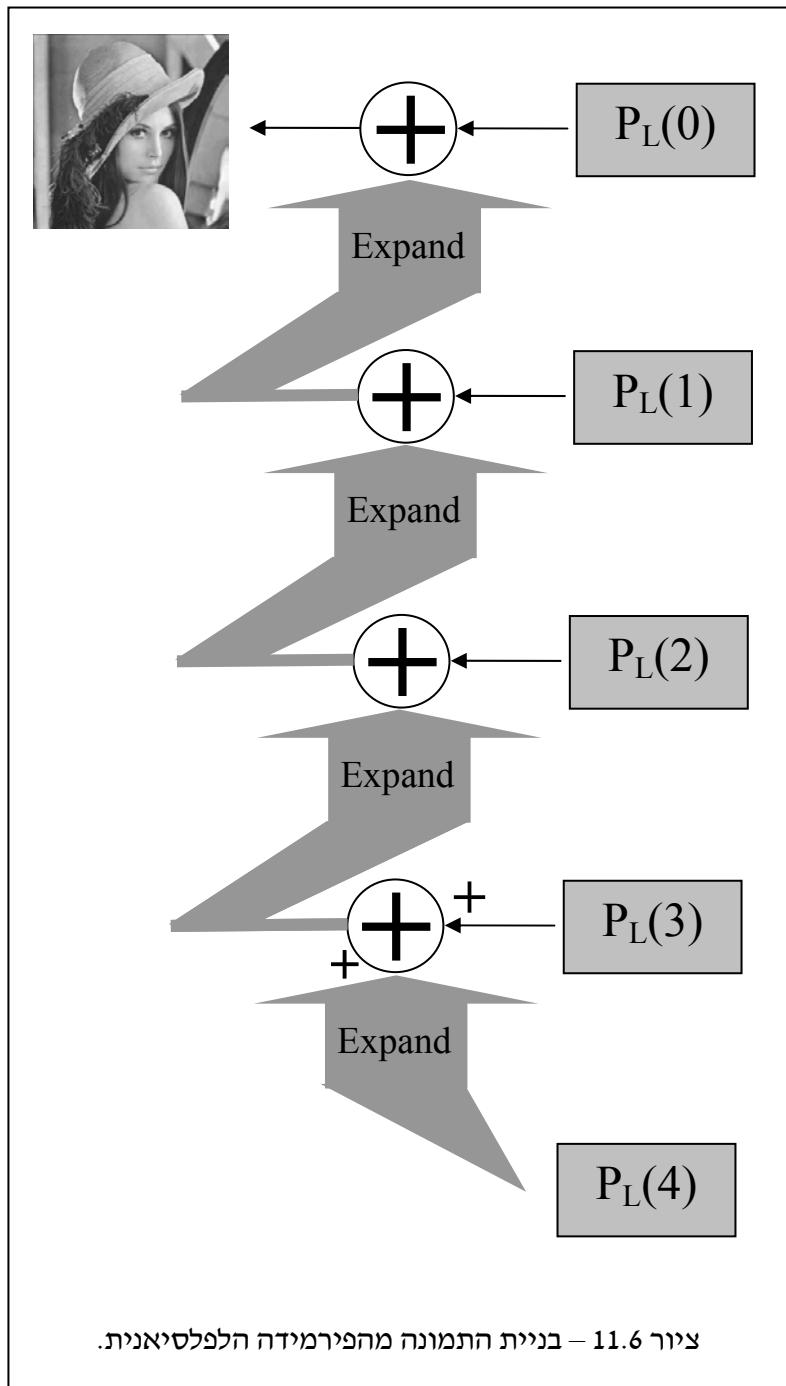
הלפלסיאנית. כשאנו בונים את התמונה $P_L(k-1)$ אנו עושים זאת ע"י המשוואה

$$P_L(k-1) = P_G(k-1) - \text{Expand}\{P_G(k)\} = \\ = P_G(k-1) - \text{Expand}\{\text{Reduce}\{P_G(k-1)\}\}.$$

כלומר, התמונה ה- $k-1$ בפירמידה הגאוסית היא תמונת היסוד בבניה זו. אנו מחשבים את התמונה $P_L(k-1)$ ומשתמשים בה בחיזוי של התמונה המקורית $P_G(k-1)$. תמונת החיזוי היא תמונה נטולת פרטים והיא מנסה לנבא את התמונה שמתחתיה (המפורטת יותר) בפירמידה הגאוסית. אם אין פרטים חדים בתמונה $P_G(k-1)$, החיזוי יביא לשארית קטנה. חיזוי מוצלח הוא חיזוי בו התמונה $P_L(k-1)$ כוללת אנרגיה דלה. אם החיזוי מוצלח זה יכול להיות בגלל שתי סיבות שונות:

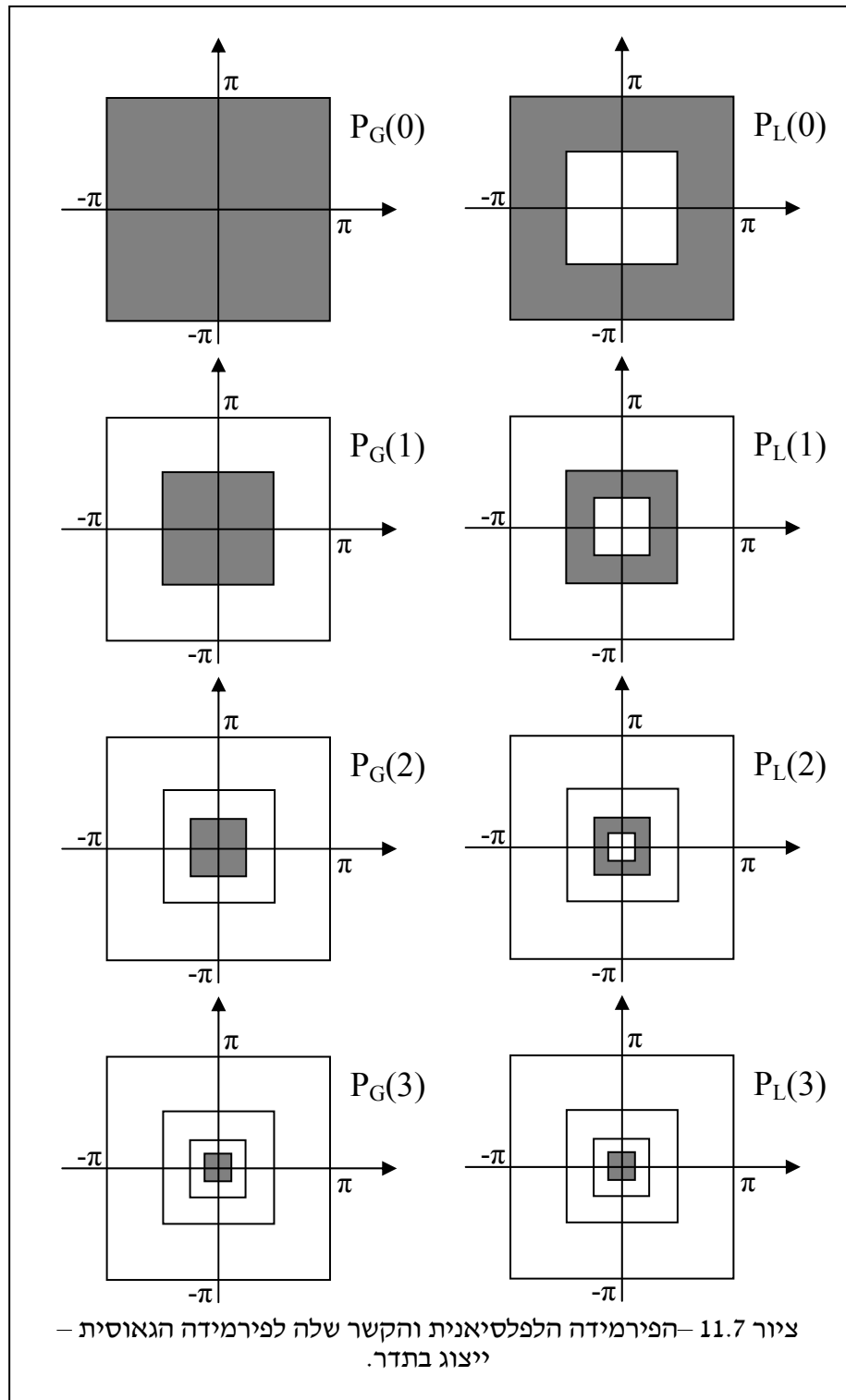
1. יתכן כי תוכן התמונה הוא כזה שאין רמת פרטים גבוהה (רזולוציה גבוהה). במקרה כזה החיזוי ייתן ערכים קרובים מאוד לאפס. במקרה כזה הפירמידה מצליחה לומר לנו את מידת הרזולוציה האמיתית של התמונה (ואפילו כתלות במקום) לפי העוצמה של תמונות הפירמידה בכל מקום. אם לדוגמה מהשכבה השנייה ומטה מתקבלים ערכי אפס בחצי העליון של התמונה בעוד שערכי שכבות אלה בחלק התחתון שונים מהותית מאפס, נדע כי החצי העליון מוגבל ברזולוציה ודי היה בפחות פיקסלים בפקטור 4 לתיאור אזור זה.

2. ללא קשר לנקודה הקודמת, ישנן דרכים שונות להגדיר את הפעולה Expand וכל דרך תיתן שגיאת חיזוי שונה. רצוי לבחור בהגדרת Expand שתיתן חיזוי מוצלח – חיזוי שיקטין את אנרגיית השארית ב- $P_L(k-1)$. עם זאת, יש גם עניין בבחירה של פעולה פשוטה שקלה לביצוע מבחינה חישובית.



ציור 11.7 ממחיש את משמעות הפירמידה הלאפלסיאנית (והגאוסיאנית) בתדר. אנו רואים את התמך התדרי $[-\pi, \pi]^2$ בו התמונה ממלאת מידע. הפירמידה הגאוסית לוקחת רביע פנימי המתקבל ע"י מסנן מעביר נמוכים הולך וצר בכל שלב וכך מתרכזות בעליית השכבות בתדרים הולכים וקטנים. הפירמידה הלפלסיאנית בונה תמונות הנבנות כמסנן מעביר פס כשהפס נע לעבר הראשית ונהיה צר יותר

ויותר עם התקרבותו לראשית. מציור זה גם ברור כי כל שעלינו לעשות הוא לסכם את רצועות התדר הללו בפירמידה הלפלסיאנית על מנת לקבל את התמונה המלאה. עניין מטעה מעט הוא נושא הדילול שלא ניכר מהציור בתדר, כיוון שמדובר בהתמרה מנורמלת.



במקור הוצעה הפירמידה הלפלסיאנית כפירוק של תמונה לשם דחיסתה. הרעיון פשוט - בניית הפירמידה לתמונה, ואז ביצוע קוונטיזציה לשכבותיה כך שערכים קרובים לאפס מאופסים, וכל ערך אחר מומר לערכי ייצוג קלים לדחיסה. ניתן אף לחשוב על קוונטיזר שונה לכל שכבה. בפועל ביצועי אלגוריתם

זה נופלים מאלה של JPEG. יש לזכור כי תמונה בגודל N על N מומרת בתהליך בניית הפירמידה ל- $4/3N^2$ פיקסלים - N^2 לתמונה $P_L(0)$, $N^2/4$ לתמונה $P_L(1)$, וכך הלאה, לקבלת הטור

$$\sum_{k=0}^{\log N} \left[\frac{N}{2^k} \right]^2 \approx N^2 \cdot \sum_{k=0}^{\infty} 4^{-k} = \frac{N^2}{1 - 1/4} = \frac{4N^2}{3}$$

לכן, כבר לאחר בניית הפירמידה יש לנו ניפוח של כמות המידע ב- 33%. לעומת זאת, המידע במבנה החדש הינו נמוך אנטרופיה כיוון שכולו מידע הפרשים שעבורו כבר ראינו כי הפילוג בעל צורה לפלסיאנית (הנה עוד סיבה לשמה של הפירמידה).

אם רצוננו בדחיסה, נוכל גם לבחור את הפעולה Expand באופן מיטבי להקטנת שגיאות החיזוי, ועימן את האנטרופיה. ניתן להציע גישות שונות לבחירת Expand. דרך אטרקטיבית היא התייחסות לתהליך הניפוח כבעיית Up-Scaling של תמונה ועשיית שימוש בכלים שלמדנו מפרק 6 על שחזור תמונות תוך שימוש ב-Prior לבניית חיזוי מוצלח יותר.

מעניין לציין כי אלגוריתם דחיסה זה מצוי תיאורטית באמצע הדרך בין גישות החיזוי וההתמרה שנדונו בהרחבה בפרק 8 כדרכים לדחיסת תמונות. עד כה נתנו אינטרפרטציה של חיזוי לפירמידה הפלסיאנית, אם כי זהו חיזוי מעניין בהיותו לא סיבתי כמו החיזויים בהם דנו בפרק 8, ומכאן צומח פוטנציאל לביצועים טובים יותר מגישות הסורקות את התמונה באופן סיבתי חד-מימדי.

ניתן גם לחשוב על בניית הפירמידה כהתמרה ואפילו התמרה ליניארית. לצורך זה נניח כי פעולות ה-Reduce וה-Expand הינן פעולות ליניאריות וככאלה הן ניתנות לייצוג כמטריצות R_k ו- E_k . המטריצה R_k מקבלת תמונה בגודל $(N/2^k) \times (N/2^k)$ לאחר סידור לקסיקוגרפי ומניבה את תוצאת הטשטוש במסנן ליניארי כלשהו (ואפילו תלוי מקום) כשלאחריו דילול. לכן מטריצה זו תהיה מלבנית עם פחות שורות מעמודות, בגודל של $(N^2/2^{2k+2}) \times (N^2/2^{2k})$. באופן דומה, המטריצה E_k מקבלת תמונה בגודל $(N/2^{k-1}) \times (N/2^{k-1})$, ומבצעת הגדלה ואינטרפולציה לקבלת תמונה גדולה פי 2 בכל ציר (שוב – כוקטור בייצוג לקסיקוגרפי). לכן מטריצה זו תהיה בגודל $(N^2/2^{2k}) \times (N^2/2^{2k+2})$.

כדוגמה לצמד מעניין של מטריצות ה-Reduce וה-Expand, ניתן לבחור את R_k כך שיבצע טשטוש כפירמידה הגאוסית עם $\{w_k\}_k = \{0.05, 0.25, 0.4, 0.25, 0.05\}$ ואחריו דילול ביחס 1:2. את המטריצה E_k נוכל לבחור כפעולת ה-Adjoint דהיינו $E_k = R_k^T$. בשל דרך בניית ערכי המסנן לפירמידה הגאוסית ייצא כי אינטרפולציה זו מתאימה (אם כי נדרש לכפול ב-4 את תוצאתה, כפי שנראה). אם נכתוב מספר שורות אופייניות ב- R_k כפעולה חד-מימדית, הן יראן כך

$$R_k = \begin{bmatrix} & & & & \vdots & & & & \\ & 0.05 & 0.25 & 0.4 & 0.25 & 0.05 & & & \\ \dots & & 0 & 0.05 & 0.25 & 0.4 & 0.25 & 0.05 & \dots \\ & & & & 0.05 & 0.25 & 0.4 & 0.25 & 0.05 \\ & & & & \vdots & & & & \end{bmatrix}$$

פעולת השחלוף תיתן מטריצה בעלת המבנה

$$E_k = \begin{bmatrix} & & & & \vdots & & & & \\ & & & & 0.05 & & & & \\ & & & & 0.25 & & & & \\ & & & & 0.4 & 0.05 & & & \\ & & & & 0.25 & 0.25 & & & \\ \dots & & 0.05 & 0.4 & 0.05 & \dots & & & \\ & & & & 0.25 & 0.25 & & & \\ & & & & 0.05 & 0.4 & & & \\ & & & & & 0.25 & & & \\ & & & & & 0.05 & & & \\ & & & & \vdots & & & & \end{bmatrix}$$

אנו רואים כי שורות מטריצה זו מסתכמות כולן ל- 0.5 – דבר המבטיח כי צירופים ליניאריים שנעשים ע"י מטריצה זו יכולים להיחשב כאינטרפולציה. לכן יש לכפול ב- 2 את התוצאה לשם תיקון הערך הממוצע. תכונת סכומי השורות הקבועים נובעת מהדרישה הרביעית של תרומה שווה בתכנון המסנן לפירמידה הגאוסית. במעבר לדו-מימד, כיוון שכל הפעולות פרידות, מכפלת קרונקר של מטריצות אלה יתנו את הפעולה הדו-מימדית הכוללת, ולכן נדרש גורם תיקון של 4.

בעזרת מטריצות אלו, בניית הפירמידה הגאוסית תיראה כך, כשאנו מתחילים בתמונת המקור X

בגודל N על N (ניקח דוגמא בה עומק הפירמידה הוא 5):

$$\begin{aligned} \underline{P}_G(0) &= \underline{X} \\ \underline{P}_G(1) &= R_0 \underline{X} \\ \underline{P}_G(2) &= R_1 R_0 \underline{X} \\ \underline{P}_G(3) &= R_2 R_1 R_0 \underline{X} \\ \underline{P}_G(4) &= R_3 R_2 R_1 R_0 \underline{X} \end{aligned}$$

מתוך קשרים אלו, בניית הפירמידה הלפלסיאנית תיעשה ע"י

$$\begin{aligned} \underline{P}_L(0) &= \underline{P}_G(0) - E_0 \underline{P}_G(1) = (I - E_0 R_0) \underline{X} \\ \underline{P}_L(1) &= \underline{P}_G(1) - E_1 \underline{P}_G(2) = (R_0 - E_1 R_1 R_0) \underline{X} \\ \underline{P}_L(2) &= \underline{P}_G(2) - E_2 \underline{P}_G(3) = (R_1 R_0 - E_2 R_2 R_1 R_0) \underline{X} \\ \underline{P}_L(3) &= \underline{P}_G(3) - E_3 \underline{P}_G(4) = (R_2 R_1 R_0 - E_3 R_3 R_2 R_1 R_0) \underline{X} \\ \underline{P}_L(4) &= \underline{P}_G(4) = R_3 R_2 R_1 R_0 \underline{X}. \end{aligned}$$

לכן, הפירמידה הלפלסיאנית אינה אלא המכפלה

$$\begin{bmatrix} \underline{P}_L(0) \\ \underline{P}_L(1) \\ \underline{P}_L(2) \\ \underline{P}_L(3) \\ \underline{P}_L(4) \end{bmatrix} = \begin{bmatrix} I - E_0 R_0 \\ (I - E_1 R_1) R_0 \\ (I - E_2 R_2) R_1 R_0 \\ (I - E_3 R_3) R_2 R_1 R_0 \\ R_3 R_2 R_1 R_0 \end{bmatrix} \underline{X}$$

במונחי התמרה בהם עסקנו בפרק 7, זוהי התמרה ליניארית (גם ההתמרה וגם היפוכה). בשל יצירת מסדרה של $4N^2/3$ מקדמי ייצוג, התמרה זו כוללת יתרות (Over-Complete).

יש לנו גם את ההתמרה ההפוכה – תהליך קריסת הפירמידה הלפלסיאנית לתמונת המקור. המשוואה לשימוש תהיה

$$\underline{X} = E_0 [E_1 [E_2 [E_3 \underline{P}_L(4) + \underline{P}_L(3)] + \underline{P}_L(2)] + \underline{P}_L(1) + \underline{P}_L(0)$$

בייצוג מטריצי, קשר זה יראה כך:

$$\underline{X} = \begin{bmatrix} I & E_0 & E_0 E_1 & E_0 E_1 E_2 & E_0 E_1 E_2 E_3 \end{bmatrix} \begin{bmatrix} \underline{P}_L(0) \\ \underline{P}_L(1) \\ \underline{P}_L(2) \\ \underline{P}_L(3) \\ \underline{P}_L(4) \end{bmatrix}$$

ומכפלת מטריצת ההתמרה ההפוכה במטריצת ההתמרה תיתן

$$\begin{aligned} \begin{bmatrix} I & E_0 & E_0 E_1 & E_0 E_1 E_2 & E_0 E_1 E_2 E_3 \end{bmatrix} \begin{bmatrix} I - E_0 R_0 \\ (I - E_1 R_1) R_0 \\ (I - E_2 R_2) R_1 R_0 \\ (I - E_3 R_3) R_2 R_1 R_0 \\ R_3 R_2 R_1 R_0 \end{bmatrix} &= \\ = I - E_0 R_0 + E_0 (I - E_1 R_1) R_0 + E_0 E_1 (I - E_2 R_2) R_1 R_0 + & \\ + E_0 E_1 E_2 (I - E_3 R_3) R_2 R_1 R_0 + E_0 E_1 E_2 E_3 R_3 R_2 R_1 R_0 &= I \end{aligned}$$

אנו רואים כי בדרך הגדרת ההתמרה ההפוכה בנינו בפועל את מטריצת הפסאודו-היפוך של מטריצת ההתמרה, ובכך הבטחנו חזרה לתמונה ממנה יצאנו. מכאן גם ברור כי בבחירת צמד הפעולות – התמרה קדימה והתמרה הפוכה – בחרנו בהתמרה שמביאה למינימום את מקדמי ההתמרה – דבר שמסתדר יפה עם הרצון להשתמש בתוצאה לדחיסה.

יישום מעניין אחר בו עושים שימוש בפירמידה לפלסיאנית הוא מיזוג תמונות ברעיון שזכה לשם Multi-Resolution Spline. נניח כי נתונות שתי תמונות שהתחבר ביניהן (לשם פסיפס). לרוב חיבור ישיר של התמונות מניב תפר צורם בשל אי התאמה של תנאי ההארה, סטיות קלות בתיאום ביניהן ועוד. באמצעות פירמידה לפלסיאנית ניתן למזג תמונות תוך כדי טשטוש נעים של התפר. לשם פשטות נניח כי שתי תמונות שוות גודל צריכות להתחבר כך שאחת תורמת את החצי הימני, והאחרת את החצי השמאלי. חיבור שני חצאים אלו ישירות יגרום כאמור לתפר בולט ולא טבעי. האלטרנטיבה המוצעת היא פירוק שני התמונות לפירמידות לפלסיאניות, ובניית פירמידה משולבת בה כל שכבה בנויה כמיזוג ישיר

של חצאי התמונה מהפירמידות של שתי התמונות. כאשר משחזרים את התמונה מהפירמידה החדשה, ראינו כבר כי תהליך הבניה מתואר ע"י

$$\begin{aligned} \underline{X} &= E_0[E_1[E_2[E_3 P_L(4) + P_L(3)] + P_L(2)] + P_L(1) + P_L(0) = \\ &= P_L(0) + \\ &\quad + E_0 P_L(1) + \\ &\quad + E_0 E_1 P_L(2) + \\ &\quad + E_0 E_1 E_2 P_L(3) + \\ &\quad + E_0 E_1 E_2 E_3 P_L(4). \end{aligned}$$

לכן, השכבה העליונה בעלת הרזולוציה הנמוכה תעבור סידרה ארוכה של צעדי Expand שבהם היא מטושטשת. לכן, אם שני החצאים אינם תואמים בשכבה זו, תבוא מריחה ותיצור תפר נקי בין השתיים. כשנרד עם שכבות הפירמידה, כל שיכבה עוברת סידרה הולכת וקצרה של פעולות Expand ולכן הטשטוש הולך ופוחת. בשכבה האחרונה נמזג את הפרטים ללא כל טשטוש. בצורה זו מידע נמוך תדר יטושטש חזק יותר ממידע גבוה תדר, ממש כפי שאנו רוצים שיקרה. זאת כיוון שכשמידע בתמונה חלק מאוד, נחוץ טשטוש רחב תמך לשם חיבור שני הנתחים, ולעומת זאת, כשהמידע פחות חלק, החלקה רחבה תעלים פרטים ותזיק. ציור 11.8 מראה תוצאת תהליך זה על מיזוג שני תמונות זהות שכוללות כל אחת "תנאי הארה אחרים" (ובפועל, תוספת של שתי שדות אקראיים וחלקים). פרטים נוספים על יישום זה והפירמידה הלפליסיאנית מצויים במאמריהם של Burt ו- Adelson.



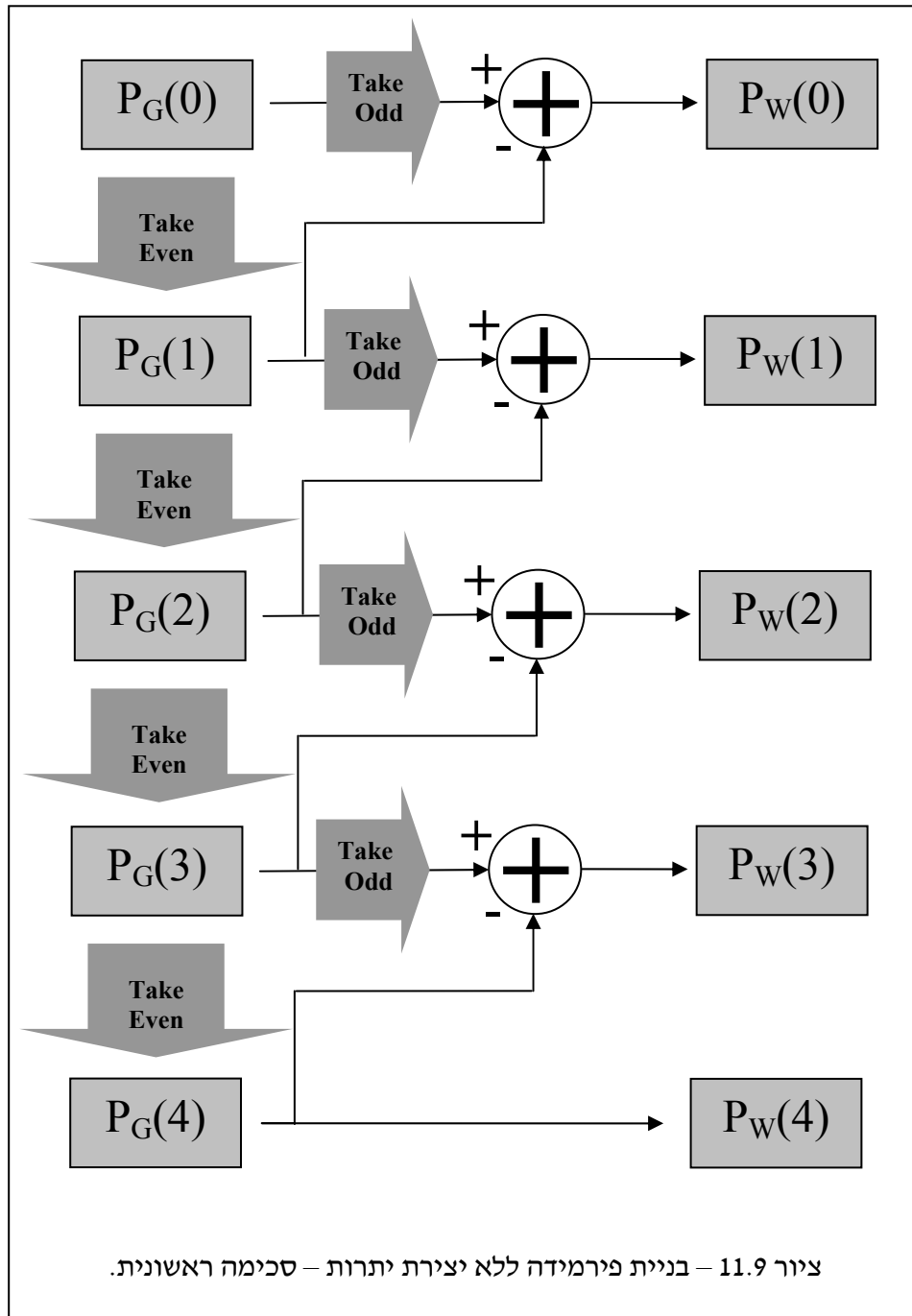
Lifting Scheme ע"י Wavelet - התמרת ה-11.3

הפירמידה הלפלסיאנית נתנה לנו ייצוג נאות של מגוון הרזולוציות הקיימות בתמונה, אך ראינו כי בפירוק זה יצרנו יותר מידע ביחס של $4/3$. דבר זה פעל נגדנו, למשל, כשביקשנו לדחוס בצורה יעילה את התמונה. האם ניפוח מידע זה חיוני? בפרק זה נראה כי ניתן להציע סכימה דומה מאוד לפירמידה הלפלסיאנית בה אין ניפוח מידע. הדבר המעניין שיצוץ מתוך הסכימה המוצעת אינו אלא התמרת ה-Wavelet בה כבר פגשנו בפרק 7. חשוב להבהיר כי פרק זה אינו בא להציג את התמרת ה-Wavelet מהזווית הקלאסית ולא יוכל להקיף את כל הידע בתחום – נקודת מבטנו צרה ותתמקד בייצוג מולטי-רזולוציוני ויעיל של התמונה הנתונה, תוך בניית פירמידה הדומה לפירמידה הלפלסיאנית.

נציע את התהליך הבסיסי הבא שיתואר לאות חד-מימדי X - וקטור באורך $N = 2^n$. כזכור, בבניית פירמידה גאוסית אות זה הוא גם השכבה התחתונה - $P_G(0)$. נפעיל פעולת Reduce על $P_G(0)$ ונקבל אות באורך $N/2 = 2^{n-1}$ לו קראנו $P_G(1)$. בשלב זה נניח כי פעולת ה-Reducer פשוט מבצעת דילול ובוחרת את האיברים הזוגיים בסידרה. בעזרת אות זה נבצע חיזוי של איבריו האי-זוגיים של הוקטור המקורי $P_G^{Odd}(0)$, ונחשב את שגיאת החיזוי. נגדיר וקטור שגיאות אלו כ- $P_W(0) = P_G^{Odd}(0) - P_G(1)$. וקטור זה יהיה באורך $N/2 = 2^{n-1}$ והוא יישמר כתיאור מרכיבי המידע העדין ביותר באות המטופל וזוהי שכבת האפס בפירמידה שאנו בונים כעת.

בשלב זה, בהתייחס ל- $P_G(1)$, נוכל לחזור על התהליך – פעולת Reduce לקבלת $P_G(2)$, חיזוי החלק האי זוגי מ- $P_G(1)$ (שנסמנו - $P_G^{Odd}(1)$), קבלת וקטור שגיאות חיזוי $P_W(1) = P_G^{Odd}(1) - P_G(2)$, וחוזר חלילה. תהליך זה יכול להימשך עד לשלב ה- $(n-1)$ - י בו $P_G(n)$ הוא סקלר. השכבה האחרונה בפירמידה שנוצרה, $P_W(n)$, תהיה פשוט $P_G(n)$. אוסף כל הדגימות שייווצרו יכלול רק N דגימות, ויצא שאין ניפוח בכמות המידע, כל השכבות למעט האחרונה כוללות ערכי הפרשים. פסי התדר המכוסים בשכבות אלו זהים למתואר בציור 11.7. ציור 11.9 ממחיש את דרך בניית הפירמידה החדשה המוצעת.

בהינתן הפירמידה P_W , בניית האות המקורי תיעשה ע"י היפוך סדר הפעולות. נתחיל ב- $P_W(n)$ אשר הינו האיבר השני $P_G(n-1)$ - זהו סקלר. נוסיף את ערכו ל- $P_W(n-1)$ (סקלר אף הוא), ונקבל את האיבר הראשון של $P_G(n-1)$. כך השלמנו את בניית $P_G(n-1)$. באופן דומה, הוספת $P_G(n-1)$ ל- $P_W(n-2)$ תיתן את החלק האי-זוגי ב- $P_G(n-2)$, והחלק הזוגי אינו אלא $P_G(n-1)$. תהליך זה יעלה כל הדרך עד לבניית $P_G(0)$ הוא אות המקור ממנו התחלנו. ציור 11.10 מראה תהליך זה.



דוגמה 11.1: ניקח את הסדרה הבאה ונפעיל עליה את התהליך שתואר.

$$\underline{X} = [1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7 \ 8 \ 9 \ 10 \ 11 \ 12 \ 13 \ 14 \ 15 \ 16] = P_G(0)$$

ההפרדה לאיברים הזוגיים והאי-זוגיים תיתן

$$P_G(1) = [2 \ 4 \ 6 \ 8 \ 10 \ 12 \ 14 \ 16] = \underline{X}_E$$

$$\underline{X}_O = [1 \ 3 \ 5 \ 7 \ 9 \ 11 \ 13 \ 15]$$

לאחר החיזוי נקבל

$$P_W(0) = [-1 \ -1 \ -1 \ -1 \ -1 \ -1 \ -1 \ -1]$$

הפרדת $P_G(1)$ לזוגיים ואי-זוגיים תיתן באופן דומה

$$\underline{P}_G(2)=[4 \ 8 \ 12 \ 16] \Rightarrow \underline{P}_W(1)=[-2 \ -2 \ -2 \ -2]$$

ובאופן דומה

$$\underline{P}_G(3)=[8 \ 16] \Rightarrow \underline{P}_W(2)=[-4 \ -4]$$

$$\underline{P}_G(4)=[8] \Rightarrow \underline{P}_W(3)=[-8]$$

$$\underline{P}_W(4)=[8]$$

ולכן פירמידת המוצא תהיה

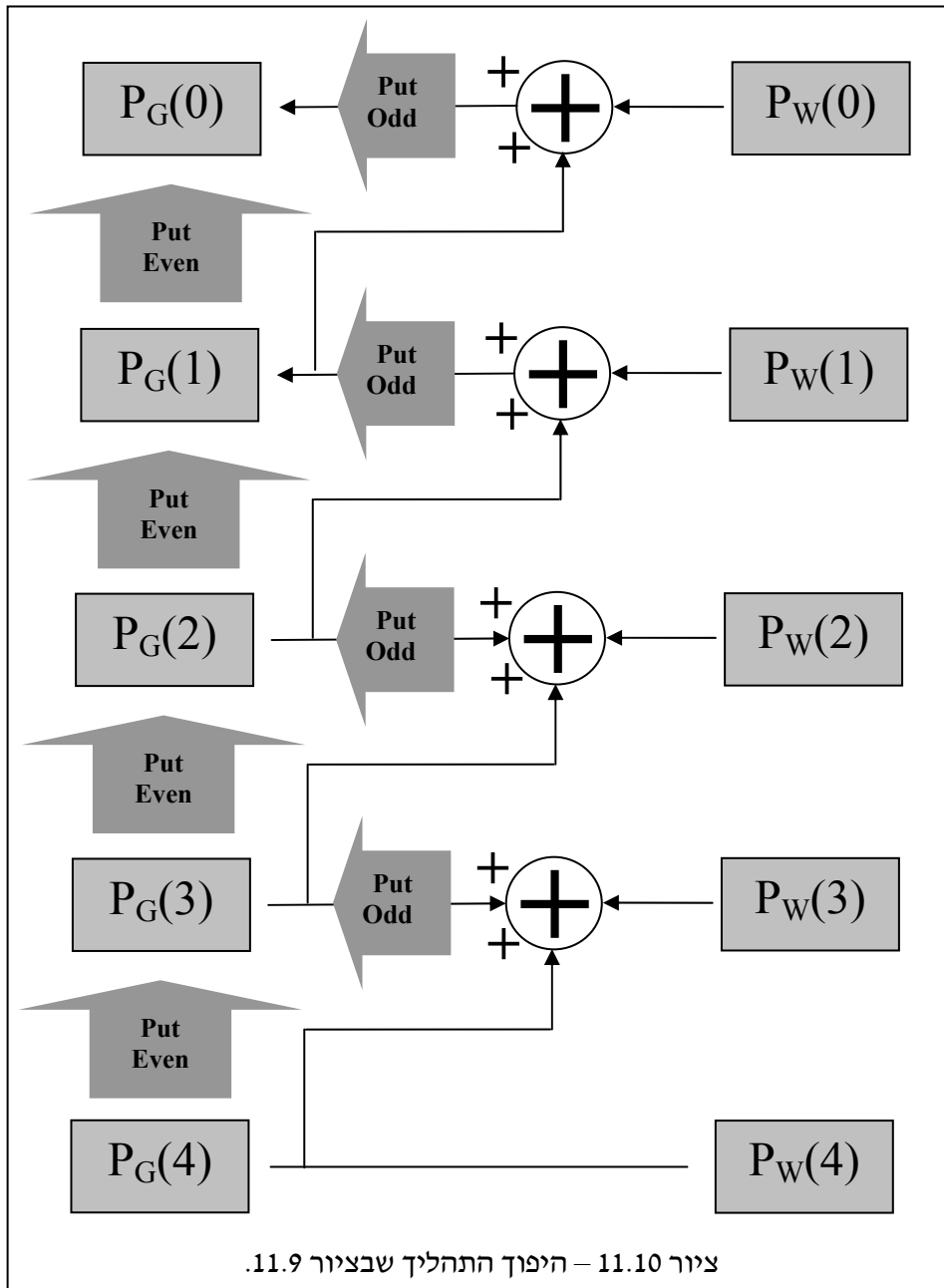
$$\underline{P}_W(0)=[-1 \ -1 \ -1 \ -1 \ -1 \ -1 \ -1 \ -1]$$

$$\underline{P}_W(1)=[-2 \ -2 \ -2 \ -2]$$

$$\underline{P}_W(2)=[-4 \ -4]$$

$$\underline{P}_W(3)=[-8]$$

$$\underline{P}_W(4)=[8]$$



נשים לב לכך שבחירת פעולת Reduce מורכבת יותר מאשר דילול פשוט תיצור בעיה בהיפוך, כיוון שביצועה בכיוון ההפוך אינה בהכרח אפשרית. עובדה מצערת זו מובילה לשאלה: האם נוכל להכליל סכימה זו להשתמש בפעולות Reduce ו-Expand כלליות יותר מאשר בחירת זוגיים ואי-זוגיים? הכללה כזו הוצעה תחת השם Lifting Scheme ע"י Sweldens באמצע שנות ה-90' והוצגה כדרך אלטרנטיבית מפתח לבניית התמרת ה-Wavelet. הרעיון הוא הבא – בהינתן סידרת המקור \underline{X} מוצע לפצלה לאיברים זוגיים \underline{X}_E ואי-זוגיים \underline{X}_O כמקודם. בעזרת סידרת הזוגיים \underline{X}_E יש לבנות חיזוי לאי-זוגיים ולהחליפם בערכי שגיאת החיזוי, דהיינו החיזוי יהיה פונקציה מהצורה $\text{Pred}\{\underline{X}_E\}$, וערכי האי-זוגיים יהפכו להיות $\underline{X}_O^{\text{New}} = \underline{X}_O - \text{Pred}\{\underline{X}_E\}$. למעשה יכולנו לעצור כאן ואז ציורים 11.9 ו-11.10 היו משתנים לכלול אבן בניין חדשה Pred באופן המתואר בציור 11.11, כשבציור זה אנו מתרכזים בשלב הראשון. ברור כי בהינתן $P_G(1)$ יכול כל התהליך לחזור על עצמו שוב ושוב עד להגעה לראש הפירמידה עם $P_G(n)$ סקלר.

בסכימה חדשה זו הבלוק Pred ייתן שיפור בביצועים כיוון ששגיאת החיזוי יכולה להיות קטנה יותר עבור בחירה נאותה. למשל, ניתן לבחור למצע שני שכנים מימינו ומשמאלו של האיבר האי-זוגי לשם חיזוי טוב. עבור הבחירה הכללית יותר

$$\text{Pred}\{\underline{X}_E\}[j] = \sum_{j=-L}^L \alpha_j \underline{X}_E[k-j] \Rightarrow \underline{X}_O^{\text{New}}[k] = \underline{X}_O[k] - \sum_{j=-L}^L \alpha_j \underline{X}_E[k-j]$$

נוכל לבחור מקדמים לבניית חיזוי אופטימלי ע"י מינימיזציה מהצורה

$$\text{Minimize}_{\alpha_{-L}, \dots, \alpha_L} \sum_{k=1}^{N/2} \left(\underline{X}_O[k] - \sum_{j=-L}^L \alpha_j \underline{X}_E[k-j] \right)^2$$

דוגמה 11.2: ניקח את הסדרה הבאה ונפעיל עליה את התהליך שתואר, כשנניח כי החיזוי Pred ממצע שכנים (ובקצוות חיסוי פשוט של שכן יחיד).

$$\underline{X} = [1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7 \ 8 \ 9 \ 10 \ 11 \ 12 \ 13 \ 14 \ 15 \ 16] = P_G(0)$$

ההפרדה לאיברים הזוגיים והאי-זוגיים תיתן

$$P_G(1) = [2 \ 4 \ 6 \ 8 \ 10 \ 12 \ 14 \ 16] = \underline{X}_E$$

$$\underline{X}_O = [1 \ 3 \ 5 \ 7 \ 9 \ 11 \ 13 \ 15]$$

לאחר החיזוי נקבל

$$\text{Pred}\{\underline{X}_E\} = [2 \ 3 \ 5 \ 7 \ 9 \ 11 \ 13 \ 15]$$

$$\underline{P}_w(0) = [-1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0]$$

הפרדת $P_G(1)$ לזוגיים ואי-זוגיים תיתן באופן דומה

$$P_G(2) = [4 \ 8 \ 12 \ 16] \Rightarrow \underline{P}_w(1) = [-2 \ 0 \ 0 \ 0]$$

ובאופן דומה

$$\begin{aligned} \underline{P}_G(3) &= [8 \ 16] \Rightarrow \underline{P}_w(2) = [-4 \ 0] \\ \underline{P}_G(4) &= [8] \Rightarrow \underline{P}_w(3) = [-8] \\ \underline{P}_w(4) &= [8] \end{aligned}$$

ולכן פירמידת המוצא תהיה

$$\begin{aligned} \underline{P}_w(0) &= [-1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0] \\ \underline{P}_w(1) &= [-2 \ 0 \ 0 \ 0] \\ \underline{P}_w(2) &= [-4 \ 0] \\ \underline{P}_w(3) &= [-8] \\ \underline{P}_w(4) &= [8] \end{aligned}$$

וזוהי תוצאה טובה יותר משום שמרבית האיברים הם אפסים. למעשה, גם את בעיית הקצה השמאלי בוקטורי שגיאת החיזוי ניתן לפתור ע"י שימוש בשני שכנים וביצוע אקסטרפולציה, ובשל פשטות הוקטור ממנו יצאנו, ניתן לקבל כי כל השכבות כוללות אפסים למעט שתי השכבות האחרונות. בוקטורי מידע בעלי תוכן מורכב יותר אנו צפויים לקבל תוכן שונה מאפס.

כל שתואר עד כה עדיין אינו סכימת ה-Lifting Scheme המלאה. לצורך השלמתה מוצע שלב

חישוב נוסף בו לאחר חישוב העדכון $\underline{X}_O^{New} = \underline{X}_O - \text{Pred}\{\underline{X}_E\}$ נבצע עדכון לערך האיברים הזוגיים לפי

$$\cdot \underline{X}_E^{New} = \underline{X}_E + \text{Update}\{\underline{X}_O^{New}\}$$

הרעיון בבסיס צעד נוסף זה הוא להיטיב את צורת האות \underline{X}_E לפני סיום החישוב כך שיהיה חלק יותר וככה בעל פוטנציאל דחיסות טוב יותר לצעד הבא. זהו בעצם המקור לשם Lifting כי פעולה זו היא שהופכת את המערכת להתמרת Wavelet. ציור 11.12 מראה את הסכימה המלאה.

אם נניח כי פעולות Predict ו-Expand הינן ליניאריות, נוכל לתאר את פעולתן כמטריצות P ו-U

המכפילות וקטורים. במקרה זה נקבל כי הסכימה המוצעת ניתנת לרישום ע"י

$$\begin{aligned} \underline{X}_O^{New} &= \underline{X}_O - P\underline{X}_E \\ \underline{X}_E^{New} &= \underline{X}_E + U\underline{X}_O^{New} = \underline{X}_E + U\underline{X}_O - UP\underline{X}_E \end{aligned}$$

או ברישום נוח יותר

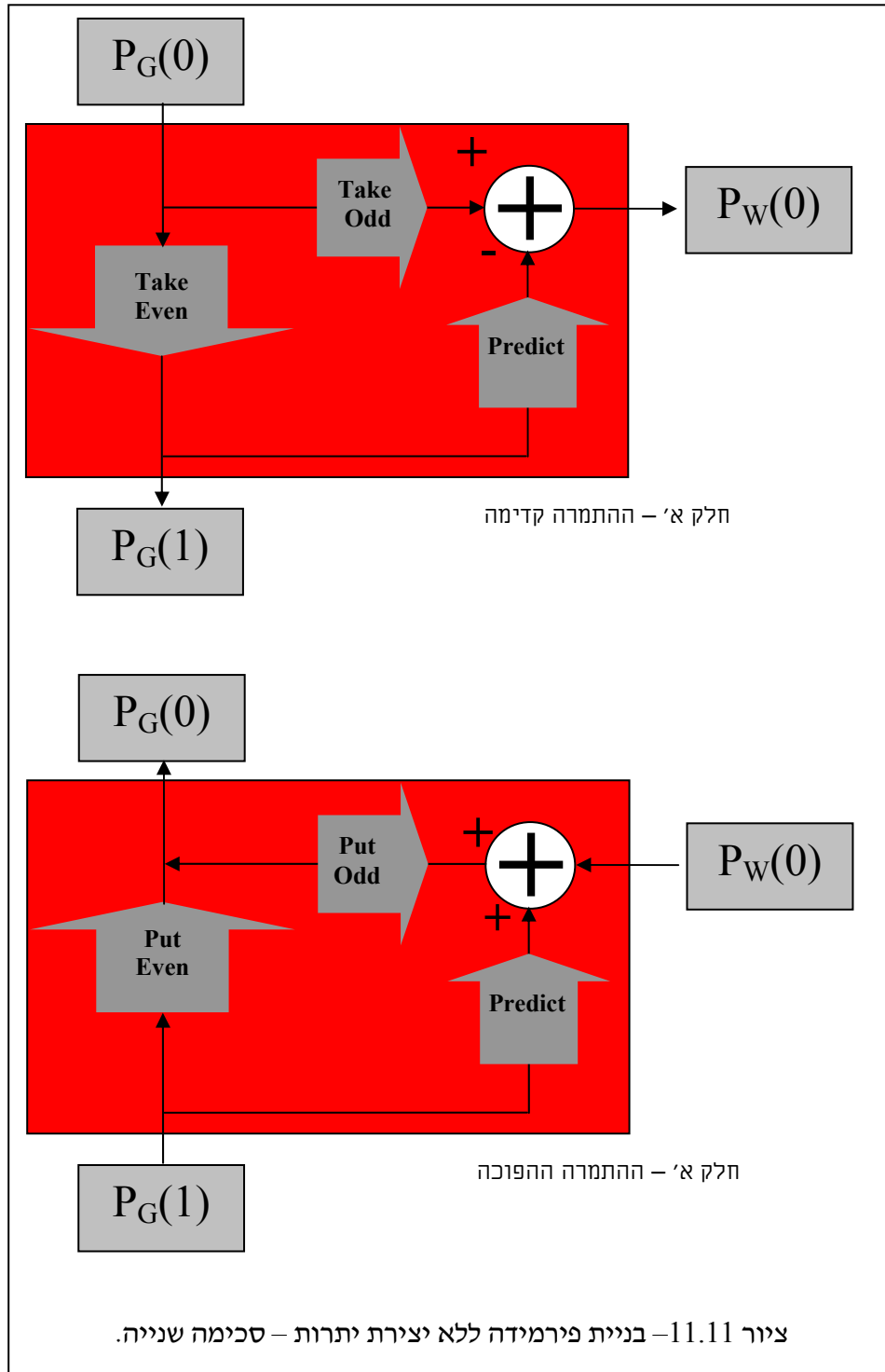
$$\cdot \begin{bmatrix} \underline{X}_O^{New} \\ \underline{X}_E^{New} \end{bmatrix} = \begin{bmatrix} I & -P \\ U & I - UP \end{bmatrix} \begin{bmatrix} \underline{X}_O \\ \underline{X}_E \end{bmatrix}$$

דוגמה 11.3: נחבר $P=I$ ו- $U=0.5I$ ונקבל

$$\cdot \begin{bmatrix} \underline{X}_O^{New} \\ \underline{X}_E^{New} \end{bmatrix} = \begin{bmatrix} I & -P \\ U & I - UP \end{bmatrix} \begin{bmatrix} \underline{X}_O \\ \underline{X}_E \end{bmatrix} = \begin{bmatrix} I & -I \\ 0.5I & 0.5I \end{bmatrix} \begin{bmatrix} \underline{X}_O \\ \underline{X}_E \end{bmatrix}$$

פירוש הדבר הוא שמוצא האי-זוגיים יהיה הפרשים של צמדים סמוכים, ומוצא הזוגיים יהיה ממוצעם.

אם נכפול את \underline{X}_E^{New} ב-2 נקבל כי זהו בדיוק השלב הראשון בהתמרת Haar.



ברור כי נוכל לבחור פעולות מורכבות יותר ל- P ו- U וכך לייצר התמרות Wavelet שונות. יש לשים לב כי אם היעד הוא התמרה אורתונורמלית, נוצר קשר בין בחירת P ו- U אותו יש לקיים, ומסתבר כי אז תכנון הצמד אינו קל. אבל, חשוב להבהיר כי לכל בחירה שרירותית של P ו- U , אף כי תוביל להתמרה שאינה יוניטרית, עדיין נוכל לטעון כי יש התמרה הפוכה פשוטה שניתנת לרישום כ:

$$\underline{X}_O^{New} = \underline{X}_O - P\underline{X}_E \Rightarrow \underline{X}_O = \underline{X}_O^{New} + P\underline{X}_E$$

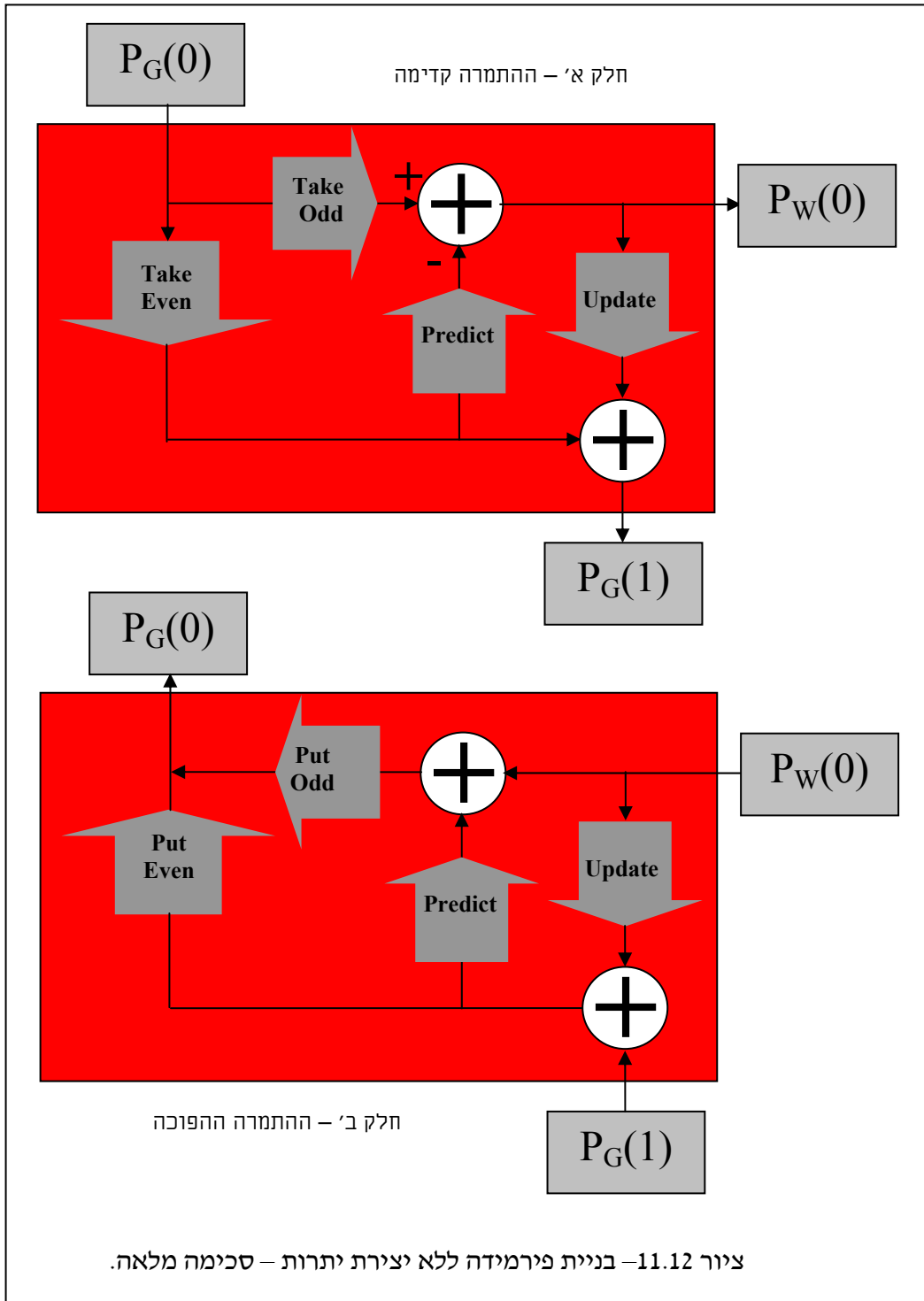
$$\underline{X}_E^{New} = \underline{X}_E + U\underline{X}_O - U P \underline{X}_E$$

$$\Rightarrow \underline{X}_E^{New} = \underline{X}_E - U P \underline{X}_E + U \underline{X}_O^{New} + U P \underline{X}_E \Rightarrow \underline{X}_E = \underline{X}_E^{New} - U \underline{X}_O^{New}$$

$$\Rightarrow \underline{X}_O = \underline{X}_O^{New} + P \underline{X}_E^{New} - U \underline{X}_O^{New}$$

או ברישום נוח יותר

$$\begin{bmatrix} \underline{X}_O \\ \underline{X}_E \end{bmatrix} = \begin{bmatrix} I - PU & P \\ -U & I \end{bmatrix} \begin{bmatrix} \underline{X}_O^{New} \\ \underline{X}_E^{New} \end{bmatrix}$$



מכיוון שפעולת ה-Update נועדה להחליק את וקטור הזוגיים הנותר, נוכל להציע מציאת מסנן

ליניארי אופטימלי להשגת יעד זה. מהמשוואה $\underline{X}_E^{New} = \underline{X}_E + U\underline{X}_O^{New}$ ניכר כי אפ פעולת ה-Update ניתנת לרישום כצירוף ליניארי

$$, \text{Update}\{\underline{X}_O^{New}\}_k = \sum_{j=-L}^L \beta_j \underline{X}_O^{New}[k-j]$$

הרי שניתן לאמץ גם רישום מטריצי בו S היא מטריצת הזזה בדגימה אחת ימינה, ולכתוב

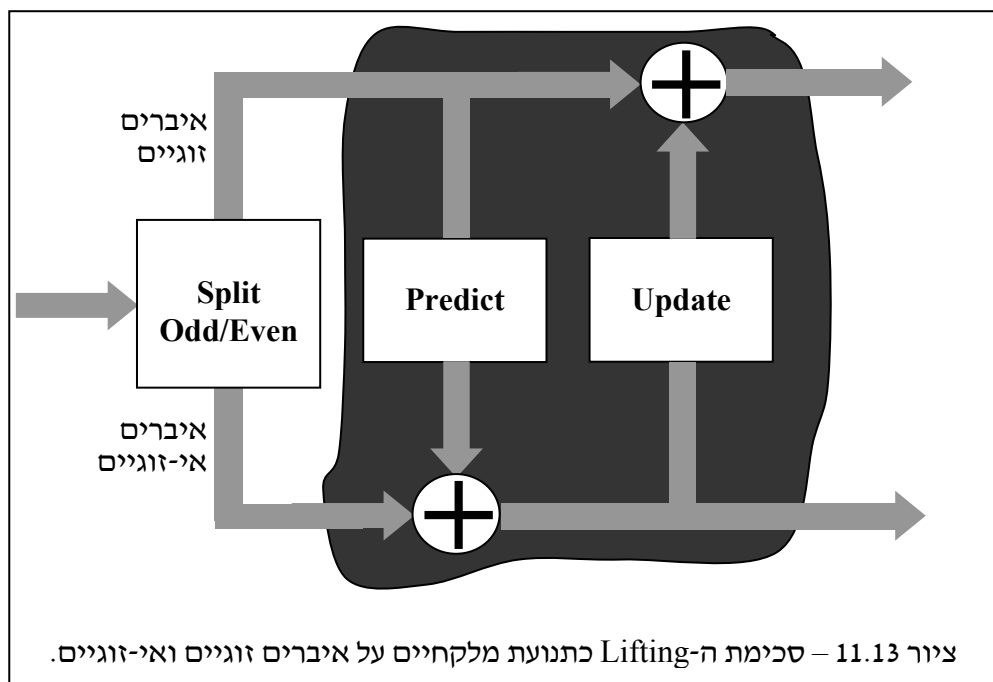
$$. U\underline{X}_O^{New} = \begin{bmatrix} \vdots & \vdots & \vdots \\ S^{-L} \underline{X}_O^{New} & \dots & \underline{X}_O^{New} & \dots & S^L \underline{X}_O^{New} \\ \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} \beta_{-L} \\ \beta_0 \\ \beta_L \end{bmatrix} = M_X \underline{\beta}$$

אזי, נוכל להציע קביעת מקדמי המסנן $\underline{\beta}$ ע"י קביעת חלקות מרבית על התוצאה של פעולת ה-Update תוך שימוש באופרטור גזירה D ולקבל

$$\text{Minimize}_{\underline{\beta}} \|D(M_X \underline{\beta} + \underline{X}_E)\|_2^2 \Rightarrow \underline{\beta}_{opt} = [D^T M_X^T M_X D]^{-1} D^T M_X^T \underline{X}_E$$

מסתבר כי סכימת ה- Lifting ניתנת לשכלול עוד ועוד ע"י הוספת פעולות Predict ו-Update, כתנועת מלקחיים כמתואר בציור 11.13. כל פעולה כזו הפיכה, ולכן קיימת ההתמרה ההפוכה. עם זאת, לא ברור כיצד יש לתכנן פעולות אלו לביצועים מיטביים. בהינתן התמרת Wavelet ניתן לגלות את פעולות ה-P וה-U אשר תבאנה לאותו יעד, ומימוש ההתמרה בפועל ע"י Lifting נחשב יעיל יותר.

כנקודה אחרונה נזכיר כי פעולות ה-Predict וה-Update לא חייבות להיות ליניאריות – דבר זה פותח פתח להתמרות חדשות ומעניינות – נושא זה עדיין בחיתוליו ויש בו פוטנציאל עצום.



פרק 12 – מבוא לתורת הצבע

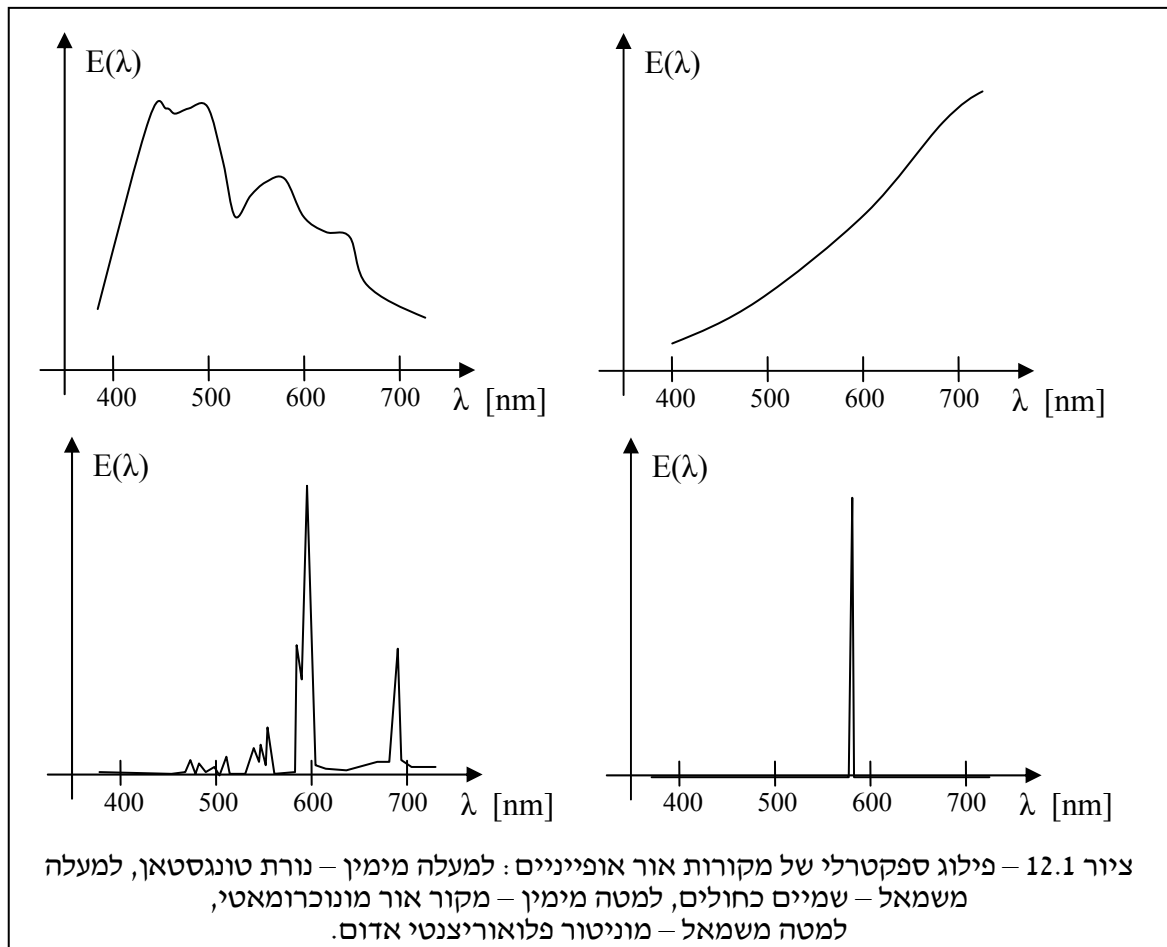
257 פרק 12 – מבוא לתורת הצבע
258 12.1 צבע - מהפיזיקה לעין האנושית
261 12.2 התאמות צבע
265 12.3 שיטות ייצוג צבע בתמונות
271 12.4 ייצוג תמונות צבע
276 12.5 שיפור ושחזור תמונות צבע
283 12.6 דחיסת תמונות צבע

רוב הדיון בספר זה מתרכז בטיפול בתמונות Gray-Scale בהן כל פיקסל כולל ערך סקלרי. מן הבחינה המתמטית, הכללת הנאמר לתמונות וקטוריות בהן כל פיקסל כולל וקטור ערכים אינה פשוטה ואינה חד-משמעית. ברור שניתן לטפל במידע תמונתי זה כאוסף בלתי תלוי של תמונות ולבצע פעולות על כל תמונה כפי שלמדנו – האם זה מספק? ברוב המקרים לא! אם כן, איך ניקח בחשבון את הקשרים בין הערכים השונים בוקטורי הפיקסלים?

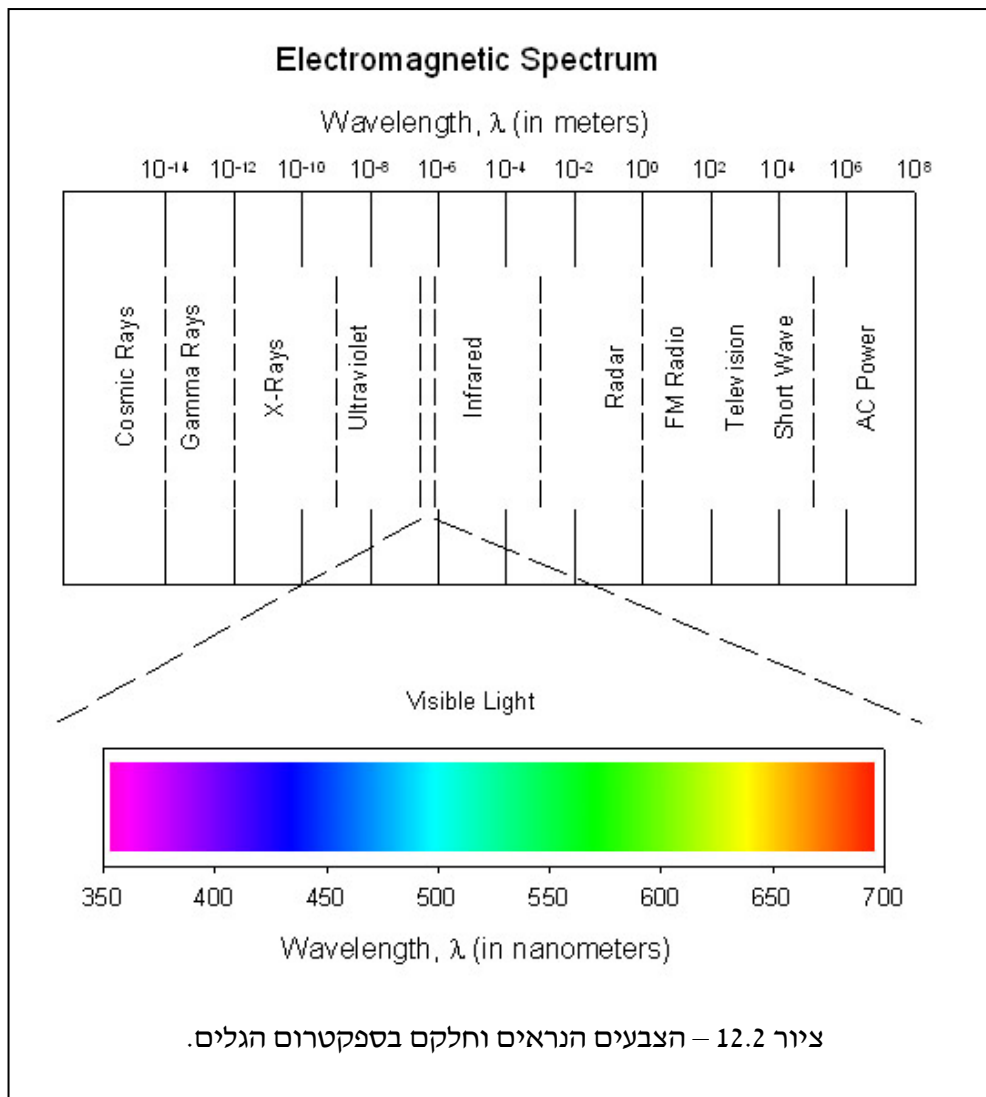
כאשר וקטור ערכי הפיקסל הינו שכבות הצבע R, G, ו-B, מעבר למורכבות שבטיפול בוקטורי ערכים, נכנס מרכיב מסבך חדש למשחק והוא מערכת הראיה האנושית. השלשה RGB אינה מקבץ תזדמן של ערכים – יש קשר ביניהם וקשר זה יש להגדיר לפי הדרך בה מערכת הראיה האנושית ממזגת שכבות אלו. ואמנם, הטיפול בצבע בתמונות מורכב, ומרבית העוסקים בעיבוד תמונה משתמשים בגישות אד-הוקיות ליישומים המופעלים על תמונות צבע. תורת הצבע היא עניין מורכב מכיוון שהיא דורשת עירוב של ידע בפיזיקה (לשם תיאור היווצרות הצבע מגלים אלקטרומגנטיים), הבנה במערכת הראיה האנושית (על מנת להבין כיצד הצבע נמדד בעין), ויכולת בעיבוד תמונות הבאה מעל כל אלה. בשל כך, שאלות רבות בעניין הטיפול בתמונות צבע עומדות עדיין פתוחות. אנו נציג בפרק זה את היסודות הברורים בתחום, ובעזרת אלו נציע הכללה לפרקי היסוד בעיבוד תמונה – ייצוג, שיפור, שחזור, ודחיסה של תמונות צבע.

12.1 צבע – מהפיזיקה לעין האנושית

היווצרה של תמונה בעין האנושית מתחיל במקור אור המאיר את היעד הנצפה. אור זה הינו גל אלקטרומגנטי שניתן לתיאור כצירוף ליניארי של אורכי גל שונים λ ובעוצמות שונות. אורך הגל מחושב כמנה בין מהירות האור ותדר הגל - $\lambda = c/f$. ניתוח של גל זה ייתן לנו את פילוגו הספקטרלי $E(\lambda)$ – מעין גרף של העוצמות בכל אורך גל. הפילוג הספקטרלי מלמד הרבה (לא הכל! ישנו עניין הפאזה שבה לא נעסוק) על אופי הגל. לשם דוגמה, בציור 12.1 מוצגים פילוגים ספקטראליים (מקורבים) של מספר מקורות אור אופייניים.



גל מונוכרומאטי הוא גל בעל אנרגיה באורך גל יחיד (כמו פונקציה של λ). אם ניצור מקור אור כזה עם אורך גל יחיד וניתן לשליטה ונציג אותו ישירות לצופה אנושי, יראה הצופה צבעים שונים כפונקציה של אורך הגל. ציור 12.2 מראה את הצבעים המתקבלים מניסוי זה. מעניין לציין שגלים בעלי אורך גל של פחות מ-350nm או מעל 750nm אינם נראים כלל. התחום $[350\text{nm}, 750\text{nm}]$ הוא התחום הנראה. בציור 12.2 אנו גם רואים מיהם הגלים שמחוץ לתחום זה, כשהשכנים הם האינפרא-אדום בגלים ארוכים והאולטרא סגול בגלים קצרים. לדוגמה, גל באורך של 480nm יראה כחול, גל באורך של 570nm יראה צהוב, וגל באורך של 610nm יראה כתום.



בשלב שני של היווצרות התמונה פוגע גל האור באובייקטים שונים, כשלכל חפץ ולכל חומר יש תכונות בליעה שונות של הגל. הבליעה תהיה פונקציה של אורך הגל, ופונקציה של הזווית בין כיוון מקור האור ובין המשטח עליו מאירים. כתוצאה מהמפגש בין גל האור ונקודה על פני האובייקט המצולם ייבלע חלק מהגל וחלק יוחזר בפילוג ספקטרלי חדש שנתון ע"י

$$E_{out}(\lambda) = E_{in}(\lambda)R(\lambda)\cos\theta$$

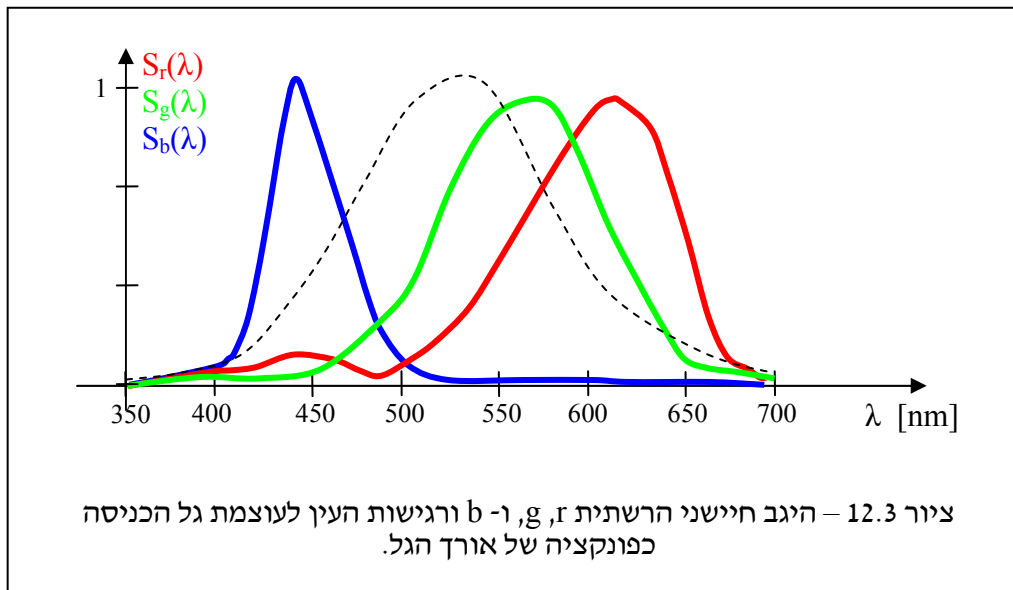
כשהפונקציה $R(\lambda)$ נובעת מתכונות הבליעה של החומר ו- θ היא הזווית בין מקור האור לפני שטח האובייקט. כך למשל, אם הזווית היא 90° , מוחזרת אנרגיה מירבית משטח האובייקט. הכנסת הגורם $\cos\theta$ נובעת מהנחה כי האובייקט בעל פני שטח למברטיים (בניגוד למראה למשל).

בשלב שלישי ואחרון בשלבי היווצרה של התמונה משתלבת מערכת הראיה האנושית - אל העין מגיע הגל $E_{out}(\lambda)$ לאחר הפגיעה בחפצים הנצפים. היה זה נפלא לו ברשתית העין היו חיישנים המבצעים ניתוח ספקטרלי מלא של הגל הנכנס ורואים אל מלוא המידע המתאר את הגל, אך בפועל אין זה כך. כבר באמצע המאה ה-19 הסיקו מדענים וביניהם Helmholtz ו-Maxwell כי חישת הצבע נעשית ע"י שלושה

חיישנים החשים צבעי יסוד שונים. נסמן חיישנים אלו ב- R, G, B . לכל אחד מחיישנים אלו עקומת היגב לאורכי הגל S_r, S_g, S_b , ומוצא חיישנים אלו נתון לפי

$$R = \int_{\lambda} S_r(\lambda)E(\lambda)d\lambda \quad G = \int_{\lambda} S_g(\lambda)E(\lambda)d\lambda \quad B = \int_{\lambda} S_b(\lambda)E(\lambda)d\lambda$$

כאשר $E(\lambda)$ הוא הגל הנופל על הרשתית. פעולה זו מבצעת שלושה ממוצעים ממושקלים של הפילוג הספקטרי, כשפונקציות ההיגב כפי שנמדדו בניסויים נראים כבציור 12.3. יש לזכור כי בהתייחסותנו לגל זה אנו חושבים על נקודה בודדה בתמונה – בפועל יש להתייחס לגל זה גם כפונקציה של המקום (x, y) , כך שבכל נקודה ברשתית פועל המודל הנ"ל עם גל $E(\lambda)$ פעל פילוג שונה.



התפיסה המוצעת לפיה שלוש אינטגרציות ממצעות יוצרות שלושה צבעי יסוד קרויה TriChromatic Theory (תיאורית שלושת הצבעים) והיא רווחת במגוון תחומים בהם נדון צבע וחישתו.

גרף נוסף שמקובל להציג בדיון על תכונות מערכת הראיה הוא גרף רגישות העין לעוצמת גל האור כפונקציה של אורך הגל. גם גרף זה מוצג בציור 12.3 כקן מקווקוו. גרף זה פירושו שאם גל מונוכרומטי בעל אנרגיה אחידה יוזן ישירות לעין, העין תבחין בו חזק יותר סביב הירוקים וחלש עד כדי אי-ראיה בקצוות התחום הנראה.

מתוך התיאור שניתן ברור כי יתכנו אינספור פילוגים ספקטריים עבור אותה שלשה של R, G, B ו- B שנמדדים ע"י החיישנים, כלומר, מידע אובד במעבר מפילוג ספקטרי מלא לשלושה ערכים מייצגים. הבחירה בשם Blue, Green, Red אינה בגלל ששלושת מרכיבים אלו הם צבעים מונוכרומטיים, אלא בגלל שהגוון הכללי אותו הם חשים (מרכז הכובד של פונקציה ההיגב) נופלת באופן כללי על צבעים אלו.

נוח להציג את האינטגרציה הנ"ל בייצוג דיסקרטי ע"י דגימה מספיק צפופה של ציר ה- λ בה יש N דגימות. וקטור העמודה \underline{E} הוא וקטור ערכי העוצמה כפונקציה של אורך הגל באורך N , והוקטורים

הם רגישויות התאים השונים בחישה ה- R, G, B . אנו מניחים כי וקטורים אלו בעלי סכום מנורמל,

$$\underline{S}_r \underline{1} = \underline{S}_g \underline{1} = \underline{S}_b \underline{1} = 1$$

נגדיר את המטריצה S כמיווג שלושת וקטורים אלה למטריצה בגודל 3 על N (מספר נקודות הדגימה על הציר λ). אזי

$$R = \underline{S}_r \underline{E}, \quad G = \underline{S}_g \underline{E}, \quad B = \underline{S}_b \underline{E} \Rightarrow \begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} - & \underline{S}_r & - \\ - & \underline{S}_g & - \\ - & \underline{S}_b & - \end{bmatrix} \underline{E} = \underline{S} \underline{E} = \underline{C}$$

מהנאמר למעלה ברור כי בהינתן \underline{C} הכולל את שלושת ערכי ה- RGB , ישנם אינסוף אפשרויות ל- \underline{E} שיניבו צבע זה, כזו מערכת של 3 משוואות ב- N נעלמים.

על רקע התיאור הנ"ל ברור מדוע טבעי לייצג תמונה ספרתית כשלושת RGB בכל פיקסל. גם מרבית המצלמות הדיגיטליות מפעילות שלושה סוגי חיישנים באופן דומה אם כי עם עקומות היגב שונות מעט. דרך השגת חיישנים שונים היא פשוטה – משתמשים בחיישנים בעלי אפיון היגב דומה, אך יוצקים על שטח פניהם מסיכת צבע המסננת חלק מאורכי הגל, וכך מעצבים את ההיגב השונה שלהם ל- R, G, B .

אין כל מניעה ליצור מצלמות בהן ההיגב לאורכי הגל שונה בתכלית מזה של מערכת הראיה האנושית. בפרט, ניתן לבנות חיישנים ש-"יראו" גלים ארוכים מ- 700nm ואלו הן מצלמות האינפרא-אדום. ישנן גם מצלמות בהן במקום שלושה סוגי חיישנים יש יותר (עד כמה מאות), כשמטרתן להתקרב לחישה הפילוג הספקטרלי במלואו, ומסתבר כי חישה זו חשובה ביישומים של חישה מרחוק (תצלום לווייני) ובצילום רפואי של רקמות ועוד. תחום זה קרוי לא במקרה "חישה מולטי-ספקטרלית".

12.2 התאמות צבע

המרות צבע נחוצות אם ישנו אי-תיאום בין חישה הצבע במערכת אחת והצגתה במערכת אחרת. נתחיל את הדיון בחישה צבע ע"י מצלמה וביצוע התאמה לשם הזנת הצבע הנמדד לעין אנושית.

נניח כי במצלמה דיגיטלית נתונים לנו שלושה סוגי חיישנים בעלי מטריצת חישה P (מקביל למטריצה S אותה הגדרנו עבור מערכת הראיה האנושית אך עם עקומות היגב שונות), ואת מוצא המצלמה אנו מזינים למסך כך שבהינתן שלושת RGB המסך מציגה בצורה נאמנה לנצפה במערכת הראיה האנושית (כמילים אחרות – בהינתן שלושת RGB פולט המסך אור בעל פילוג ספקטרלי כך שכשיוכפל ב- S ייתן שוב את אותה שלשה).

נרצה כמובן לבצע תיקון על ערכי ה- RGB הנמדדים במצלמה לפני העברתם למסך, לשם מעבר מערכי החישה שהתקבלו ב- \mathbf{P} (במצלמה) לאלו שיתקבלו ב- \mathbf{S} (בעין). עבור פילוג ספקטרלי $\underline{\mathbf{E}}$ שהוצג למצלמה, כל שיש לנו הוא סט של שלושה ערכים שייקרא $\underline{\mathbf{C}}_{\text{camera}} = \mathbf{P}\underline{\mathbf{E}}$ ורצונו בסט אחר שייקרא $\underline{\mathbf{C}}_{\text{Eye}}$. נגדיר מטריצה \mathbf{M} בגודל 3 על 3 שתמיר את האחד לאחר ונרצה לקבוע את זהותה. לו היה נמדד ישירות, הצבע בעין היה $\underline{\mathbf{C}}_{\text{Eye}} = \mathbf{S}\underline{\mathbf{E}}$. בפועל צבע זה יקורב ע"י $\underline{\mathbf{C}}_{\text{Eye}} \approx \mathbf{M}\mathbf{P}\underline{\mathbf{E}}$. לכן, נרצה מטריצה \mathbf{M} שתביא למינימום את הביטוי

$$f\{\mathbf{M}\} = \int_{\underline{\mathbf{E}} \in \Omega} \|\mathbf{M}\mathbf{P}\underline{\mathbf{E}} - \mathbf{S}\underline{\mathbf{E}}\|_2^2 d\underline{\mathbf{E}}$$

הרעיון הוא למצוא \mathbf{M} שתהיה נאמנה לכל המקבץ האפשרי של פילוג הספקטרום הכניסה $\underline{\mathbf{E}}$. גזירת ביטוי זה ליחס ל- \mathbf{M} נותנת

$$\frac{\partial f\{\mathbf{M}\}}{\partial \mathbf{M}} = \int_{\underline{\mathbf{E}} \in \Omega} (\mathbf{M}\mathbf{P} - \mathbf{S})\underline{\mathbf{E}}\underline{\mathbf{E}}^T \mathbf{P}^T d\underline{\mathbf{E}} = 0 = \mathbf{M}\mathbf{P} \left[\int_{\underline{\mathbf{E}} \in \Omega} \underline{\mathbf{E}}\underline{\mathbf{E}}^T d\underline{\mathbf{E}} \right] \mathbf{P}^T - \mathbf{S} \left[\int_{\underline{\mathbf{E}} \in \Omega} \underline{\mathbf{E}}\underline{\mathbf{E}}^T d\underline{\mathbf{E}} \right] \mathbf{P}^T$$

נגדיר את האינטגרל בביטוי הנ"ל כ- $\mathbf{R} = \int_{\underline{\mathbf{E}} \in \Omega} \underline{\mathbf{E}}\underline{\mathbf{E}}^T d\underline{\mathbf{E}}$ ונקבל

$$\mathbf{M}_{\text{Opt}} = \mathbf{S}\mathbf{R}\mathbf{P}^T \left[\mathbf{P}\mathbf{R}\mathbf{P}^T \right]^{-1}$$

המטריצה \mathbf{R} היא מטריצת האוטוקורלציה של $\underline{\mathbf{E}}$. עבור פילוג ערכים אקראי בתחום $[0,1]$ בהסתברות אחידה, וללא קורלציה בין ערכים סמוכים (כמו ברעש לבן!), מטריצה זו תהיה

$$\int_{\underline{\mathbf{E}} \in \Omega} \underline{\mathbf{E}}\underline{\mathbf{E}}^T d\underline{\mathbf{E}} = \mathbf{R} = \frac{1}{12}\mathbf{I} + \frac{1}{4}\mathbf{1}\mathbf{1}^T$$

דרך החישוב של מטריצה זו פשוט – כל איברי האלכסון הראשי זהים ושווים ל-

$$\int_0^1 x^2 dx = \frac{x^3}{3} \Big|_0^1 = \frac{1}{3}$$

כל האיברים מחוץ לאלכסון הראשי גם הם זהים ושווים ל-

$$\int_0^1 \int_0^1 xy dx dy = \int_0^1 x dx \cdot \int_0^1 y dy = \frac{x^2}{2} \Big|_0^1 \cdot \frac{y^2}{2} \Big|_0^1 = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

ניתן גם לחשב את \mathbf{R} בהנחת פילוג אחרת על שדה ההארה $\underline{\mathbf{E}}$, ולקחת בחשבון למשל משקל שונה לפילוגים בהתאם לתמונות בהן אנו עוסקים. ניתן גם לבצע את בניית \mathbf{R} מתוך סט דיסקרטי של הארות צפויות.

בביטוי של \mathbf{M} האופטימלי אנו הופכים מטריצה בגודל 3 על 3 אשר תהיה לא סינגולרית אם $\text{rank}\{\mathbf{P}\} = 3$ ואם \mathbf{R} מטריצה חיובית מוגדרת. שתי דרישות אלו מתקיימות ($\text{rank}\{\mathbf{P}\} = 3$) משמעו ששלושת סוגי החיישנים במצלמה פורסים מרחב תלת מימדי – אם לא – זהו בזבוז של חיישנים). כמקרה פרטי מעניין, אם $\mathbf{S} = \mathbf{P}$ – כלומר לא נדרשת המרה, וזה כמובן צפוי.

בעיית התאמה אחרת היא בעיית התאמת צבע כך שפילוג ספקטרלי נתון יוצג במסך בעל מאפיינים מסויימים כך שהעין האנושית תראה את הגוונים הנכונים. עניין זה דומה מעט בניתוחו והוא מוביל בניית עקומות ה-CIE שנוכיר בהמשך.

נניח כי בידנו התקן פלט (למשל, מסך מחשב טלויזיה) המציג תמונה. דרך הצגת התמונה היא ע"י בחירת מספר (אולי שלוש ואולי יותר) צבעי יסוד, בניית צירוף ליניארי שלהם והצגתם. ברור כי התקן פלט זה אינו יכול להציג כל פילוג ספקטרלי אפשרי של אור בשל השימוש במעט צבעי יסוד. עם זאת, אין כל חשיבות להפקת הפילוג הספקטרלי המדויק – די בהפקת פילוג ספקטרלי אשר כשיוצג לעין אנושית הוא ייתן אותה שלשת RGB, ובכך וידאנו כי הפקת הצבעים מהתקן הפלט תואמת את ציפיותיה של מערכת הראיה האנושית.

ב-1930 קבעה ועדה בשם Commission Internationale de L'Éclairage (CIE) סטנדרטיזציה של הפקת צבע בהתקני פלט לפי האופן הבא. נבחרו שלושה צבעים מונוכרומטיים כצבעי יסוד מהן מורכב כל צבע בצירוף ליניארי. ועדת ה-CIE קבעה צבעי יסוד אלה לאורכי הגל 700nm (אדום), 546.1nm (ירוק) ו-435.8nm (כחול). נוכל לייצג סידרת פילוגים ספקטראליים אלו כמטריצה \mathbf{P} בגודל 3 על N בה בכל שורה יש איבר אחד השווה 1 והוא המיקום אורך הגל הנאות. לכן, בדומה להנחה שנעשתה קודם, סכום איברי כל שורה מנורמל להיות 1.

בהינתן פילוג צבע אמיתי כלשהו \underline{E} (וקטור באורך N), נציע צירוף ליניארי של שלושת צבעי ה- \mathbf{P} לשם הפקת הצבע בעין - צירוף ליניארי זה יהיה $\mathbf{P}^T \underline{\beta}$ - וזהו פילוג ספקטרלי חדש. בעוד שישירות העין תחוש את שלשת ה-RGB הנתונה ע"י \underline{SE} , הרי שע"י הצבעים היסודיים נקבל כי שלשת ה-RGB שהעין תחוש היא $\mathbf{SP}^T \underline{\beta}$. נרצה כמובן ששתי שלשות ה-RGB תלכדנה (או תהיינה קרובות) לכל צבע \underline{E} אפשרי ע"י בחירת הווקטור $\underline{\beta}$ בצורה נאותה כפונקציה של \underline{E} .

הדרך בה בחרה ועדת ה-CIE לתקוף את הבעיה היא הבאה: במקום לקבוע פונקציה כללית מהצורה $\underline{\beta}(\underline{E})$, נניח כי את קביעת ה- $\underline{\beta}$ נעשה לצבעים מונוכרומטיים בלבד. לצבע מונוכרומטי \underline{E}_k (דהיינו, '1' במקום ה-k בווקטור ואפס בכל מקום אחר) נקבל את המשוואה

$$\mathbf{SP}^T \underline{\beta}_k = \underline{SE}_k \Rightarrow \underline{\beta}_k = [\mathbf{SP}^T]^{-1} \underline{SE}_k$$

וממשוואה זו נחלץ את שלושת המשקלות $\underline{\beta}_k$. בסריקת ה-k מ-1 ועד N נקבל שלוש פונקציות משקל הקרויות Tristimulus Values. פונקציות אלו מייצגות את המשקל היחסי של כל צבע יסוד בבניית צבע יעד מונוכרומטי בעין. ציור 12.4 מראה את הפונקציות הללו, וכפי שניכר, בצבעים מסויימים נדרש משקל שלילי לצבע היסוד האדום לקבל האפקט הרצוי.

ברישום אחר, ניתן לצרף את כל המשוואות למשוואה מטריצית אחת בה הצבעים \underline{E} ימלאו עמודותיה של מטריצה \underline{E} בגודל N על N וזו כמובן תהיה מטריצת היחידה בשל היות כל צבעים אלו מונוכרומטיים עם סריקה מלאה של אורך הגל. אז נקבל כי פונקציות המשקל הנ"ל אינן אלא

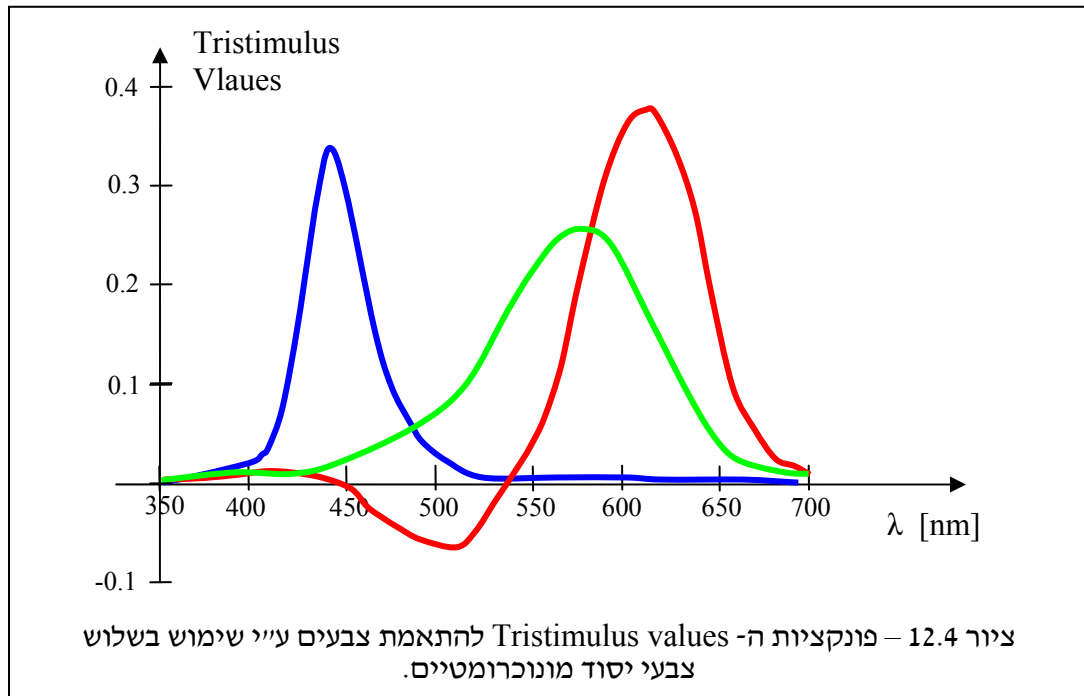
$$\left\{ \begin{array}{l} \mathbf{SP}^T \underline{\beta}_1 = \mathbf{SE}_1 \\ \vdots \\ \mathbf{SP}^T \underline{\beta}_k = \mathbf{SE}_k \\ \vdots \\ \mathbf{SP}^T \underline{\beta}_N = \mathbf{SE}_N \end{array} \right\} \Rightarrow \mathbf{SP}^T \mathbf{B} = \mathbf{SE} \Rightarrow \mathbf{B} = [\mathbf{SP}^T]^{-1} \mathbf{SE} = [\mathbf{SP}^T]^{-1} \mathbf{S}$$

בהגיע צבע כללי \underline{E} , הוצע ע"י ה-CIE להתייחס אליו כצירוף ליניארי של אותות מונוכרומטיים, לכל צבע מונוכרומטי להשתמש במשקלות הנאותים, וכך לקבל את סט המשקלות הכללי. בתיאור אלגברי זה אינו אלא לומר כי

$$\underline{E}_{New} = \sum_{k=1}^N w_k \underline{E}_k$$

$$\Rightarrow \underline{\beta}_{New} = \sum_{k=1}^N w_k \underline{\beta}_k = \sum_{k=1}^N w_k [\mathbf{SP}^T]^{-1} \mathbf{SE}_k = [\mathbf{SP}^T]^{-1} \mathbf{S} \sum_{k=1}^N w_k \underline{E}_k = [\mathbf{SP}^T]^{-1} \mathbf{SE}_{New} = \mathbf{BE}_{New}$$

כלומר, בפשטות, בהינתן צבע חדש יחושבו מקדמי הצירוף הליניארי כממוצע משוקלל של משקלות ה-Tristimulus Values.



האם זוהי מטריצת ה- \mathbf{B} הטובה ביותר? נציע גישה כללית יותר למציאת \mathbf{B} המיטבית באופן הבא – נניח כי המטריצה \mathbf{B} נקבעה ולכל צבע המגיע נשתמש בשלשת המשקלות $\underline{\beta}_{New} = \mathbf{BE}_{New}$. בהגיע הצבע \underline{E}_{New} , העין תחוש ישירות את השלשה \mathbf{SE}_{New} , בעוד שבגישת שילוב צבעי היסוד המוצעת נקבל

את השלשה $\underline{\beta}_{New} = \underline{SP}^T \underline{BE}_{New}$. מטבע הדברים, נרצה כי שתי שלשות RGB אלה יתלכדו לכל צבע. לכן נדרוש

$$\forall \underline{C}, \underline{SP}^T \underline{BE} = \underline{SE} \Rightarrow \underline{SP}^T \underline{B} = \underline{S} \Rightarrow \underline{B} = [\underline{SP}^T]^{-1} \underline{S}$$

ותוצאה זו זהה לתוצאה שקיבלנו קודם.

אם נחזור לרגע להתחלה, משמעות תוצאה זו היא שבהינתן מסך מחשב המייצג צבע ע"י מיוזג של שלושה צבעי יסוד ע"י צירוף ליניארי, ניתן "לרמות" את העין ולספק במסך פילוג ספקטרלי אשר יתאים לצבע אותה העין הייתה רואה לו צפתה בגל האור המקורי. אבל, תוצאה זו בעלת ערך תיאורטי כיוון שמשקל שלילי יוצר בעיה במימוש. שאלה מיידית שעולה מכך היא השאלה הבאה – האם קיים סט של צבעי יסוד פיזיקליים (כלומר בעלי ספקטרום אי-שלילי) שיתנו כי \underline{B} אי-שלילית אף היא? ננסח זאת כאוסף הדרישות

$$\forall i, j, \underline{B}[i, j] \geq 0, \quad \underline{P}[i, j] \geq 0, \quad \underline{B} = [\underline{SP}^T]^{-1} \underline{S}$$

מסתבר כי על מנת לתת תשובה חיובית לשאלה זו נדרשים מספר גדול משלוש של צבעי יסוד (יש לשים לב שאז נוסחת החישוב של \underline{B} אינה כמתואר כי המטריצה \underline{SP}^T אינה ריבועית, וישנן דרגות חופש שמביאות לקיומו של פתרון). לעומת זאת, אם נתיר שימוש בצבעי יסוד לא פיזיקליים (כאלה שכוללים בספקטרום שלהם ערכי אנרגיה שליליים), הדבר אפשרי, וייצוג זה מוביל למפת ה- CIE-XYZ שתוזכר בהמשך. ברור כי שימוש בפתרון מעין זה לא יינתן לשלם בניה בפועל של צבע (שהרי צבעי היסוד אינם ברי מימוש), אך חשיבות מפה זו גדולה לשם הצגת פריסת הצבעים האפשרית.

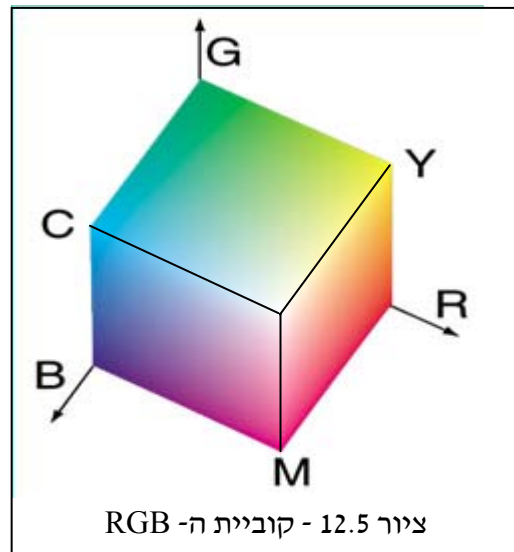
12.3 שיטות ייצוג צבע בתמונות

כאמור, שיטת ייצוג צבע טבעית לתמונות כשהן במדיה דיגיטלית היא שיטת ה-RGB. אם נניח כי כל שכבת צבע זוכה לייצוג ב-8 סיביות, ניתן לחשוב על שלשה זו כמגדירה קובייה של צבעים, וכל צירוף של 24 סיביות יוצר אחד מתוך כ-16 מיליון צבעים אפשריים בתוך חלל הקובייה הזו. קוביית ה-RGB מוצגת בצירור 11.5. נשים לב כי הצבע $[R=255, G=255, B=0]$ נותן את הצבע צהוב המתואר כפינה אחת של הקובייה. פינות מעניינות אחרות הן $[R=255, G=0, B=255]$ - Magenta, ו- $[R=0, G=255, B=255]$ שנותן את ה- Cyan (תכלכל). השילוב $[R=255, G=255, B=255]$ נותן לבן. צבעי שילוב אלו חשובים בהקשר של החסרות צבעים אשר יידון בהמשך בקצרה.

הערה: יש מקום להיות יותר מדויקים, ולצייר את הקובייה הנ"ל כשהיא מעוותת. הרושם שנובע מהצירור הנ"ל הוא שצבעי היסוד R, G, ו-B אורתונורמליים זה לזה – ובכך חבויה הנחה על חוסר תלות ביניהם, ואין זה כלל כך. אם נניח למשל ששלשה זו היא כאלו המתקבלים במערכת הראיה האנושית, הרי שמצירור 12.3 ברור לנו כי מכפלה פנימית בין פונקציות המשקל שתוארו תיתן קורלציה כלשהי בין הצירים שיצרנו. דרך הצירור הנכונה יותר תביא לפוליטופ (מצולע תלת-מימדי) שהינו עיוות של הקובייה כך שהזווית בין הצבעים תהיה חדה יותר מ-90°. פוליטופ זה יהיה בצורת פירמידה (או יותר נכון - קונוס)

שראשה בראשית, כיוון שאם שלשת RGB קיימת ונראית, הרי שכל הכפלה שלה בקבוע ייתן צבע נראה גם כן.

ניתן גם להסתכל על קובייה זו בדרך שונה, כפי שמציע ציור 11.5, ולמתוח קו דמיוני בין נקודת הראשית $[0,0,0]$ ובין $[255,255,255]$. לכל אורך ישר זה נקבל צבע אפור בבהירות משתנית משחור מוחלט ללבן מוחלט. הצבע הלבן בגל מתייחס לפילוג קרוב לאחיד של כל אורכי הגל (להזכירכם, כשאנו משתמשים במושג "לבן" באפיון רעש למשל, אנו מתכוונים כי פילוגו בתדר אחיד), ועבור פילוג זה $R=G=B$ (כלומר, השטח מתחת הגרפים בציור 12.3 שווה בשלושת העקומים, כפי שכבר הנחנו).



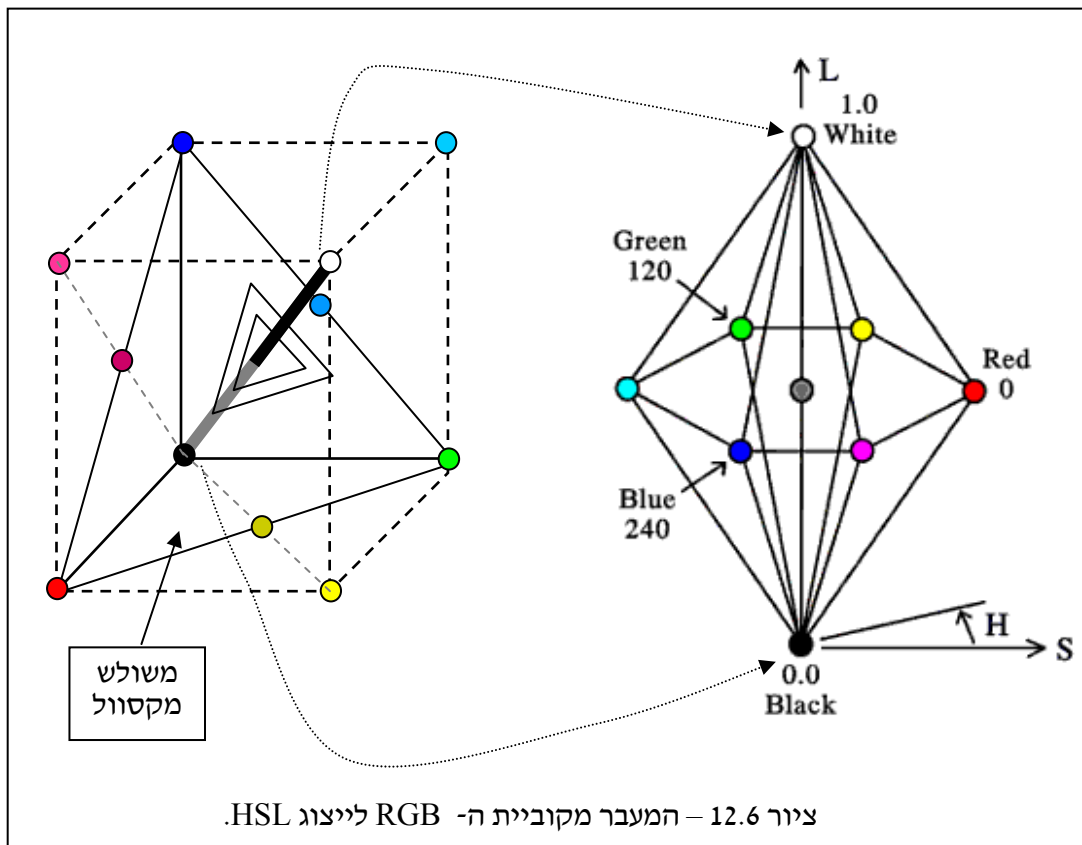
משתמע מכך כי אם נרצה להמיר תמונת צבע לתמונת גווני אפור, נחליף כל שלשת ערכים $[R,G,B]$ בנקודה הקרובה ביותר על האלכסון שחוצה את הקובייה משחור ללבן, וברור כי זהו פשוט מיצוע $(R+G+B)/3$. אבל נוכל לעשות יותר - במקום שימוש ב- $[R,G,B]$ ניתן לייצג את הצבע ע"י שלשת ערכים חדשה שתיתן הפרדה טבעית בין בהירות לתכונות הצבע. כפי שהוצע קודם, המיקום לאורך ציר הבהירות (נקרא Lightness או Brightness או Intensity או Luminance) יהיה הערך הראשון בשלשה החדשה. בקביעת הבהירות התמקמנו על החתך המתואר כמשולש אשר הינו אנך לקו הציר שחור-לבן - משולש שנקרא משולש מקסוול.

על פני משולש מקסוול, מיקום ייצג את הצבע. דרך אחת לעשות זאת היא קביעת המרחק בערך מוחלט מקו הציר (נקודת המרכז של המשולש) כעדות לעומקו של הצבע (נקרא saturation). בנוסף, הכיוון אליו פנינו - או במילים אחרות - אורך הגל הדומיננטי - יעיד על הגוון (קרוי Hue). כך נקבל ייצוג חדש לצבע הקרוי HSL (או HSB או HSI) והוא מתואר בציור 12.6. ישנן מספר דרכים לבצע את החישוב האנליטי העובר משלשת RGB ל-HSL, אך אנו נוותר על התענוג שבהצגת המשוואות שמיישמות מעבר זה בפועל.

הרעיון להפריד בין בהירות לפירוטי הצבע מפתה וחוזר על עצמו במגוון ייצוגים אלטרנטיביים כגון YIQ, CIE-XYZ, La*b*. בחירה פופולרית מבין ייצוגי הצבע קרויה YIQ. ייצוג זה הוצע ע"י ועדה שעסקה בייצוג צבע לשידורי טלוויזיה (National Television Systems Committee - NTSC) וייצוג זה הוא המיושם באותות טלוויזיוניים. מתוך רצון להציע המרה פשוטה מ- RGB (צבעי היסוד של CIE ולא אלו של מערכת הראיה) ל- YIQ הוצעה ההמרה הליניארית הבאה:

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & -0.322 \\ 0.211 & -0.523 & 0.312 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

בייצוג זה, מרכיב ה- Y הוא המקביל לבהירות והוא מוצע כממוצע משוקלל של R, G, ו- B עם דגש על הירוק – זהו האות לפיו תיראה התמונה בטלוויזיה שחור לבן. שני המרכיבים האחרים מאפיינים צבע כשהבחנה אינה כמקודם – גוון ועומקו, אלא שני צירים כמעט אנכיים זה לזה שפורסים את התנועה על פני שטח משולש מקסוול שווי הבהירות שהגדרנו בציור 12.6. מטריצה זו הפיכה ולכן ההתמרה חזרה פשוטה אף היא. הערכים במטריצה באים לשקף מחד את שלושת צבעי היסוד שנתונים במסך הטלוויזיה, ומאידיך את הרצון לסובב את צירי קוביית ה- RGB בזווית מסוימת לקבלת הפרדה טובה בין צבע ובין בהירות.

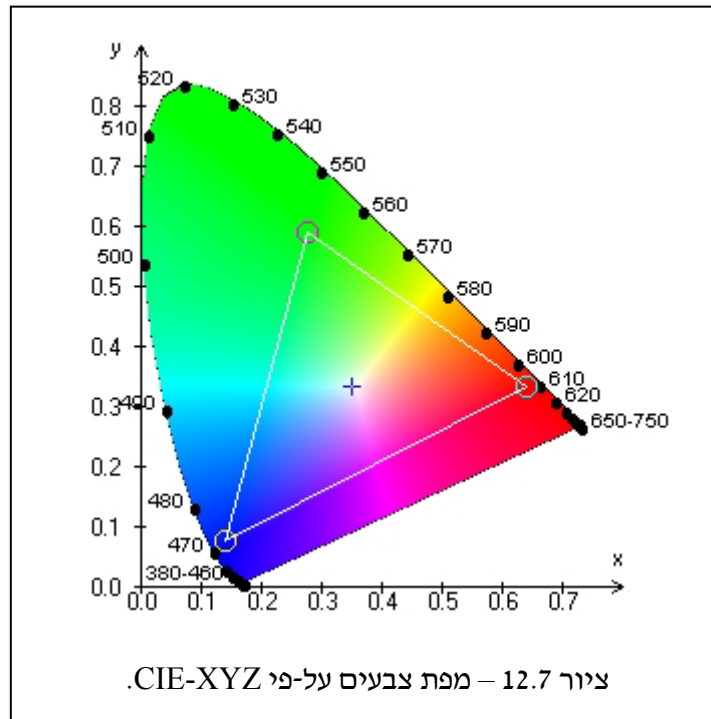


ייצוג אחר שהוזכר הוא ה- CIE-XYZ והוא כבר הוזכר קודם. הרעיון הוא הבא: אם אנו מוותרים על העניין הבהירות, הרי שנרצה להתרכז בנקודות על פני המשולש לו קראנו משולש מקסוול.

משולש זה מאופיין במשוואה $R+G+B=Const$. בהינתן שלשת RGB, די בשני ערכים על מנת לתאר את המיקום של הצבע,

$$.X = \frac{R}{R+G+B}; Y = \frac{G}{R+G+B}; Z = \frac{B}{R+G+B} \Rightarrow X+Y+Z=1$$

כנגד כל צמד ערכי $Z, (Y, X)$, ידוע ומיותר, ולכל צמד כזה במישור נוכל ליחס צבע על כל מאפייניו (גוון בסיסי ועומק הצבע). נוכל לייצר מפת צבעים באופן זה, ומפה מעין זו מתוארת בציור 12.7. חשוב לציין כי בניית מפה זו תלויה בבחירת צבעי היסוד R, G, ו-B שמהם התחלנו, ובבנייתה מעורבת המטריצה B שהצגנו קודם על מנת לעבור לחישה צבע כפי שהעין תיתן. כפי שכבר ראינו, שימוש בצבעי היסוד של CIE (המונוכרומטיים) לשם השגת צבעים מסויימים מוביל למשקל שלילי. המפה שמתוארת כאן מציעה שילובים של ערכי X ו-Y חיוביים לשם קבלת כל צבע בספקטרום הצבעים, ועובדה זו נשענת על שימוש בצבעי יסוד (B,G,R) שאינם פיזיקליים, כפי שהבהרנו קודם.

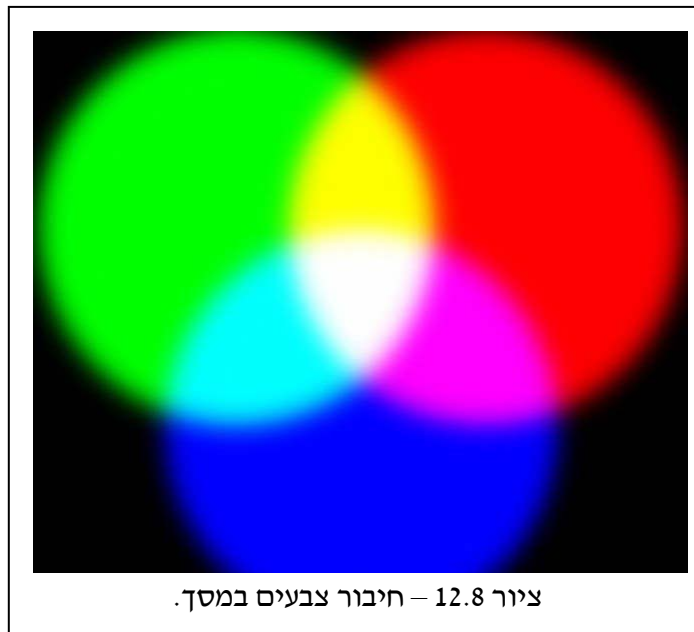


בעיה שמתעוררת עם מפה זו היא שמרחקים שווים על-פניה אינם משקפים מרחקי צבע שווים ביחס למערכת הראיה האנושית. דרך אחת להציג סוגיה זו היא הצגת אליפסות שוות "חישה" סביב נקודות הפזורות על פני מפה זו. מסתבר כי בחלק הימני עיון של המפה נדרש מרחק גדול יותר על פני המפה לשם קבלת שינוי שווה בעין האנושית. עניין זה חשוב למדי במערכות בהן מרחקי צבעים משמשים (וכמעט כל נושא שיוצג בהמשך נופל לקטגוריה זו). ייצוג La^*b הוצע כעיוות המפה הנ"ל כך שמרחקים אויקלידים פשוטים על-פניה יהיו מרחקים הוגנים.

כנושא אחרון במסגרת הדיון בייצוגי צבע, ניגע בנושא חיבור וחיסור של צבעים. מבלי לומר זאת במפורש הנחנו עד כה כי צביעת נקודה אחת בשני צבעים מהווה את חיבורם בייצוג RGB, כלומר חיבור מרכיביהם באופן אלגברי פשוט. כמו כן, כשמיזגנו שני צבעי יסוד התקבל צבע ביניים כקומבינציה ליניארית פשוטה. תפיסה זו מתאימה למסך בו נקודת המוצא היא צבע שחור, והכנסת אור מקרבת את

גווינו ללבן. בגישה זו, חיבור צבעים כמתואר בציור 12.8 טבעי ונכון. לשם דוגמה, אנו רואים כי כאשר מצרפים את הירוק לאדום מתקבל צהוב, ועם הוספת הכחול יתקבל לבן.

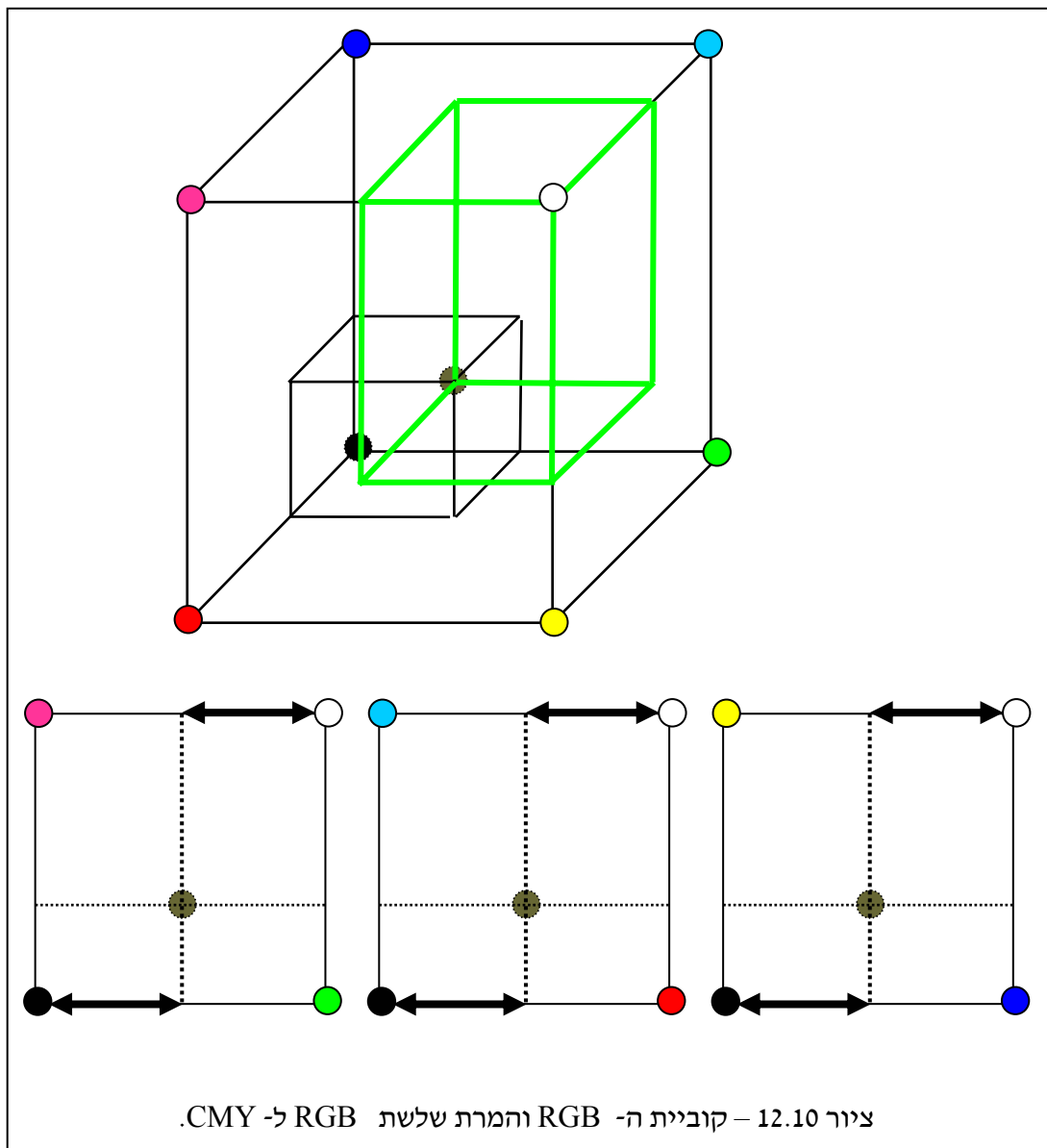
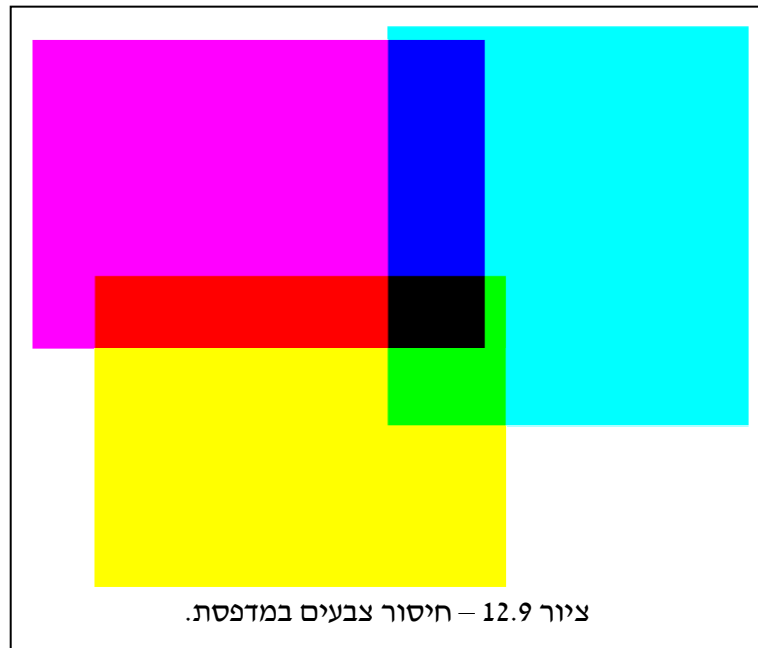
לעומת זאת, כשאנו מדפיסים על דף נייר, נקודת המוצא הפוכה – זהו הצבע הלבן – וכל הוספת צבע מקרבת אותנו לשחור. אם כך, בהינתן שלשת RGB אותה רצונו להפיק על גבי נייר, באיזה צבעי יסוד נשתמש וכיצד נמזגם? מוצע להשתמש בצבעים המשלימים אותם כבר פגשנו בתיאור קוביית ה-RGB בציור 12.4. הצבעים המשלימים הם Cyan, Magenta, ו-Yellow, ומכאן השם CMY. השם 'משלימים' נובע מהעובדה שהוספת צבע משלים ללבן יוצר בפועל את הצבע המקורי – למשל, בהינתן דף לבן ועליו מודפס גוון ה-Magenta, אם נצבע מעליו בצבע שני Cyan נקבל צבע כחול (ראה ציור 12.9). פירוש הדבר הוא שבעבודה עם קוביית ה-RGB עלינו להתייחס לראשית כצבע הלבן וממנו אנו פועלים לבניית השחור, בעוד שבגישה המקורית ראינו את השחור כראשית וממנו בשילובי צבעים התקרבו ללבן.



נניח כי נתונה לנו השלשה $[R,G,B]$ המיוצגים ב-8 סיביות, ורצונו להדפיסה על דף נייר לבן. לו היה זה מוניטור, היינו מצרפים את הגוונים R באדום, G בירוק ו-B בכחול שבמסך והגוון הרצוי היה מתקבל. בהדפסה הגישה תהיה שונה. אם נצייר את נקודת היעד בתוך קוביית ה-RGB, עלינו למזג את הצבעים המשלימים CMY לפי

$$C = 255 - B, \quad M = 255 - G, \quad Y = 255 - R$$

ותהליך זה מוסבר בציור 12.10. לדוגמה, כשאנו נעים מנקודת הלבן לעבר נקודת ה-Magenta לשם קביעת הכמות הנחוצה של צבע זה, אנו נעים בפועל בקו מקביל לקו הנע מהשחור לירוק. לכן מידת ה-Magenta הנחוצה היא מלוא הלבן פחות הירוק שבצבע הנתון.

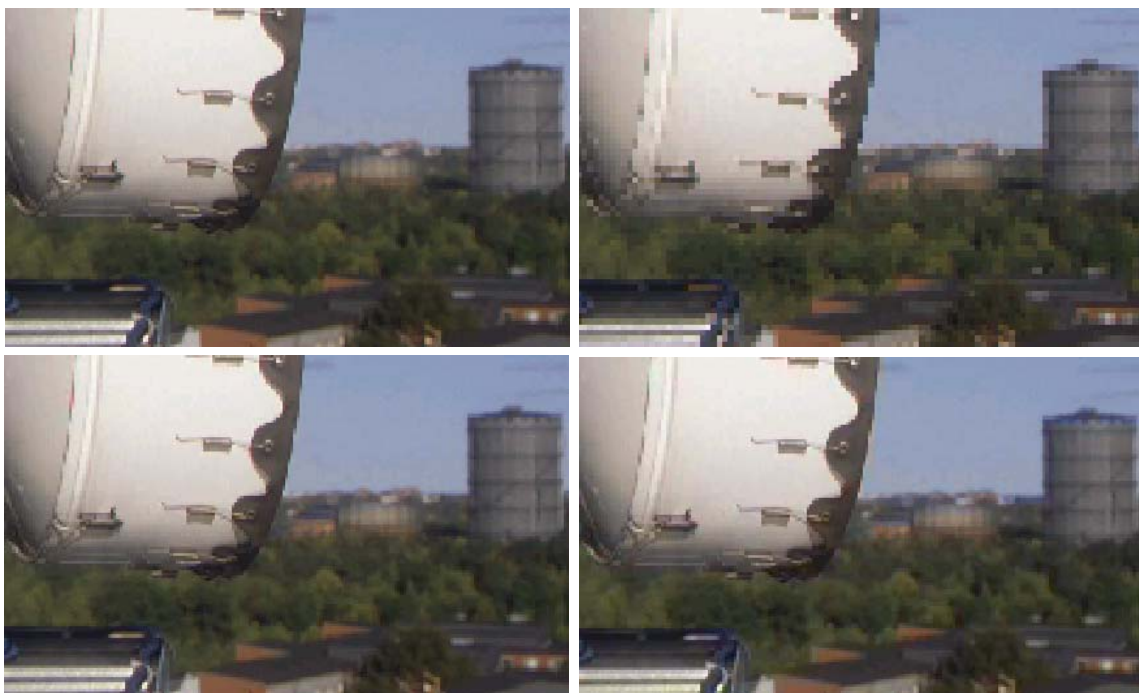


12.4 ייצוג תמונות צבע

בדיון בייצוג תמונות עסקנו בשני נושאים מרכזיים – דגימה וקוונטיזציה. נחזור לשני נושאים אלו בקצרה ונראה מה השוני הנובע מהטיפול בצבע.

דגימה של תמונת צבע יכולה לקבל ניתוח זהה לניתוח שנתנו לתמונת שחור-לבן עם ההבחנה כי יש לבצע את תהליך הדגימה לכל סקלר בייצוג הוקטורי של הצבע. אם התמונה מיוצגת ב-RGB, לרוב נגלה כי מרווח הדגימה הנחוץ קרוב מאוד ואף זהה בשלושת שכבות התמונה.

לעומת זאת, אם הייצוג יכלול הפרדה של בהירות וגוון נקבל לרוב כי מרווח הדגימה בבהירות נדרש להיות קטן יותר, בעוד שבשכבות הגוון המרווח יכול להיות גס יותר. תופעה זו נובעת מתכונות מערכת הראייה האנושית בה אין רזולוציה טובה לצבע בעוד שנדרשת רזולוציה גבוהה לבהירות. ציור 12.11 ממחיש תופעה זו ע"י הצגת תמונה (כל ההצגות הינן בסופו של דבר ב-RGB) שיוצגה ב-YIQ, כשבאחת התמונות כל השכבות נתונות באותה רמת פירוט, באחת התמונות ישנו דילול ואינטרפולציה ב-Y (פעולת הדילול והאינטרפולציה יוצרות בלוק של 2 על 2 או 4 על 4 פיקסלים בעלי אותו ערך) עם השאר I ו-Q, ללא קלקול, ותמונה שלישית נתונה כששכבת ה-Y שמורה כבמקור בעוד שכבות ה-I וה-Q עברו דילול ואינטרפולציה.



ציור 12.11 – השפעת דילול ואינטרפולציה על חדות תמונת צבע. למעלה משמאל – תמונת המקור.

- למעלה מימין – תמונת המקור עם דילול ואינטרפולציה ביחס 1:2 בכל ציר לשכבת ה-Y.
- למטה משמאל – תמונת המקור עם דילול ואינטרפולציה ביחס 1:2 בכל ציר לשכבות ה-I וה-Q.
- למטה מימין – תמונת המקור עם דילול ואינטרפולציה ביחס 1:4 בכל ציר לשכבות ה-I וה-Q.

אנו רואים כי פגיעה (ואפילו קלה) בשכבת הבהירות ניכרת מיד. לעומת זאת, אבדן $\frac{3}{4}$ משכבות ה-I וה-Q נותנת אבדן זעום שקשה לאבחנה (בסך-הכל חצי ממידע התמונה הושלך כמעט ללא נזק). כאשר הדילול בשכבות הכרומה הינו 1:4 בכל ציר (דהיינו, נותר $\frac{1}{16}$ מכלל המידע בהן) הנזק מתחיל להיות מורגש, אך עדיין לא באותה עוצמה בה נראה הנזק שהסבנו ל-Y עם דילול פחות אגרסיבי. בדיוק תכונה זו היא שעמדה לנגד עיניהם של חברי ועדת ה-NTSC כשבחרו בייצוג YIQ. מטרתם העיקרית היתה חסכון ברוחב פס לשידור, ועם המעבר ל-YIQ השידור כולל פס רגיל לשכבת ה-Y ופסי תדר צרים הרבה יותר לשתי שכבות הכרומה. אנו נראה כי תכונה זו תשוב ותשמש אותנו בבואנו לבצע שחזור או דחיסה של תמונות צבע.

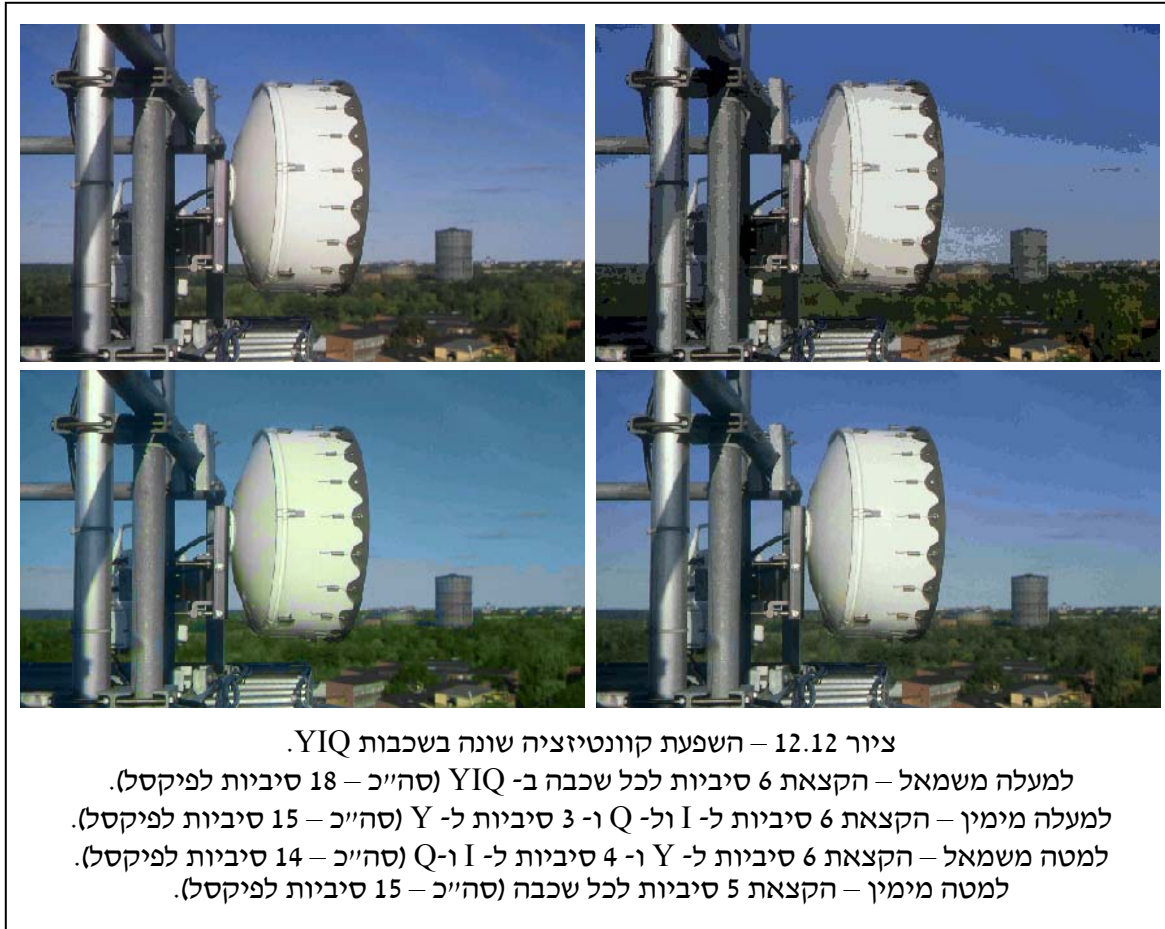
קוונטיזציה של תמונת צבע יכולה להיעשות ע"י קוונטיזציה של כל אחת מהשכבות הסקלריות בנפרד. בפרט, כבר הזכרנו כי בייצוג RGB מקובל לתת 8 סיביות לכל שכבה – דבר המוביל להיווצרן של 16 מיליון צבעים אפשריים בייצוג. אם נתבונן בקוביית ה-RGB נראה קוונטיזציה זו כקוביות שוות נפח אשר גבולותיהן מקבילות לצירים כשריג קרטזי אחיד. בגישה ספרבילית זו, הורדת כמות הסיביות באחת השכבות תביא לרוב לפגיעה מורגשת, וכך אנו תקועים על 24 סיביות לפיקסל.

האם באמת לא נוכל לרדת מ-16 מיליון צבעים? התשובה היא כנראה לא אם הגישה ממשיכה להיות ספרבילית ב-RGB ואם נדרוש התאמה לכלל כל התמונות. לעומת זאת, אם נרשה עירוב הצירים או בניית קוונטיזציה תלוית תמונה, נוכל בהחלט לחסוך.

עירוב הצירים משמעו שתאי הקוונטיזציה בקוביית ה-RGB יהיו בעלי צורה כללית יותר, ואת זאת ניתן להשיג למשל ע"י עיוות הקובייה ע"י מעבר לייצוג אחר (כגון YIQ) וביצוע של קוונטיזציה אחידה על הצירים החדשים עם בחירת פחות סיביות לאזורים פחות רגישים. למשל, בייצוג YIQ נידרש ל-8 סיביות ל-Y, אך ברוח הדיון הקודם נוכל לפגוע בשכבות ה-I וה-Q בייצוג ב-6 סיביות בלבד. ציור 12.12 מראה תוצאות ניסוי מעין זה בו רואים כי אם יש להפחית מכמות הסיביות, הרבה פחות נזק נגרם אם הפחתה זו נעשית על שכבות הצבע.

הערה: בקוונטיזציה של שכבות ה-YIQ בניסוי הנ"ל נעשה שימוש בקוונטיזר יוניפורמי ממורכז. לשם יישום קוונטיזר כזה יש לקבוע את ערך המינימום והמקסימום. בהנחה ששכבות ה-RGB נתונות בתחום $[0,1]$, לשכבת ה-Y הטווח יהיה $[0,1]$, לשכבת ה-I הטווח יהיה $[-0.6,0.6]$ ולשכבת ה-Q הטווח יהיה $[-0.53,0.53]$. ערכי מינימום ומקסימום אלה מתקבלים בפשטות מתוך מטריצת ההמרה.

למעשה, ניתן ללכת הרבה יותר רחוק אם מוכנים להכניס את המרכיב השני שהוזכר למעלה, הוא עניין התלות בתמונה. בגישה הקודמת מוצעת קוונטיזציה אחידה, ואם ניתן לפילוג הערכים בתמונה להשפיע, הרי שנוכל לתכנן קוונטיזר שאינו אחיד, וכבר כזאת ניצור אדפטציה של תהליך הקוונטיזציה לתמונה אותה אנו מעבדים. ברור כי בשמירת תמונה זו יש לשמור מידע נלווה על מדרגו הקוונטיזרים השונים לשם בניית התמונה מערכי האינדקסים שלה.

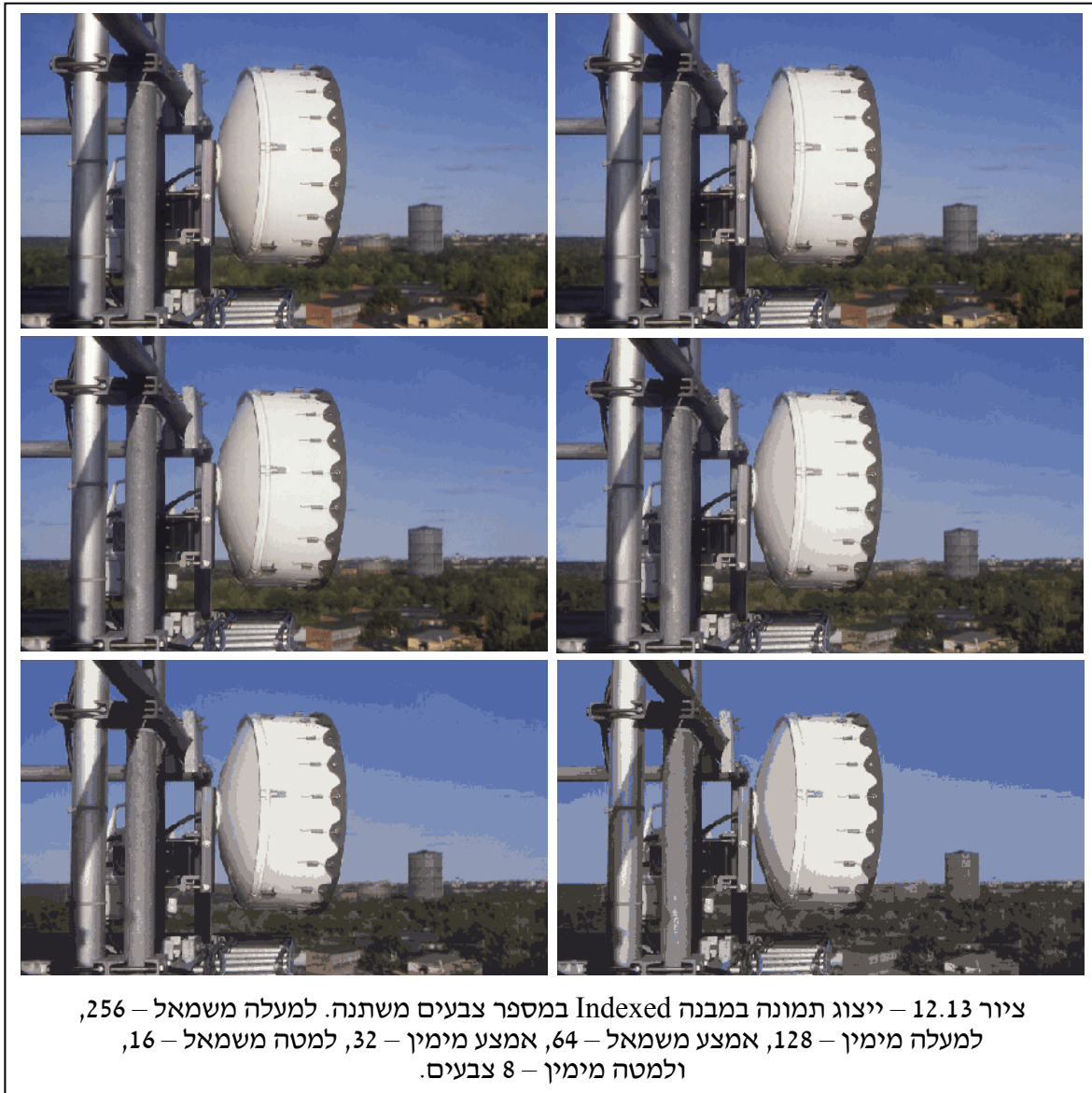


גישה כללית יותר להורדת כמות הצבעים היא גישה שתיפרד מהתפיסה הספרבילית גם בייצוג האלטרנטיבי ותפעל ע"י אישכול (Clustering). אם נחזור לרגע לייצוג התמונה ב-RGB, נוכל לבנות היסטוגרמה תלת-מימדית בה לכל תא $[R,G,B]$ יש מונה הסופר את מספר ההופעות של שלשה זו. לתמונה נתונה ניתן לצפות כי יהיו צבירים – קבוצה גדולה של הופעות של אותו צבע או צבעים קרובים למדי, ועבורם נוכל לבחור צבע מייצג יחיד.

זו בדיוק הגישה המובילה לקוונטיזציה וקטורית שנדונה בפרק 4. אנו נתייחס לאוסף הפיקסלים כקבוצת וקטורים בתלת-מימד, ונחפש מילון ייצוג בעל מספר צבעי ייצוג שייקבע מראש. פירוש הדבר שמבנה התמונה יהפוך למבנה חדש בו יש טבלת צבעים לבניית התמונה שתכלול N צבעים (זהו המילון שנמצא), ולכל פיקסל יינתן אינדקס אל טבלת הצבעים על מנת לומר מהו הצבע המאכלס פיקסל זה.

ב-Matlab מבנה זה נקרא Indexed והפקודה `rgb2ind` ממירה מתמונת RGB למבנה החדש ומספקת מפת Index וסט צבעים Palette. הצגת תמונת Indexed נעשית ע"י הצגת מערך האינדקסים כתמונה ע"י הפקודה `image`, ואז מיפוי הערכים לצבעיהם המתאימים ע"י הפקודה `colormap`. בציור 12.13 מוצגת התמונה שכבר הצגנו בציורים קודמים כשהיא מומרת למבנה Indexed עם מספר משתנה של צבעים ב-Palette. בעוד שבגישה הקודמת נדרשנו לכ-14 סיביות לשמור את מראה התמונה, הרי שכאן

אנו רואים כי די ב-6 סיביות והתוצאה קרובה מאוד. אנו נראה כי גישה זו מובילה לאלטרנטיבה מעניינת בדחיסת תמונות צבע.



כיצד תיעשה פעולת תכנון המילון במעבר מתמונת RGB לתמונת Indexed? גישה אפשרית אחת היא האלגוריתם שהוצג בפרק 4 – זהו האלגוריתם המכליל את אלגוריתם Max-Lloyd ונקרא K-Means או בתחום של דחיסת מידע – LBG על שם שלושת יוצריו (Linde, Buso, Gray). עם זאת, ישנן אלטרנטיבות פשוטות יותר אשר לא פעם נותנות תוצאה ויזואלית נעימה יותר. אלגוריתמים כמו Octree, Median-Cut ואחרים מוכרים בספרות כאלטרנטיבות טובות, אך אנו נעצור את הדיון בנושא זה כאן.

מרחב צבע לבניית ה-Palette מקובל להיות מרחב ה-RGB. הסיבה לכך היא הרצון למדוד מרחקים בין צבעים בצורה הוגנת. כזכור (ראה פרק 4), בבניית המילון נעשה שימוש איטרטיבי בצעד חישוב מרכז כובד הדורש מיצוע של צבעים, וצעד מיפוי הדורש חישוב מרחקים בין צבעים. במרחב RGB שתי פעולות אלה קלות – בהינתן שתי שלשות RGB, מרחק אויסקלידי ביניהן (סכום ריבועי הפרשים)

מהווה הערכה לא רעה למידת המרחק הויזואלי בין צבעים אלו, ולכן שימוש ב- RGB נראה מתאים. לעומת זאת, למשל, במעבר ל- YIQ, מרחב הצבע מתעוות וחישוב מרחק אויקלידי אינו הוגן. מסתבר כי נושא זה של מרחקים אויקלידים ונאמנותם לפערי צבעים בעין מהותי ומושך עניין רב. בספרות ישנו דיון מעמיק בסוגיה זו והוצעו מרחב צבע כגון La^*b^* שנועדו לפתור בעיה זו בדיוק.

אם כבר נבנה ה-Palette, נוכל לשפר את התמונה ע"י בחירת אינדקסים בצורה שונה מעט – דבר בו נפגשנו בעבר כשדיברנו על Half-Toning. נניח כי בסריקה שורה לאחר שורה קבענו את האינדקסים עד לפיקסל ה- (k,j) , וכעת עלינו לקבוע את האינדקס במקום זה. בגישה רגילה נקבע את האינדקס להצביע על הצבע ב- Palette אשר יהיה הקרוב ביותר לצבע בפיקסל זה. בבחירה זו אנו מתעלמים מהבחירה שנעשתה לסביבת הפיקסל, ובכך יכול להיות שאנו מנציחים שגיאה מצטברת. תופעה זו בדיוק היא העומדת מאחורי מראה השמיים בציור 12.12 בבחירת 8 צבעי Palette. נוצר מצב בו נוצר קו תפר בין שתי רמות צבע – קו שאינו שם במציאות. האלטרנטיבה היא גישה הקרויה Error Diffusion, בה אנו בוחרים אינדקסים תוך כדי צבירת שגיאה והתייחסות אליה.

בשיטת ה- Error-Diffusion אנו נתחיל בכך שנגדיר שכנות סיבתית לפיקסל ה- (k,j) – קבוצת פיקסלים עבודה כבר נעשתה החלטה לגבי האינדקס – נסמן שכנות זו ב- Ψ . נוכל להניח לשם פשטות כי שכנות זו כוללת את שלושת הפיקסלים מעל ואחד משמאל לפיקסל הנוכחי. בכל פיקסל p בשכנות זו יש שלשה של RGB מקורית שתסומן $\{R, G, B\}_p$. לכל פיקסל p יש גם את הצבע שנבחר מתוך ה-Palette, אשר יסומן כ- $\{\hat{R}, \hat{G}, \hat{B}\}_p$. השגיאה הממוצעת על-פני שכנות זו נתונה כ:

$$\Delta R = \sum_{p \in \Psi} R_p - \hat{R}_p \quad \Delta G = \sum_{p \in \Psi} G_p - \hat{G}_p \quad \Delta B = \sum_{p \in \Psi} B_p - \hat{B}_p$$

כעת, בהינתן הפיקסל החדש x בעל השלשה $\{R, G, B\}_x$, נרצה לבחור את הצבע המתאים לפיקסל זה. אנו ניחס לפיקסל זה את הצבע הקרוב ביותר ל- $\{R - \Delta R, G - \Delta G, B - \Delta B\}_x$ מתוך צבעי ה-Palette. בצורה זו, אם השגיאה היא אפס זהותית, נימצא מפעילים את האלגוריתם הפשוט של בחירת השכן הקרוב מתוך המילון.

כאשר למשל השגיאה באדום שלילית, פירוש הדבר שהגוון האדום הכללי בסביבה כהה מהדרוש ויש להבהירו. לכן, החסרת שגיאה זו תיתן ערך יעד חדש בהיר יותר מולו נחפש שכן קרוב ביותר במילון. ציור 12.14 מראה 4 תמונות בהן נעשתה פעולת ה- Error-Diffusion בבחירת האינדקסים. בתמונות אלו יש 64, 32, 16 ו- 8 צבעים, כמו רביעיית התמונות התחתונות בציור 12.13, והשיפור ביחס לגישה הרגילה ניכר.

שיטת ה- Error Diffusion רבת פנים, וניתן להכלילה בדרכים שונות. ניתן, למשל, לדרוש כי בבחירת צבע שאינו השכן הקרוב המידי, לכפות רק תת-קבוצה של צבעים אפשריים על מנת למנוע קפיצות צבע עצבניות מידי. שיטה זו רווחת במדפסות צבע בהן יש מספר מחסניות צבע (יותר מ- 3) בבואן להדפיס תמונה (בניגוד לגרפיקה או טקסט, בהם מנגנון זה מכובה).



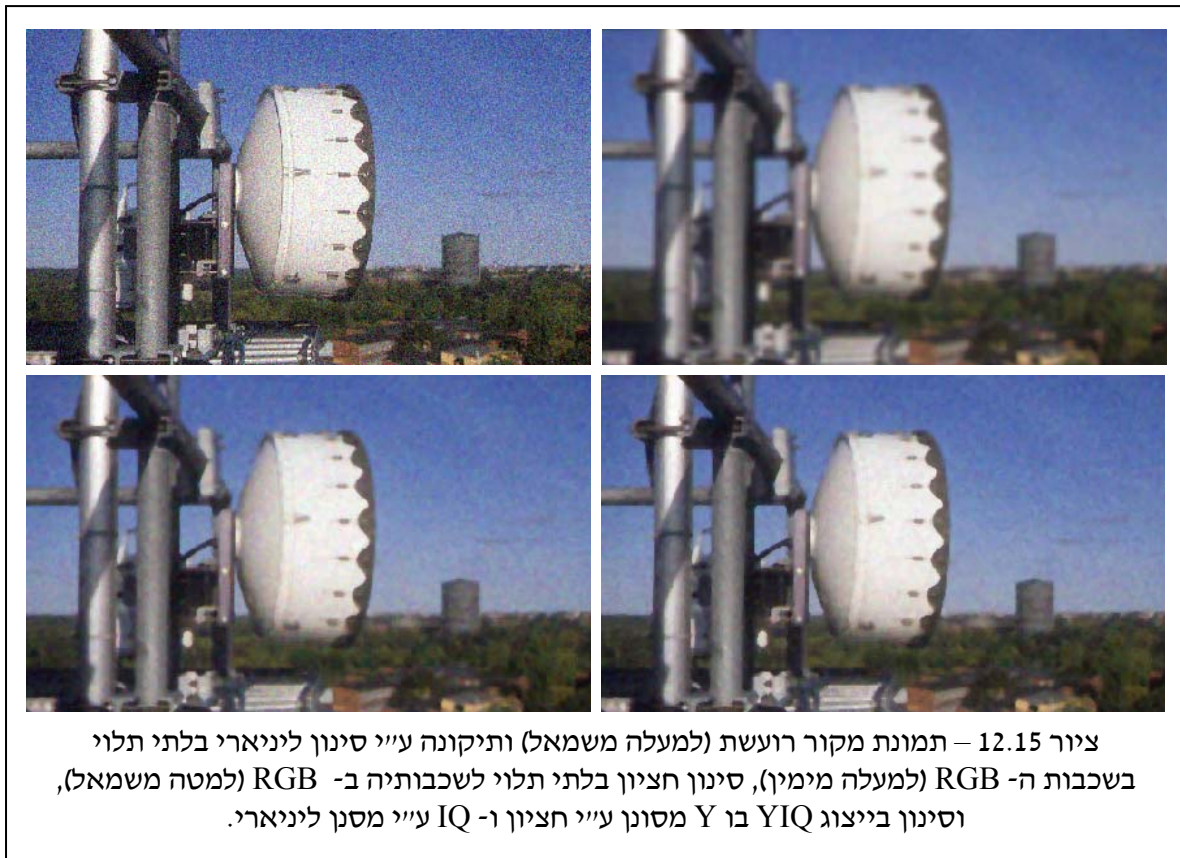
בהקשר של תמונות בעלות מבנה Indexed מקובל לעיתים קרובות השימוש בפסאודו-צבע. אם נתונה לנו תמונה שאינה מתארת תצלום אלא גודל כלשהו (טמפרטורה כפונקציה של המקום, עומק מים, שדות אלקטרומגנטיים ועוד), ניתן ליחס לערכיה צבעים לשם קבלת אפקט ויזואלי חזק יותר. אין כל חוקיות בדרך הבחירה, וזוהי בדרך-כלל יותר אומנות מאשר מדע (כמו הרבה דברים בוויזואליזציה). אנו נסתפק רק באיזכור המושג.

12.5 שיפור ושחזור תמונת צבע

פעולות היסטוגרמה וסינון של תמונת צבע יכולים להיעשות על כל אחת משכבות ה-RGB בנפרד, ובהרבה מקרים כך זה אמנם מופעל. בגישה זו ייתכנו עיוותי צבע מטרידים. לדוגמה, אם רצוננו לנקות רעש מתמונה ע"י החלקה פשוטה, ונפעיל מסנן 5 על 5 אחיד על התמונה, בממוצעים המחושבים ימוזגו צבעים ליצירת צבע חדש שאינו בהכרח קיים בתמונה המקורית. גישה אלטרנטיבית מקובלת היא הפעלת הסינון על תמונה לאחר המרה לייצוג המפריד בהירות ממאפייני צבע.

אם נתייחס לרגע ליישום הבסיסי של הורדת רעש מתמונת צבע תוך כדי שמירה על שפות, נוכל להציע שלוש אלטרנטיבות – פעולת סינון ליניארי על שכבות ה-RGB באופן בלתי תלוי, פעולת חציון על שכבות ה-RGB באופן בלתי תלוי, או המרה ל-YIQ, ביצוע חציון על שכבת ה-Y, וטשטוש פשוט על שכבות ה-I וה-Q. תוצאות גישות אלה מוצגות בציור 12.15, ואנו רואים שוב כי הפרדה של בהירות

מצבע משרתת את המטרה של פישוט הפתרון ע"י הקצאת התהליך המורכב (חציון) לבהירות, והסתפקות בפתרון פשוט (סינון ליניארי) בשכבות הצבע.



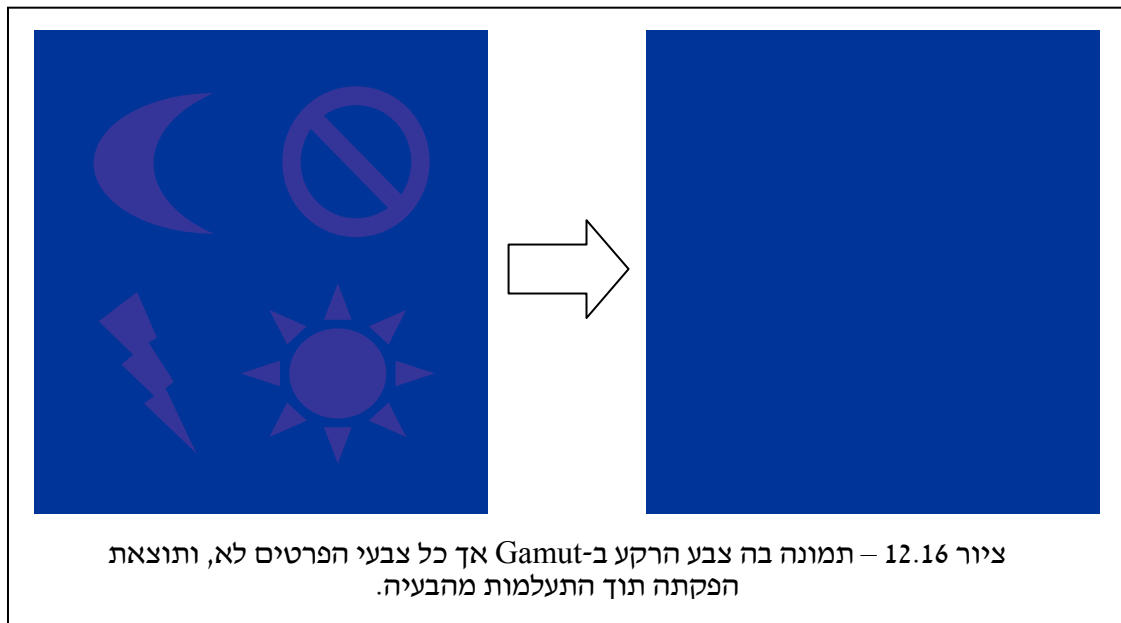
באופן דומה נוכל לטפל בתמונות עם מגוון הכלים שהוצגו בפרק 5 – פעולות היסטוגרמה, סינון וסינון אדפטיבי – כל אלה יכולים להיעשות ע"י שימוש בייצוג נוח המפריד את הבהירות אשר עליה ייעשה עיקר העיבוד. גם שחזור תמונות מטשטוש ידוע יכול להיעשות בצורה זו- ביצוע השחזור על שדה הבהירויות והשארת שכבות הצבע כמות שהן, שהרי ממילא לא רואים את חדותן.

במסגרת פרק זה רצוננו להתבונן בבעיות מיוחדות הקשורות לשיפורה של תמונה הנובעות ממעורבות הצבע, וכאלה הדורשות פתרונות חדשים. אנו נתמקד בשלושה נושאים מסוג זה - מגבלות Gamut והתייחסות אליהן, פונקציות הסתברות לתמונות צבע – הקשר בין שכבות ה-RGB, ושחזור מתת-דגימה – Demosaicing.

המושג Gamut פירושו טווח הצבעים הניתנים לשימוש בהתקן פלט. לפיכך, Gamut של מדפסת (או מסך) הוא אוסף כל שלשות ה-RGB אותן יכולה המדפסת לייצר. לרוב ה-Gamut מוגבל ואינו כולל את מלוא קוביית ה-RGB. בהינתן תמונה אותה רצוננו להפיק בהדפסה, מה עלינו לעשות לתמונה בטרם הפנייתה למדפסת, על מנת לקחת בחשבון את מגבלותיה? אם נתעלם והתמונה אמנם מכילה צבעים מחוץ ל-Gamut, התוצאה תהיה קיצוץ של צבעים אלה לצבע הקרוב ביותר אשר ניתן להדפסה – זוהי הטלה

פשוטה של צבעי התמונה על טווח ה-Gamut. במקרים רבים זה מספיק, ובעיקר אם התמונה לא כוללת פיקסלים רבים עם גוונים מחוץ ל-Gamut.

במקרים אחרים, פעולת קיצוץ זו מעלימה פרטים חשובים. אם באזור שמרביתו מחוץ ל-Gamut ישנן שפות המתארות פרטים חשובים, מרבית שפות אלה תימחקנה לחלוטין בעקבות פעולת ההטלה הפשוטה. דוגמה סינטטית לתופעה זו נתונה בציור 12.16.



אם ננסח את תהליך ההטלה כפעולת מינימיזציה, ישפוך הדבר אור על דרך פתרון אפשרית. נסמן את תמונת המקור כ- \underline{Y} , ואת תמונת היעד כ- \underline{X} . ה-Gamut יתואר כקבוצה Θ . לכן, פעולת ההטלה פירושה שבהינתן התמונה \underline{Y} , עלינו למצוא את \underline{X} הקרובה אליה ביותר (מחושב פיקסל מול פיקסל) כך שהתמונה \underline{X} כולה בתוך הקבוצה Θ . בניסוח מתמטי,

$$\underline{X}_{\text{opt}} = \underset{\underline{X}}{\text{ArgMin}} \|\underline{X} - \underline{Y}\|_2^2 \quad \text{Subject to } \underline{X} \in \Theta$$

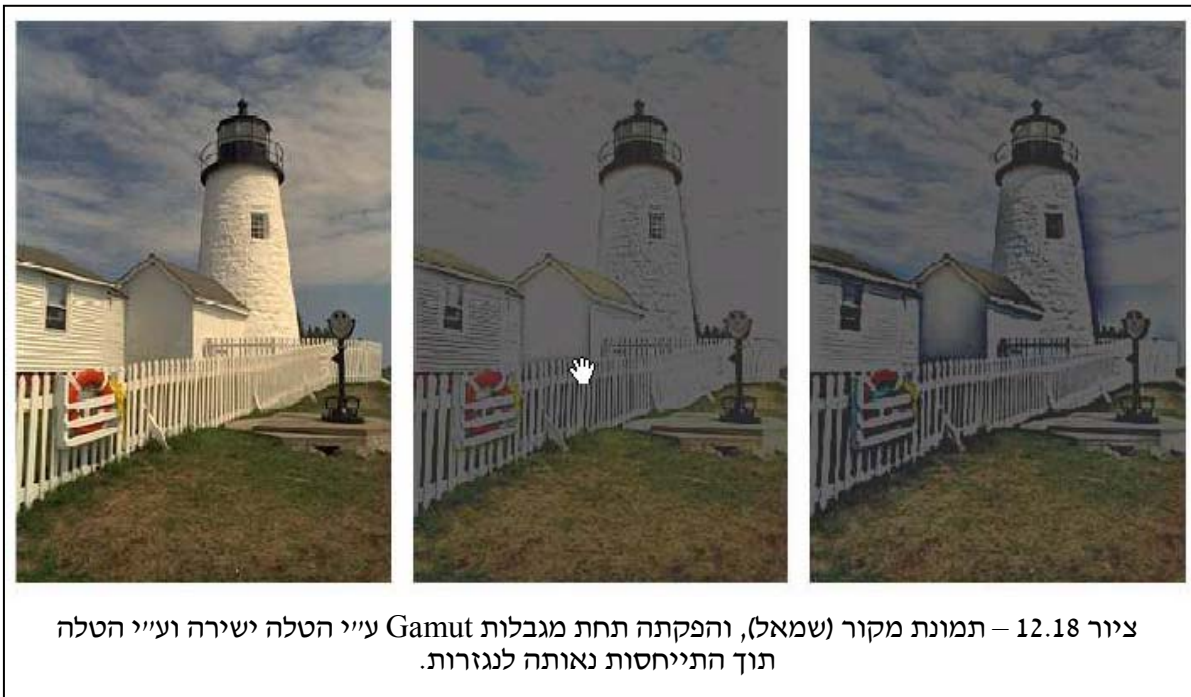
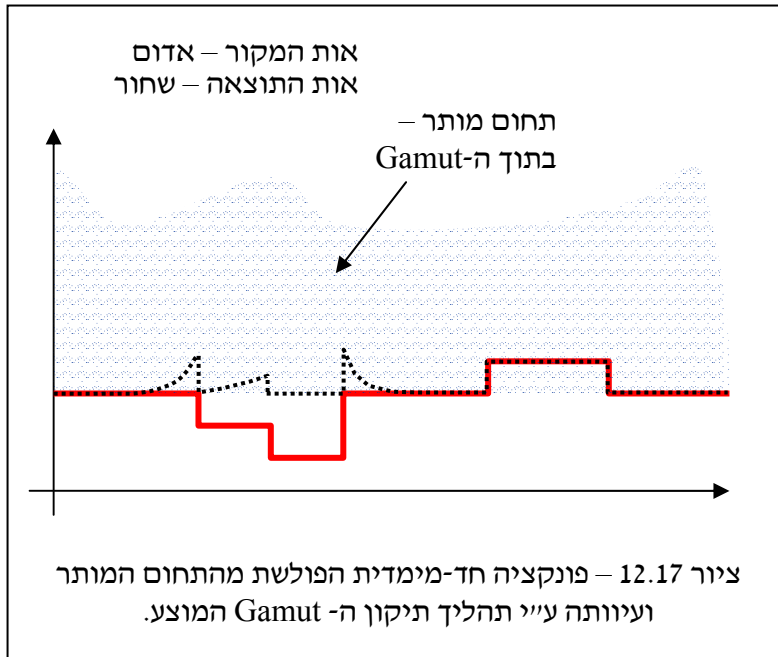
וכאמור, הפתרון פשוט למדי ופועל על כל פיקסל באופן בלתי תלוי בשכניו.

שיפור של תהליך זה יבוא אם ניתן לשכנים כוח להתערב בתוצאת הקביעה של הפיקסל הנוכחי. זאת נוכל להשיג ע"י מידת מרחק בין התמונות \underline{X} ל- \underline{Y} אשר תראה לא רק הפרשי רמות RGB בין הפיקסלים בשתי התמונות, אלא גם את הפרשים בנגזרות. נוכל להציע ביטוי מהצורה

$$\underline{X}_{\text{opt}} = \underset{\underline{X}}{\text{ArgMin}} \|\underline{X} - \underline{Y}\|_2^2 + \lambda \|\mathbf{D}(\underline{X} - \underline{Y})\|_2^2 \quad \text{Subject to } \underline{X} \in \Theta$$

הגורם החדש מחשב נגזרות מרחביות על שכבות ה-RGB (כל אחת בנפרד, ויתכן אף שבכמה כיוונים), והדרישה היא שהקירבה שתהיה בין \underline{X} ל- \underline{Y} תהיה גם בין נגזרותיהן המרחביות. לכן, אם יש שפה בתמונה אחת, מן הראוי שתהיה שפה גם באחרת. עבור הדוגמה שבציור 12.15 המשמעות היא שהרקע ישנה את ערכו קרוב לשפות, על מנת לאפשר קפיצה שתראה במסגרת ה-Gamut המוגבל. ציור 12.17

ממחיש גרפית את השפעת השימוש בנגזרות על פונקציה חד מימדית (למשל, שורה אחת בתמונה שבציור 12.16). ציור 12.19 מראה דוגמה מעשית עם התוצאה הצפויה.



הפתרון הנומרי של בעיית האופטימיזציה שניתנה יכול היעשות ע"י איטרציות השיפוע המירבי, תוך כדי הטלה על מרחב האילוצים. לכן

$$\hat{\mathbf{X}}_{k+1} = \text{Proj}_{\Theta} \left\{ \hat{\mathbf{X}}_k - \mu \left[\hat{\mathbf{X}}_k - \mathbf{Y} + \lambda \mathbf{D}^T \mathbf{D} (\hat{\mathbf{X}}_k - \mathbf{Y}) \right] \right\}$$

במהותה, פעולת ההטלה היא אותה פעולה שעשינו קודם – חישוב הצבע הקרוב ביותר מתוך ה-Gamut כשזה נעשה לכל פיקסל בנפרד. השוני הוא במעורבותן של נגזרות מרחביות בעדכון הפתרון. ישנן כמובן דרכים להאיץ התכנסות התהליך, אך אנו לא ניכנס לנושא זה. אם ה-Gamut יוצר קבוצה קמורה, אלגוריתם זה מובטח להתכנס.

אנו עוברים כעת לנושא של פונקציות הסתברות לתמונות כשהפעם נרצה לקחת בחשבון צבע. זהו נושא הנמצא עדיין בחיתוליו, והידע המצוי בידנו כיום, לרבות מה שיוצג כאן ראשוני בלבד. עבור תמונות שחור-לבן כבר ניסחנו פונקציית הסתברות זו כפילוג Gibbs מהצורה

$$P(\underline{X}) = C \cdot \exp\{-\alpha^2 \rho\{\underline{DX}\}\}$$

(ראה פרק 6) כאשר C נועד לנרמול פונקציה זו, α הוא פרמטר שקובע את מידת פיזור ההסתברות, ו- $\rho(x)$ היא פונקציה רובסטית כגון פונקצית הערך המוחלט, על מנת להרשות שפות בתמונה. אנו רוצים להכליל פונקציה זו לתמונת צבע וההשערה הראשונה שבאה בחשבון היא המבנה

$$\begin{aligned} P(\underline{X}) &= C \cdot \exp\{-\alpha^2 [\rho\{\underline{DX}_R\} + \rho\{\underline{DX}_G\} + \rho\{\underline{DX}_B\}]\} = \\ &= C \cdot \exp\{-\alpha^2 \rho\{\underline{DX}_R\}\} \cdot \exp\{-\alpha^2 \rho\{\underline{DX}_G\}\} \cdot \exp\{-\alpha^2 \rho\{\underline{DX}_B\}\} \end{aligned}$$

כלומר, בבחירה זו הצענו כי שכבות ה-RGB כל אחת לגופה חלקה למקוטעין. הפגם בבחירה זו הוא שאנו לא מציעים קשר בין השכבות השונות. למעשה, מהעובדה שההסתברות נכתבת כמכפלה של שלוש פונקציות הסתברות לכל שכבה נובע כי אנו מניחים אי-תלות. האמנם שכבות ה-RGB בלתי תלויות זו בזו?

בחינה של סוגיה זו מגלה כי ישנו קשר הדוק הקובע כי מיקומן של שפות וכיווניהן צריך להתאים בין שכבות ה-RGB. ישנן דרכים מגוונות לנסח זאת מתמטית. אחת הדרכים הפשוטות היא הרישום

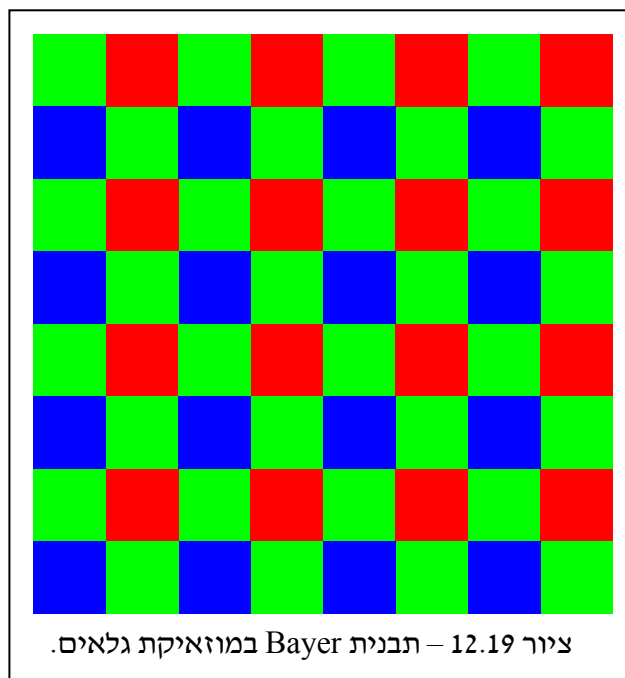
$$\begin{aligned} P(\underline{X}) &= C \cdot \exp\{-\alpha^2 [\rho\{\underline{DX}_R\} + \rho\{\underline{DX}_G\} + \rho\{\underline{DX}_B\}]\} \cdot \\ &\cdot \exp\{-\beta^2 [\eta\{\underline{D}(\underline{X}_R - \underline{X}_G)\} + \eta\{\underline{D}(\underline{X}_G - \underline{X}_B)\} + \eta\{\underline{D}(\underline{X}_B - \underline{X}_R)\}]\} \end{aligned}$$

המרכיב הראשון דורש כמקודם את החלקות למקוטעין של כל שכבה בנפרד, והמרכיב השני דורש התאמת שפות בין צמדי שכבות. הפונקציה η מתמחרת את שינויי השפות בין השכבות, וגם היא יכולה להיבחר כפונקציה רובסטית כלשהי. דרך אחרת להסתכל על הביטוי השני היא הדרך הבאה – גם תמונת הפרש בין שתי שכבות צריכה להיות חלקה למקוטעין ועם מעט שפות ככל האפשר. ברור כי אם השפות תתלכדנה נקבל כי אנרגיה זו תוקטן, כפי שאנו רוצים. ניתן להמשיך ולשפר פונקציה זו לקחת בחשבון כי עוצמת הנגזרת בשפה אינה זהה בין השכבות ולכן יידרש גורם תיקון שיהפוך לחלק מהנעלמים.

בהינתן פונקציית הסתברות מוצלחת לתמונת צבע, אנו ממושים לפתרון בעיות היפוך בהן הנעלם היא תמונת צבע. נוכל בדרך זו לפתח סינון מבוסס MAP לשם שחזור תמונה מטשטוש ורעש, ובכך לעקוף את ההצעה האמפירית להמיר התמונה ל-YIQ בטרם השחזור.

בעיית שחזור מעניינת המחייבת פונקציית Prior מהצורה הנ"ל היא בעיית ה-Demosaicing. מצלמות דיגיטליות המצלמות תמונות צבע צריכות בפועל שלושה מערכים שווים גודל אשר יחשו את שכבות ה-R, ה-G וה-B בנפרד. אם לא די בכך, הרי ששלושת מערכי גלאים אלה צריכים "לראות" את היעד עם תיאום גיאומטרי מלא – שכל שלשת פיקסלים מתואמת RGB תתייחס למקום זהה באובייקט המצולם. דבר זה אפשרי ע"י הכנסת האור למצלמה ופיצולו בפריזמה לשלושת הגלאים.

כל הגישה הזו יקרה למדי אך יש מצלמות שפועלות כך והן נחשבות למצלמות מיקצועיות יותר והן כמובן גם יקרות יותר. האלטרנטיבה היא שימוש במערך גלאים אחד שבו חלק מהגלאים רואים ירוק, חלק אדום וחלק כחול, כשהבחירה נעשית ע"י יציקת מסיכה קטנה על פני כל גלאי לבחירת הצבע המתאים. אחת הדרכים ליצור מוזאיקה זו היא ה-Bayer Pattern אשר נראה כבציור 12.19.



בעיית השחזור אם כך היא: בהינתן מוזאיקת הערכים (שפירוש תת-דגימה של תמונת הצבע המלאה – RGB), שחזר את הערכים החסרים. שיטות לשחזור התמונה המלאה קרויות אלגוריתמי Demosaicing, וכאלה קיימים רבים. יש לזכור כי כל מצלמה דיגיטלית ביתית כיום כוללת בתוכה תהליך Demosaicing ייחודי, ולא פעם איכות התמונה במצלמה נקבע על-פי האלגוריתם שיושם.

ניתן להציע גישה פשטנית שמבצעת שחזור של הירוקים מהירוקים, האדומים מהאדומים והכחולים מהכחולים. שחזור זה יכול להיעשות ע"י שיטות אינטרפולציה שלוקחות ערכים ידועים בסביבה, מקרבים פולינום דו-מימדי מסדר כלשהו, ומתוך קירוב זה קובעים את הערך החסר. גישה זו תיתן לרוב תוצאה גרועה בסביבת שפות, בשל אי-תיאום של מיקום השפה בשכבות הצבע – דבר היוצר הד מטריד.

שיטה מבוססת על שחזור MAP אשר תעשה שימוש בפונקציית ההסתברות שהוצעה קודם

תציע את בעיית המינימיזציה הבאה :

$$\cdot \text{Min}_{\underline{X}_R, \underline{X}_G, \underline{X}_B} \left\{ \begin{aligned} & \|\mathbf{S}_R \underline{X}_R - \underline{Y}_R\|_2^2 + \|\mathbf{S}_G \underline{X}_G - \underline{Y}_G\|_2^2 + \|\mathbf{S}_B \underline{X}_B - \underline{Y}_B\|_2^2 + \\ & + \alpha^2 [\rho\{\mathbf{D}\underline{X}_R\} + \rho\{\mathbf{D}\underline{X}_G\} + \rho\{\mathbf{D}\underline{X}_B\}] + \\ & + \beta^2 [\eta\{\mathbf{D}(\underline{X}_R - \underline{X}_G)\} + \eta\{\mathbf{D}(\underline{X}_G - \underline{X}_B)\} + \eta\{\mathbf{D}(\underline{X}_B - \underline{X}_R)\}] \end{aligned} \right\}$$

בביטוי זה, הסימון \underline{Y}_R ניתן למדידות בפיקסלים בהם יש ערך אדום, ובאופן דומה אנו מגדירים את \underline{Y}_G

ו- \underline{Y}_B . האופרטור \mathbf{S}_R הוא האופרטור המקבל תמונת אדום מלאה ובוחר ממנה את הערכים שקיימים

במוזאיקה. גזירת ביטוי זה ביחס ל- $\underline{X}_R, \underline{X}_G, \underline{X}_B$ נותנת

$$\frac{\partial f}{\partial \underline{X}_R} = \mathbf{S}_R^T (\mathbf{S}_R \underline{X}_R - \underline{Y}_R) + \alpha^2 \mathbf{D}^T \rho'\{\mathbf{D}\underline{X}_R\} + \beta^2 \mathbf{D}^T [\eta'\{\mathbf{D}(\underline{X}_R - \underline{X}_G)\} + \eta'\{\mathbf{D}(\underline{X}_R - \underline{X}_B)\}]$$

$$\frac{\partial f}{\partial \underline{X}_G} = \mathbf{S}_G^T (\mathbf{S}_G \underline{X}_G - \underline{Y}_G) + \alpha^2 \mathbf{D}^T \rho'\{\mathbf{D}\underline{X}_G\} + \beta^2 \mathbf{D}^T [\eta'\{\mathbf{D}(\underline{X}_G - \underline{X}_R)\} + \eta'\{\mathbf{D}(\underline{X}_G - \underline{X}_B)\}]$$

$$\frac{\partial f}{\partial \underline{X}_B} = \mathbf{S}_B^T (\mathbf{S}_B \underline{X}_B - \underline{Y}_B) + \alpha^2 \mathbf{D}^T \rho'\{\mathbf{D}\underline{X}_B\} + \beta^2 \mathbf{D}^T [\eta'\{\mathbf{D}(\underline{X}_B - \underline{X}_R)\} + \eta'\{\mathbf{D}(\underline{X}_B - \underline{X}_G)\}]$$

ובתהליך איטרטיבי ניתן להפעיל עדכון מבוסס SD שיתכנס למינימום של הפונקציה שתוארה. אם נתבונן בביטויי הגזירה נגלה, בדומה לשחזור תמונות, כי אין כל צורך בייצוג וקטורי מטריצי ובניסוח אופרטורים כמטריצות. אבני הבניין הנחוצים פשוטים למדי :

א. החסרת תמונות – כגון ביצוע $\underline{X}_R - \underline{X}_B$ או $(\mathbf{S}_R \underline{X}_R - \underline{Y}_R)$. פעולה זו תיעשה כהחסרה פיקסל

מפיקסל כששתי התמונות שמורות כמעריך דו-מימדי.

ב. הפעלת אופרטור גזירה כלפליסיאן – ההכפלה ב- \mathbf{D} או \mathbf{D}^T . פעולה זו תיעשה כקונבולוציה על

התמונה, עם הכנסת שיקולי קצוות כגון המשכה מחזורית או תמונת מראה.

ג. הפעלת פונקציה סקלארית על פיקסלים בתמונה – כגון $\rho'\{\mathbf{D}\underline{X}_R\}$ או $\eta'\{\mathbf{D}(\underline{X}_R - \underline{X}_B)\}$.

פעולות אלו אינן אלא פעולות LUT על כל פיקסל. למשל, עבור הבחירה $\rho(x) = |x|$, הפעולה

הנדרשת היא נגזרת פונקציה זו – דהיינו פעולת סימן. אם ערך הפיקסל שלילי, -1 ואם חיובי +1.

ד. הפעלת האופרטור \mathbf{S}_R פירושה בחירת הפיקסלים שישנם במוזאיקה והתעלמות מאחרים.

הפעולה המשוחלפת פירושה הכנסת אפסים לערכים החסרים.

קיימות כאמור גישות שונות ומגוונות לבניית תמונת ה-Demosaic. המשותף למוצלחות שבהן

הוא ניסוח נאות של הקשר בין שכבות הצבע. בצירוף 12.20 מוצגות שתי תוצאות של תהליך

Demosaicing מבוסס למידה חישובית¹ (כלומר, מראים למערכת מספר דוגמאות של תמונת מוזאיקה

והתמונה שהיתה צריכה להתקבל, והמערכת לומדת על הקשר בין שכבות הצבע מתוך דוגמאות אלה).

התוצאות מתייחסות לטיפול בכל שכבה בנפרד וטיפול משולב שבו רואים הרבה פחות פגעי צבע.

¹ תוצאות אלו נוצרו בעבודת המחקר של דר' חגית הל-אר מאוניברסיטת חיפה.



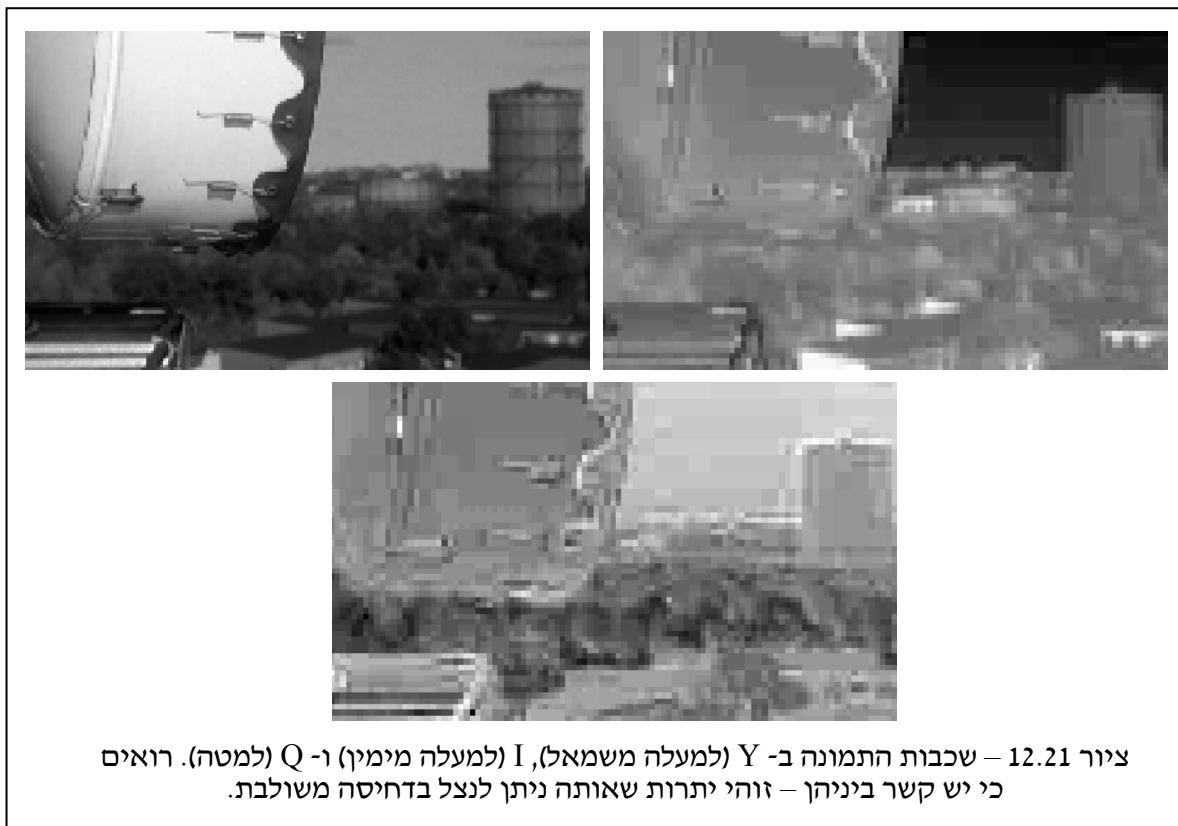
12.6 דחיסת תמונות צבע

כמו בכל סוגיה אחרת בעיבוד תמונות, כשהדבר נוגע לתמונת צבע, מוצע הפתרון האד-הוקי של הפרדת הבהירות ממאפייני הצבע. כבר ראינו כי ניתן להחזיק תמונת צבע בייצוג YIQ בן אנו זורקים 50% מסך המידע ע"י דילול ביחס 2:1 בכל ציר בשכבות ה-I וה-Q ללא נזק ניכר. הגישה מכאן יכולה להמשיך ע"י דחיסת כל אחת מהשכבות בנפרד בכל אחת מהגישות שנדונו קודם, כגון גישת חיזוי או התמרה, כשהם מלווים בקוונטיזציה מתאימה. בצורה זו אנו מנצלים במידה מסויימת את הקשר בין השכבות ע"י ההמרה והדילול של I ו-Q. עם זאת, אם נסתכל על התמונות הללו, כמו שמתוארות בציור 12.21, ברור כי עדיין קיים קשר ביניהן, גם אם חלש, ודחיסתן בנפרד מבזבזת יתרות זו.

נוכל גם להציע אלגוריתמי דחיסה שרואים את הקשר בין השכבות בצורה ישירה. כשהתחלנו את הדיון בדחיסת תמונה הזכרנו את שיטת ה-PCM שאינה אלא קוונטיזציה סקלרית ישירה על רמות האפור. בהכללה טבעית, נציע קוונטיזציה וקטורית על שלשות ה-RGB למספר מצומצם של רמות ייצוג ובכך לדחוס את התמונה – זהו בדיוק התהליך שעשינו ממילא לשם ייצוג נוח במסגרת הדיון בתמונות Indexed. מסתבר כי ניתן אף לעשות יותר – אם נתייחס למשפחת התמונות הסינטטיות כגון ציורים, גרפיקה, ועוד, הרי שמקבץ הצבעים שיידרש יהיה קטן. דבר זה יביא לדחיסות גבוהה גם בשל היות ה-Palette קטן יותר, וגם (וכנראה בעיקר) בשל היות סט האינדקסים בעל אלף-בית קטן יותר.

בהינתן סט האינדקסים עבור תמונה סינטטית, יתקבל כי אלה יוצרים תמונה חלקה למקוטעין. כלומר, יהיו אזורים רבים בהם האינדקס לא משתנה בין שכנים, ופירוש הדבר פוטנציאל דחיסה נוסף. שימוש בכלים כמו Run-Length-Encoding או וריאציות שונות של אלגוריתם Lempel-Ziv יכולים

לשמש לשם דחיסה ללא עיוות של האינדקסים, וזוהי בדיוק הגישה המובילה לפורמטי ה-GIF וה-PNG שמשמשות לתמונות גרפיקה ול-Icons. למעשה שני פורמטים אלה רווחים למדי באינטרנט בשל פשטותם ויעילותם.



אם נמשיך עם אותו קו מחשבה שננקט בפרק 8, הרי שקבענו כי דחיסת PCM (קוונטיזציה ישירות לערכי התמונה) אינה מוצלחת כמו דחיסת ההפרשים הסמוכים – דבר שהוביל אותנו לגישות חיזוי. נוכל להציע סכימת DPCM בה נחזה את שלשת ה-RGB בפיקסל על פי השכנים בדרך כלשהי. נגדיר סט שכנים סיבתי בסריקה בדר כלשהו של הפיקסלים, נשתמש בשכנים הסיבתיים לבנות חיזוי של שלושת ערכי ה-RGB, ונחסיר החיזוי מהערכים האמיתיים שישנם בפיקסל. את אות השגיאה נוכל להזין לקוונטיזר וקטורי שאותו נתכנן ע"י בניית סט גדול של הפרשי חיזוי צפויים. מעניין לציין שאפילו אם החיזוי של כל שכבה ייעשה בנפרד (דבר שמוביל לשגיאות חיזוי סקלריות שניתנות לחישוב בנפרד כשלוש סכמות DPCM נפרדות), עדין קידוד וקטורי תלת-מימדי עדיף על פני קוונטיזציה סקלרית. זאת כיוון שבקוונטיזציה סקלרית אנו מנצלים קשר בין השכבות, ומעבר לכך, מרצפים את המרחב בצורה יעילה יותר.

נספחים

285	<u>נספחים:</u>
286	נספח א' – הרחבות לקוונטיזר אופטימלי
294	נספח ב' – התמרת KLT כהתמרה אופטימלית והרחבותיה
299	נספח ג' – הקצאת סיביות אופטימלית בדחיסה
302	נספח ד' – נושאים מתקדמים בשחזור תמונות

בפרק זה נוסף דיון במספר נושאים מגוונים שמעשירים את הדיון שניתן בפרקים קודמים. בחירת הנושאים הנוספים נעשתה תוך חיפוש של סוגיות בעלות יופי תיאורטי וואו תרומה מעשית חשובה לתחום. מלכתחילה הכוונה הייתה להכניס נושאים אלו לפרקים המתאימים, וכנראה שכך ייעשה לבסוף.

נספח א' – הרחבות לקוונטיזר אופטימלי

בפרק 4 למדנו על תהליך הקוונטיזציה ועל דרכים ליעלו. ראינו כי ניתן לתכנן קוונטיזר אופטימלי להבאת השגיאה הריבועית למינימום. במסגרת נספח זה נדון בשתי וריאציות של קוונטיזר אופטימלי זה. בחלק הראשון ננסה לפשט את תהליך תכנון הקוונטיזר למקרים בהם פשוט כזה אפשרי. אנו נראה כי כאשר מספר המדרגות בפונקציית הקוונטיזציה גדול דיו, קירוב מפשט כזה אפשרי ואז תכנון הקוונטיזר היא משימה קלה יותר. חלק שני בנספח זה יידון בהחלפת קריטריון האופטימליות למינימיזציה של האנטרופיה. אנו נראה כי שילוב של קוונטיזר שלאחריו קידוד אנטרופיה יוצר פתרונות שונים בתכלית.

1. קוונטיזר Max-Lloyd

נתחיל באזכור מהיר של קוונטיזר Max-Lloyd. נתון לנו אות (נניח כי הוא חד-מימדי כיוון שבכל הדיון שיובא כאן נתייחס למקרה של הקוונטיזציה הסקלרית) $x[n]$. בעקרון ישנן שתי גישות לאפיין מקור זה – האחת, דטרמיניסטית, תאמר כי $x[n]$ נתון לנו עם אינסוף דגימות (כלומר $n = 0, 1, 2, \dots, \infty$), ואנו נוכל לייצר היסטוגרמה של ערכי $x[n]$ אשר תסומן $p(x)$ והיא תיתן לנו את שכיחותו של הערך x . גישה סטוכסטית המובילה לאותו מקום היא האמירה שערכי f מפולגים לפי הפילוג הסקלרי $p(x)$ – כלומר p היא פונקציית ההסתברות להגשמת ערכי x .

הקוונטיזר אותו אנו עומדים לתכנן הינו פונקציית מדרגות כמתואר בציור A.1. הפונקציה Q מבצעת קפיצה בערכי x השווים $\{r_k\}_{k=0}^N$ (אלו הם רמות ההחלטה), וערכה באינטרוול $[r_{k-1}, r_k]$ הוא f_k (אלו הם רמות הייצוג). אנו נניח כי ערכי הפונקציה מקבלים כל ערך ממשי, ולכן $r_0 = -\infty$ ו- $r_N = +\infty$ קבועים וידועים מראש.

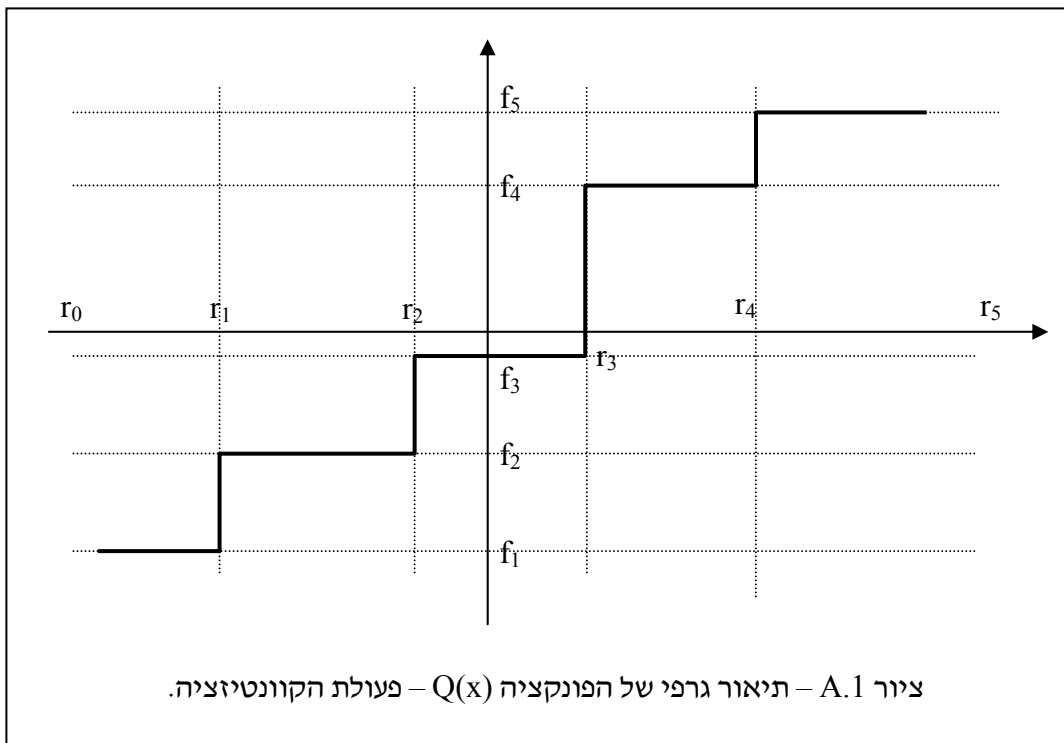
בתפיסה הדטרמיניסטית נגדיר את השגיאה הממוצעת בין האות x ובין הקוונטיזציה שלו $Q\{x\}$

ע"י

$$\bar{E} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N (x[n] - Q\{x[n]\})^2$$

כיוון שהפונקציה Q היא פונקציית הקוונטיזציה וזו סקלרית וחסרת זכרון, ניתן לשנות את דרך הסכימה בסכימה על פני הערכים האפשריים של x . המבנה החדש יהיה

$$\bar{E} = \lim_{L \rightarrow \infty} \frac{1}{L} \sum_{n=1}^L (x[n] - Q\{x[n]\})^2 = \int_{-\infty}^{\infty} p(x) (x - Q\{x\})^2 dx$$



בשינוי זה קבענו כי כל ערך של x מופיע $p(x)$ פעמים (לאחר נירמול), ולכן במקום לסכום על ציר הזמן אנו סוכמים על ציר ערכי x . עם שינוי זה קיבלנו כי ייצוג הבעיה חוזר למבנה הסטוכסטי, כיוון שהשגיאה הממוצעת לפי גישה סטוכסטית נתון ע"י

$$\bar{E} = E\{(x - Q\{x\})^2\} = \int_{-\infty}^{\infty} p(x)(x - Q\{x\})^2 dx$$

וזהו כמובן אותו ביטוי בדיוק.

כיוון שהפונקציה Q היא פונקצית מדרגות המבצעת קפיצה בערכי x השווים $\{r_k\}_{k=0}^N$, וערכה

באינטרוול $[r_{k-1}, r_k]$ קבוע ושווה f_k , נוכל לפרק את האינטגרל הני"ל לנתחים ולקבל

$$\bar{E} = \int_{-\infty}^{\infty} p(x)(x - Q\{x\})^2 dx = \sum_{k=1}^N \int_{r_{k-1}}^{r_k} p(x)(x - f_k)^2 dx$$

מטרתנו היא מציאת $\{r_k\}_{k=0}^N$ ו- $\{f_k\}_{k=1}^N$ שיביאו שגיאה זו למינימום. גזירת השגיאה לפי הנעלם f_k

מניבה

$$\begin{aligned} \frac{\partial}{\partial f_k} \sum_{k=1}^N \int_{r_{k-1}}^{r_k} p(x)(x - f_k)^2 dx &= \int_{r_{k-1}}^{r_k} p(x) \frac{\partial}{\partial f_k} (x - f_k)^2 dx = 2 \int_{r_{k-1}}^{r_k} p(x)(f_k - x) dx = 0 \\ &= \int_{r_{k-1}}^{r_k} x \cdot p(x) dx \\ \Rightarrow f_k &= \frac{\int_{r_{k-1}}^{r_k} x \cdot p(x) dx}{\int_{r_{k-1}}^{r_k} p(x) dx} \end{aligned}$$

כלומר, בידיעת הסידרה $\{r_k\}_{k=0}^N$, רמות הייצוג יימצאו כמרכזי כובד של האינטרוולים הקבועים. באופן דומה, גזירה לפי r_k נותנת

$$\begin{aligned} \frac{\partial}{\partial r_k} \sum_{k=1}^N \int_{r_{k-1}}^{r_k} p(x)(x - f_k)^2 dx &= \frac{\partial}{\partial r_k} \left[\int_{r_{k-1}}^{r_k} p(x)(x - f_k)^2 dx + \int_{r_k}^{r_{k+1}} p(x)(x - f_{k+1})^2 dx \right] = \\ &= p(r_k)(r_k - f_k)^2 - p(r_k)(r_k - f_{k+1})^2 = 0 \\ \Rightarrow r_k &= \frac{f_k + f_{k+1}}{2} \end{aligned}$$

כלומר, הפעם קיבלנו תוצאה דואלית לפיה בידיעת $\{f_k\}_{k=1}^N$ נוכל לקבוע את $\{r_k\}_{k=0}^N$ כערכי האמצע שבין רמות הייצוג.

בשל היות מערכת המשוואות הכוללת סתומה, הוצע פתרון איטרטיבי, בו מתחילים בניחוש מוצלח של רמות ההחלטה (או רמות הייצוג – זה לא משנה), ומעדכנים את רמות הייצוג לפי המשוואות הני"ל. עם הערכים המעודכנים חוזרים ומשנים את רמות ההחלטה, וחוזר חלילה.

2. קירוב קוונטיזר Max-Lloyd ע"י Companding

אחת הבעיות המרכזיות בתכנון קוונטיזר Max-Lloyd היא בעיית ההינעלות על מינימה לוקלית. להצלחת התהליך האיטרטיבי הני"ל יש קשר הדוק לבחירת האיתחול, ולשימוש בשיטות החלקה שונות. מסתבר שעבור קוונטיזרים בהם מספר המדרגות גדול דיו ניתן להשיג פתרון מקורב לבעיה הני"ל תוך הימנעות ממינימות לוקליות. גישה זו אותה נציג מיד ידועה בשם Companding בשל סיבה אותה נזכיר בהמשך. ההנחה היסודית אותה נבקש לקיים היא שגודל המדרגות כה קטן ביחס לשינויים בפונקציית הפילוג שנוכל לקבוע כי בקירוב טוב

$$\bar{E} = \int_{-\infty}^{\infty} p(x)(x - Q\{x\})^2 dx = \sum_{k=1}^N \int_{r_{k-1}}^{r_k} p(x)(x - f_k)^2 dx = \sum_{k=1}^N \int_{r_{k-1}}^{r_k} p(r_k)(x - f_k)^2 dx$$

כלומר, באינטרוול $[r_{k-1}, r_k]$ אנו מבקשים להניח כי פונקציית הפילוג בקירוב טוב קבועה ושווה לפיכך ל- $p(r_k)$ - ערך פונקציה זו בקצה אינטרוול זה (נוכל במידה דומה להניח כי פונקציית הפילוג תתייחס

לנקודת המצע של האינטרוול, דהיינו $p(0.5[r_k + r_{k-1}])$, ודבר זה נוח יותר). תחת הנחה זו נקבל כי השגיאה נתונה כ:

$$\bar{E} = \sum_{k=1}^N \int_{r_{k-1}}^{r_k} p(r_k)(x - f_k)^2 dx = \frac{1}{3} \sum_{k=1}^N p(r_k) [(r_k - f_k)^3 - (r_{k-1} - f_k)^3]$$

גזירה של ביטוי זה לפי f_k תיתן

$$\frac{\partial}{\partial f_k} \sum_{k=1}^N p(r_k) [(r_k - f_k)^3 - (r_{k-1} - f_k)^3] = p(r_k) [-(f_k - r_k)^2 + (f_k - r_{k-1})^2] = 0$$

$$\Rightarrow f_k = \frac{r_{k-1} + r_k}{2}$$

זו לא תוצאה זהה לתוצאה אותה ראינו קודם כיוון שהפעם נקבע שרמת הייצוג מתקבלת כאמצע האינטרוול $[r_{k-1}, r_k]$ (דבר מעורר תהיה אם נחשוב על קצוות הגרף! אנו נראה כי למרות צרה זו, הביטויים אליהם נגיע לא סובלים מבעיה). נוכל להציב ביטוי זה חזרה לשגיאה ולקבל כי

$$\bar{E} = \frac{1}{3} \sum_{k=1}^N p(r_k) [(r_k - f_k)^3 - (r_{k-1} - f_k)^3] = \frac{1}{12} \sum_{k=1}^N p(r_k) (r_k - r_{k-1})^3$$

וכעת עלינו להביא למינימום ביטוי זה כפונקציה של רמות ההחלטה. במקום לנקוט בגישה ישירה למינימיזציה, נציע החלפת הנעלמים. נסמן את הגודל

$$\mu_k = p(r_k)^{1/3} \cdot (r_k - r_{k-1})$$

ואז השגיאה שאותה יש להביא למינימום היא

$$\bar{E} = \frac{1}{12} \sum_{k=1}^N p(r_k) (r_k - r_{k-1})^3 = \frac{1}{12} \sum_{k=1}^N \mu_k^3$$

נשאל מהם המספרים μ_k שיביאו שגיאה זו לאפס? אלא שישנו אילוץ שערכים אלו חייבים לקיים:

$$\sum_{k=1}^N \mu_k = \sum_{k=1}^N p(r_k)^{1/3} (r_k - r_{k-1}) \cong \int_{r_0}^{r_N} p(x)^{1/3} dx = K$$

כיוון שחוק הפילוג $p(x)$ ידוע, הגודל K ניתן לקביעה מראש והוא תלוי רק בחוק הפילוג. לכן, נפתור את בעיית האופטימיזציה

$$\text{Minimize } \sum_{\mu_k} \mu_k^3 \quad \text{Subject to } \sum_{k=1}^N \mu_k = K$$

ע"י כופלי לגרנז' ונקבל

$$L\{\mu_1, \mu_1, \dots, \mu_N\} = \sum_{k=1}^N \mu_k^3 - \lambda \left(\sum_{k=1}^N \mu_k - K \right)$$

$$\frac{\partial L}{\partial \mu_j} = 3\mu_j^2 - \lambda = 0 \Rightarrow \mu_j = \sqrt{\frac{\lambda}{3}} = \frac{K}{N}$$

התוצאה מעניינת מאוד וקובעת כי להשגת אופטימום על כל ה- μ_j להיות שווים. כיוון שה- μ_j הוא בקירוב טוב השטח מתחת פונקציית ההסתברות בחזקת שליש באינטרוול $[r_{k-1}, r_k]$, פירוש הדבר הוא

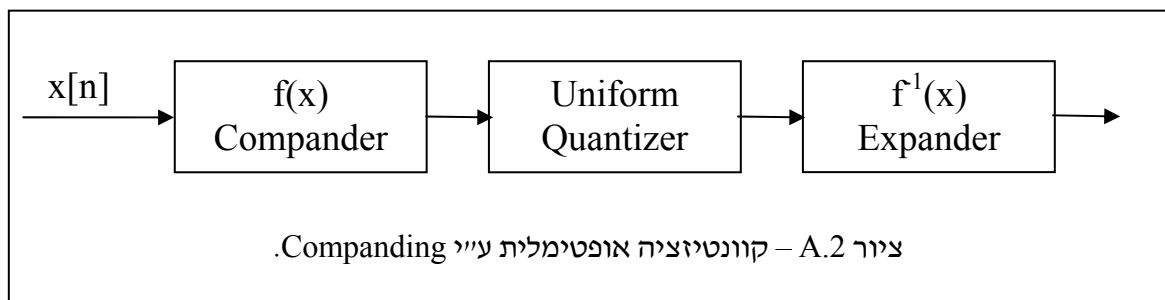
נספחים

שיש לקחת את הפונקציה $p(x)^{1/3}$ ולחלקה ל-N נתחים שווי שטח, ואלה יגדירו את מיקום רמות ההחלטה האופטימליות. בצורה סידרתית נמצא את רמות ההחלטה - נתחיל עם $r_0 = -\infty$, ונקבע את r_1 כנקודה בה צברה הפונקציה $p(x)^{1/3}$ שטח השווה ל- $1/N$ משטחה הכולל (הלא הוא K). משם נחפש את r_2 בצורה דומה וחוזר כך עד ל- r_{N-1} , כשברור כי $r_N = +\infty$. השגיאה הנובעת מקוונטיזר זה נתונה ע"י

$$\bar{E} = \frac{1}{12} \sum_{k=1}^N \mu_k^3 = \frac{1}{12} \left(\frac{K}{N} \right)^3 \cdot N = \frac{K^3}{12N^2} = \frac{K^3}{12} 2^{-2J}$$

ביטוי זה חשוב ונעשה בו שימוש תכוף – הוא נותן הערכה לשגיאת הקוונטיזציה הממוצעת כפונקציה של פונקצית ההתפלגות (חבוי ב-K) וכפונקציה של מספר הסיביות המוקצות לייצוג (ל-J סיביות יש $N = 2^J$ רמות). מביטוי זה נובע כי כל הוספת סיבית מקטינה את השגיאה הריבועית בפקטור 4.

אז למה Companding? מכיוון שישנה גישה אחרת שמובילה לאותה תוצאה בדיוק. בשנות ה-60' בהם שיקולי חומרה היו מהותיים נשאלה השאלה הבאה – האם ניתן יהיה להחליף את הקוונטיזר האופטימלי (שסביר כי יהיה בעל רוחב מדרגות משתנה) בתהליך שקול כמתואר בציור A.2, בו האות הכניסה עובר עיוות חסר זיכרון (פעולת LUT), אחריו קוונטיזר אחיד, ולבסוף עיוות הפוך?



על סמך הניתוח שהוצע לעיל למעשה קיבלנו את התשובה, לפיה ניתן בהחלט להתקרב לביצועי קוונטיזר Max-Lloyd עבור קוונטיזציה צפופה מספיק, וזאת נשיג ע"י הפונקציה

$$w = f(x) = \frac{\int_{-\infty}^x p(z)^{1/3} dz}{\int_{-\infty}^{\infty} p(z)^{1/3} dz}$$

פונקציה זו ממירה את המשתנה האקראי x (נשברנו – אין יותר ראייה דטרמיניסטית!) הקיים על פני כל ציר הממשיים במשתנה חדש w הקיים באינטרוול $[0,1]$. קוונטיזציה אחידה על פני w הוא בדיוק המירשם לקירוב קוונטיזר Max-Lloyd שקיבלנו קודם (בשל הדרישה שעל כל ה- μ_j להיות שווים).

באשר לעיוות החוזר (קראנו לו Expander), כל תכליתו היא לשנות את רמות הייצוג שייקבעו כפזורים אחיד בתחום $[0,1]$ לערכים שמשקפים את ערכי x ולכן העיוות לאחור. דרך אגב, קל לטעות ולחשוב כי העיוות הנחוץ f צריך להיות זה שימיר את המשתנה האקראי x לבעל פילוג אחיד, וכפי שרואים אין זה כך כלל!

דוגמה A.1 – נניח כי x בעל פילוג אחיד בתחום $[0,1]$. במקרה זה פונקצית העיוות המוצעת $f(x)$ היא פונקציית היחידה, דהיינו, $w=x$. לכן פילוג אחיד יקבל בשיטה זו קוונטיזציה אחידה, שהיא גם האופטימלית.

דוגמה A.2 – נניח כי x בעל פילוג מהצורה

$$p(x) = \begin{cases} 0 & x < 0 \\ 2x & x \in [0,1] \\ 0 & x > 1 \end{cases} \Rightarrow f(x) = \begin{cases} 0 & x < 0 \\ x^{4/3} & x \in [0,1] \\ 1 & x > 1 \end{cases}$$

לעומת זאת, לו רצינו להפוך את x לבעל פילוג אחיד היתה נדרשת פונקציית העיוות x^2 (חזרה על נושא יישור היסטוגרמות תוכל לרענן את זיכרוכם בדרך לקבלת תוצאה זו). כפי שאנו רואים, אלה שתי פונקציות שונות בתכלית.

3. קוונטיזר אופטימלי במונחי D-R

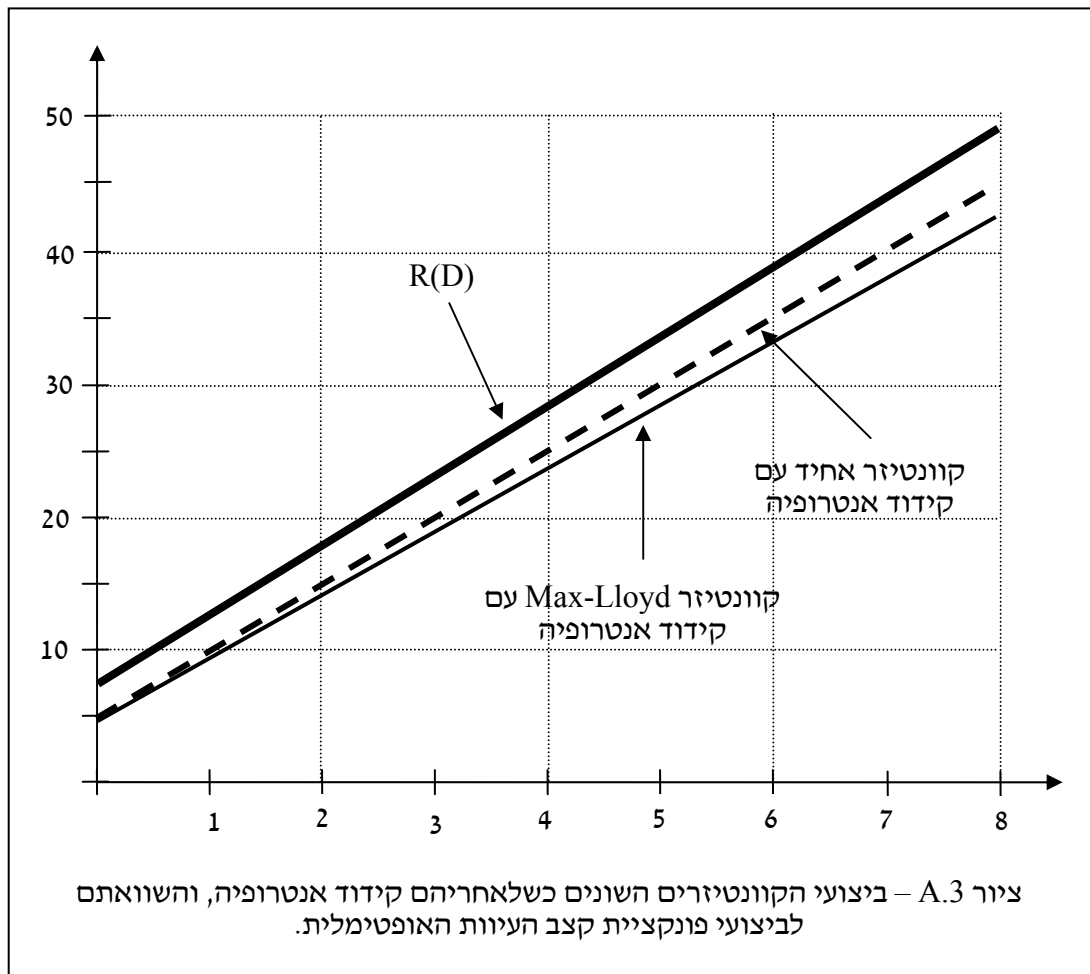
נניח כי נתון לנו משתנה אקראי x לבן (אין תלות בין דגימותיו) בעל חוק פילוג גאوسی $N\{0, \sigma\}$ ואותו רצוננו לקודד. תורת האינפורמציה מלמדת אותנו כי פונקצית קצב העיוות עבור משתנה זה נתונה ע"י

$$R(D) = \frac{1}{2} \log_2 \left(\frac{\sigma^2}{D} \right) \Leftrightarrow D(R) = \sigma^2 \cdot 2^{-2R}$$

כש- D הוא השגיאה הריבועית הממוצעת. חשוב להבהיר – חישוב פונקצית קצב העיוות בדרך כלל מורכב למדי ועל-פי רוב לא ניתן להגיע לביטוי סגור כבמקרה זה. כזכור, פונקציה זו מייצגת את הקידוד הטוב ביותר האפשרי למשתנה זה, וביטוי זה קרוי גם חסם שנון (Shannon).

נוכל כמובן לגשת לעניין קידוד המשתנה x באמצעים מעשיים – הפעלת קוונטיזציה כלשהי. כיוון שאנו חושבים במונחים של דחיסה, טבעי שננסה לסחוט הרבה ככל האפשר בייצוג המשתנה x ולכן נפעיל קידוד אנטרופיה על מוצא הקוונטיזר. כלומר, ניקח את סט רמות הייצוג כשלכל אחת סבירות הופעה, ונחשב את האנטרופיה הסקלרית מתוך הנחה שבידנו אלגוריתם שמשיג אנטרופיה זו. בפועל, כזכור, מקודד אנטרופיה כגון Huffman מתקרב כדי סיבית אחת אנטרופיה. קוונטיזר ראשון אותו נפעיל הוא הקוונטיזר האחיד, כשנשנה את גודל המדרגה עד לקבלת שגיאה ריבועית מינימלית. על התוצאה נחשב אנטרופיה (כנגד כל מספר של מדרגות) ונקבל עקום חדש של קצב-עיוות מעשי. נוכל לחזור עת התהליך כשהפעם הקוונטיזר הוא קוונטיזר Max-Lloyd, וכאן נקבל תופעה מוזרה – קוונטיזר Max-Lloyd חלש

יותר מהאחיד! תוצאות אלו מוצגות בציר A.3. השאלה המרכזית בה נרצה להתמקד היא - איך זה ייתכן?



ובכן, ביישומו של קוונטיזר Max-Lloyd ישנו כוח המתנגד לקידוד אנטרופיה. כאשר פונקציית הפילוג גבוהה יבחר בצדק הקוונטיזר Max-Lloyd לצופן מדרגות כדי להקטין את שגיאת הייצוג, בעוד שבאזורים נמוכים המדרגות תרווחנה. משמעות תוצאה זו היא שסבירות הופעתן של רמות הייצוג בקוונטיזר Max-Lloyd יהיה קרוב לאחיד, וזה אומר שקידוד אנטרופיה יהיה חסר תרומה. בניגוד לזאת, קוונטיזר אחיד ייתן שוני משמעותי בין סבירות רמות הייצוג השונות, בכך יפתח פתח לביצועים טובים לקידוד האנטרופיה.

כל זה מוביל לשאלה הבסיסית – מהו הקוונטיזר הטוב ביותר במונחי $R(D)$ בהנחה כי מבוצע קידוד אנטרופיה על רמות הייצוג? מסתבר כי ניתן לתכנן קוונטיזר אופטימלי במונחי $R-D$, אך זה אינו פשוט. עבור המקרה הגאוס, בחירת קוונטיזר אחיד אינה רחוקה כלל מהקוונטיזר האופטימלי. יתרה מזה, לא ניתן להשיג קוונטיזר סקלרי שיגיע לחסם שנון כיוון שלצורך השגת החסם נחוץ קידוד וקטורי (על אף שהאות לבן!). קיימות עבודות מתחילת שנות ה-70 שעסקו בבעיה זו וסיפקו דרכים לתכנן את הקוונטיזר האופטימלי, אך אנו נעצור את הדיון בנושא זה בשלב זה.

הערה 1: נשים לב כי חסם שנון מציע את הביטוי $D(R) = \sigma^2 \cdot 2^{-2R}$ ומתוצאה שקיבלנו עבור קוונטיזר מבוסס Companding קיבלנו $D = K^3 2^{-2R} / 12$, כשלא נלקח בחשבון קידוד אנטרופיה על מוצא הקוונטיזר. יוצא שפקטור השיפור בהוספת סיבית בשתי העקומות שווה ל-4, דבר שמעיד כי שתי עקומות אלו בעלות שיפוע זהה.

הערה 2: ממש כשם שיש מקום לשלב מקודד אנטרופיה במוצא קוונטיזר סקלרי ולתכנן את הצמד לקבלת ביצועים מיטביים במונחי R-D, יש אפשרות לדבר באופן דומה על קידוד וקטורי. לא יעלה על הדעת ששימוש באלגוריתם ה-K-Means ייתן את הביצועים הטובים ביותר מכיוון שבדומה לקוונטיזר Max-Lloyd החלוקה שהוא מציע פועלת נגד קידוד אנטרופיה יעיל.

נספח ב' – התמרת KLT כהתמרה אופטימלית והרחבותיה

בפרק 7 הצגנו את התמרת ה-KLT כהתמרה אופטימלית, והצגנו שני מובנים לפיה נבחנת אופטימליות זו. בנספח זה נציג בפירוט את ניתוח האופטימליות של התמרה זו, והרחבות מעניינות הנובעות מכך.

1. אופטימליות התמרת ה-KL

יש מגוון דרכים לבנות את התמרת ה-KLT – אנו נציע דרך סידרתית קלה יותר להבנה. נניח כי נתון לנו אוסף וקטורים $\{\underline{v}_k\} \in \mathbb{R}^N$ ורצוננו לתכנן עבורם התמרה ליניארית אשר תיוצג כמטריצה בגודל N על N הכופלת וקטורים אלה. לשם קביעת מידת הצלחת ההתמרה, נניח את התהליך הבא – בהינתן ווקטור \underline{v} מאוסף זה, תבוצע התמרה ליניארית ריבועית זו עליו עם המטריצה A לקבלת $\underline{u} = A\underline{v}$. מתוך N איברי וקטור התוצאה \underline{u} , יושארו M הראשונים ללא שינוי, וכל היתר יאופסו. וקטור חדש זה מתקבל ע"י המכפלה $\underline{w} = I_M A \underline{v} = I_M \underline{u}$, כשהמטריצה I_M היא מטריצה אלכסונית עם M '1' על אלכסונה הראשי, כשלאחריהם אפסים. את הווקטור \underline{w} נמיר חזרה ע"י מטריצה ריבועית B לקבלת \underline{z} . תפיסה זו קרויה "קירוב ליניארי" כיוון שישנו כאן קירוב של אות המקור באמצעים ליניאריים. עד כאן פעלנו לפי התיאור המופיע בפרק 7.

נניח כעת כי $M=1$, דהיינו נסתכל רק על האיבר הראשון בהתמרה כשכל היתר מאופסים. המכפלה $\underline{w} = I_M A \underline{v} = I_M \underline{u}$ תבחר למעשה את השורה הראשונה מ-A אותה נסמן כ- \underline{a}_1^T כשכל יתר השורות אינן רלוונטיות. הווקטור \underline{w} יהיה בעל איבר ראשון שונה מאפס וכל האחרים אפסים בהגדרה. לכן, בשחזור מתוך מקדם בודד זה נקבל כי רק העמודה הראשונה במטריצה B משמשת. לכן, נרצה להביא למינימום את הביטוי

$$f\{\underline{a}_1, \underline{b}_1\} = \sum_k \|\underline{v}_k - \underline{b}_1 \underline{a}_1^T \underline{v}_k\|_2^2$$

ביטוי זה מבטא את רצוננו לייצג בעזרת המקדם היחיד שנותר את המרב האפשרי מווקטור המקור. ניכר כי לא יכול להיות לביטוי זה פתרון יחיד כי אם צמד וקטורים מסוימים $[\underline{a}, \underline{b}]$ מהווים פתרון אופטימלי, אזי כך גם כל צמד $[\mu \cdot \underline{a}, \underline{b} / \mu]$. לכן, לשם יצירת יחידות פתרון נפתור הבעיה הנ"ל תוך כפיית האילוץ

$\|\underline{a}_1\|_2 = 1$. נניח לרגע כי מצאנו את \underline{a}_1 , ונשאל מהו \underline{b}_1 האופטימלי. התוצאה המתקבלת היא

$$\begin{aligned} \frac{\partial}{\partial \underline{b}_1} f\{\underline{a}_1, \underline{b}_1\} &= \frac{\partial}{\partial \underline{b}_1} \sum_k \|\underline{v}_k - \underline{b}_1 \underline{a}_1^T \underline{v}_k\|_2^2 = \\ &= \sum_k (\underline{v}_k - \underline{b}_1 \underline{a}_1^T \underline{v}_k) \underline{v}_k^T \underline{a}_1 = (\mathbf{I} - \underline{b}_1 \underline{a}_1^T) \left(\sum_k \underline{v}_k \underline{v}_k^T \right) \underline{a}_1 = 0 \end{aligned}$$

נגדיר את המטריצה R ע"י

$$R = \sum_k \mathbf{v}_k \mathbf{v}_k^T$$

מטריצה זו היא מטריצת הקווריאנס של אוסף הוקטורים. עבור מקבץ אינסופי של וקטורים אשר להם ניתן להתייחס כוקטורים אקראיים, מטריצה זו מתארת את התלויות ההדדיות בין איברי הווקטור. מתקבל,

$$(\mathbf{I} - \mathbf{b}_1 \mathbf{a}_1^T) \mathbf{R} \mathbf{a}_1 = 0 \quad \Rightarrow \quad \mathbf{b}_1 = \mathbf{R} \mathbf{a}_1 [\mathbf{a}_1^T \mathbf{R} \mathbf{a}_1]^{-1} = \frac{1}{\lambda} \mathbf{R} \mathbf{a}_1$$

כאן סימננו את הסקלר $\lambda = \mathbf{a}_1^T \mathbf{R} \mathbf{a}_1$. נציב תוצאה זו חזרה לפונקציית המחיר המקורית ונקבל כי עלינו להביא למינימום את הפונקציה

$$f\{\mathbf{a}_1\} = \sum_k \left\| \mathbf{v}_k - \frac{1}{\lambda} \mathbf{R} \mathbf{a}_1 \mathbf{a}_1^T \mathbf{v}_k \right\|_2^2$$

נניח כי הפתרון האופטימלי הוא וקטור עצמי של המטריצה R, דהיינו קיים כי $\mathbf{R} \mathbf{a}_1 = \lambda \mathbf{a}_1$. בהכפלה משמאל ב- \mathbf{a}_1^T נקבל $\mathbf{a}_1^T \mathbf{R} \mathbf{a}_1 = \lambda \mathbf{a}_1^T \mathbf{a}_1 = \lambda \|\mathbf{a}_1\|_2^2 = \lambda$, כך שהסקלר שהגדרנו קודם אינו אלא הערך העצמי שמתייחס לפתרון שנבחר. תחת הנחה זו גם נקבל כי

$$\mathbf{b}_1 = \mathbf{R} \mathbf{a}_1 [\mathbf{a}_1^T \mathbf{R} \mathbf{a}_1]^{-1} = \frac{1}{\lambda} \mathbf{R} \mathbf{a}_1 = \frac{1}{\lambda} \lambda \mathbf{a}_1 = \mathbf{a}_1$$

נוכל להציב את הקשר של היות הפתרון וקטור עצמי של R ולקבל כי פונקציית המחיר אותה יש להביא למינימום במבנה פשוט יותר,

$$\begin{aligned} f\{\mathbf{a}_1\} &= \sum_k \left\| \mathbf{v}_k - \frac{1}{\lambda} \mathbf{R} \mathbf{a}_1 \mathbf{a}_1^T \mathbf{v}_k \right\|_2^2 \stackrel{\mathbf{R} \mathbf{a}_1 = \lambda \mathbf{a}_1}{=} \sum_k \left\| \mathbf{v}_k - \mathbf{a}_1 \mathbf{a}_1^T \mathbf{v}_k \right\|_2^2 = \\ &= \sum_k \mathbf{v}_k^T (\mathbf{I} - \mathbf{a}_1 \mathbf{a}_1^T) (\mathbf{I} - \mathbf{a}_1 \mathbf{a}_1^T) \mathbf{v}_k = \\ &= \sum_k \left[\|\mathbf{v}_k\|_2^2 - 2 \mathbf{v}_k^T \mathbf{a}_1 \mathbf{a}_1^T \mathbf{v}_k + \mathbf{v}_k^T \mathbf{a}_1 \mathbf{a}_1^T \mathbf{a}_1 \mathbf{a}_1^T \mathbf{v}_k \right] = \sum_k \left[\|\mathbf{v}_k\|_2^2 - |\mathbf{a}_1^T \mathbf{v}_k|^2 \right] \end{aligned}$$

בפיתוח הנ"ל השתמשנו בעובדה ש- $\|\mathbf{a}_1\|_2 = 1$. מביטוי זה נובע כי מציאת מינימום לפונקציית מחיר זו שקולה למציאת מקסימום לפונקציית המחיר

$$g\{\mathbf{a}_1\} = \sum_k |\mathbf{a}_1^T \mathbf{v}_k|^2 = \mathbf{a}_1^T \left[\sum_k \mathbf{v}_k \mathbf{v}_k^T \right] \mathbf{a}_1 = \mathbf{a}_1^T \mathbf{R} \mathbf{a}_1$$

המקסימום של ביטוי זה (בכפוף לאילוץ $\|\mathbf{a}_1\|_2 = 1$) הינו הווקטור העצמי המתייחס לערך העצמי הגדול ביותר של המטריצה R, וכך הנחה שעשינו לפני מספר שלבים על אופיו של הפתרון מתאמת.

נניח כי עשינו כל זאת וקיבלנו את השורה הראשונה ב- A (והעמודה הראשונה ב-B). נוכל לחזור על התהליך ולחפש את השורה השנייה ע"י הבאה למינימום של פונקציית המחיר

$$f\{\mathbf{a}_2, \mathbf{b}_2\} = \sum_k \left\| \mathbf{v}_k - (\mathbf{b}_2 \mathbf{a}_2^T + \mathbf{b}_1 \mathbf{a}_1^T) \mathbf{v}_k \right\|_2^2 = \sum_k \left\| \mathbf{v}_k - \mathbf{b}_2 \mathbf{a}_2^T \mathbf{v}_k - \mathbf{b}_1 \mathbf{a}_1^T \mathbf{v}_k \right\|_2^2$$

כשהפעם אנו מניחים כי הצמד $[a_1, b_1]$ נתון. תהליך אלגברי דומה לזה שתואר לעיל יביא לבחירת a_2 כווקטור העצמי המתייחס לערך העצמי השני של המטריצה R , וכך תהליך זה נמשך. נקודה מעניינת היא שנדמה כי הפתרון שנרקם כאן יהיה תת-אופטימלי בשל היות התהליך סדרתי – למעשה אין זה כך - אם תיעשה האופטימיזציה בצורה מקבילה על כל המטריצה A (ו- B) נגיע לאותו פתרון ממש.

נחזור לרגע לתוצאה שקיבלנו במהלך הניתוח ושנראתה כפרט טכני. מצאנו כי כתחליף לבעיית מינימיזציה מורכבת יש לנו בעיית מקסימיזציה בה פונקציית המחיר היא

$$g\{a_1\} = \sum_k |a_1^T v_k|^2 = a_1^T \left[\sum_k v_k v_k^T \right] a_1 = a_1^T R a_1$$

לביטוי זה משמעות גיאומטרית מעניינת – בהינתן מקבץ הדוגמאות $\{v_k\}_k \in \mathbb{R}^N$ אנו מחפשים כיוון הטלה a_1 עליו יתקבל כי הפיזור הוא מרבי! ציור B.1 ממחיש בחירה זו בדו-מימד.



תכונה נוספת מעניינת היא התכונה הבאה: בהינתן וקטור כלשהו \underline{v} , נוכל לייצג אותו כצירוף ליניארי של וקטורים עצמיים אלה ולכתוב

$$\underline{v} = \sum_{n=1}^N v_n a_n$$

ההתמרה ע"י A פירושה הכפלה במטריצה ששורותיה הן a_n^T , ובשל היות הוקטורים העצמיים אורתונורמליים, הכפלה זו תיתן את אוסף המקדמים v_n , וזו תהיה תוצאת ההתמרה $\underline{u} = A \underline{v}$. אם נחשב כעת את מטריצת הקוריאנס של מוצא ההתמרה נקבל

$$R_U = \sum_k \underline{u}_k \underline{u}_k^T = A \sum_k \underline{v}_k \underline{v}_k^T A^T = A R A^T = \Lambda$$

כשהמטריצה Λ היא מטריצה אלכסונית המאוכלסת בערכים העצמיים של R . יצא כי פעולת ההתמרה הלבנה את המקדמים – יצרה וקטורי ייצוג חדשים בהם אין קורלציה בין איברים שונים.

2. מ-PCA ל-ICA

התמרת ה-KLT ידועה לסטטיסטיקאים בשם Principle Component Analysis – PCA, כששם זה בא לתאר בדיוק את שראינו קודם – אנו מוצאים את הכיוונים העיקריים של הנתונים הוקטורים בבניית התמרה זו. יש חסרונות רבים להתמרה זו, אך יתרונה הגדול של התמרה זו בפשטות החישובית ובניקיון המתמטי המוביל לפתרון סגור – עובדות המסבירות את הפופולאריות הרבה של שיטה זו. בעשור השנים האחרונות נעשו ניסיונות שונים להכליל את ה-PCA. המשותף לכל הכללות אלו הוא אובדן הפשטות!

הכללה ראשונה שנציג היא ניסיון למנוע מחריגים להשפיע על בניית ההתמרה. אם נתייחס לבניית כיוון ההטלה הראשון, הרי שהביטוי המקורי המובא למקסימום הוא

$$g\{\underline{a}\} = \sum_k \left| \underline{a}^T \underline{v}_k \right|^2$$

ועבור וקטור בעייתי הנותן ערך חריג יוסט כיוון ההטלה. כתחליף, שימוש בפונקציה

$$g\{\underline{a}\} = \sum_k \left[h\left(\underline{a}^T \underline{v}_k\right) \right]^2$$

כאשר $h(x)$ מרכז ערכים גדולים תוכל בהחלט לפתור ולו חלקית בעיה זו. לדוגמה, שימוש בפונקציה

$$h(x) = \begin{cases} T & x > T \\ x & |x| \leq T \\ -T & x < -T \end{cases}$$

תשרת מטרה זו. כמובן שאז צפה ועולה בעיית המקסימיזציה של הפונקציה החדשה, שהרי אין זו עדיין בעיית ערכים עצמיים. הפתרון המסתמן בעבודות רבות היא גישת הגרדיאנט הסטוכסטי בה למעשה

מעדכנים את \underline{a} לפי כל וקטור \underline{v}_k בנפרד ע"י הגרדיאנט של הפונקציה $h\left(\underline{a}^T \underline{v}_k\right)^2$. מתקבל

$$\underline{a}_{k+1} = \underline{a}_k + \alpha \left. \frac{\partial h\left(\underline{a}^T \underline{v}_k\right)^2}{\partial \underline{a}} \right|_{\underline{a}_k} = \underline{a}_k + \alpha h\left(\underline{a}_k \underline{v}_k\right) \underline{v}_k$$

השיטה נקראית גרדיאנט סטוכסטי כיוון שההתייחסות לאוסף הוקטורים היא כסט מדידות מצטבר לפונקציית מחיר שבכל שלב נתון אינה ידועה, ורק בצבירת אינסוף וקטורים יש לה משמעות דטרמיניסטית.

גישה דומה מאוד ניתן להציע בה תקבל אינטרפרטציה אחרת לגמרי – אם נניח כי במקום

שימוש בפונקציה h יחידה נציע פונקציית מחיר מהצורה

$$g\{\underline{a}\} = \sum_k [h_1(\underline{a}^T \underline{v}_k)]^p + \sum_k [h_2(\underline{a}^T \underline{v}_k)]^p + \dots$$

נקבל תוצאה מעניינת. לדוגמה, עבור $h_1(x) = x$ המופעלת לבדה ראינו כי קיבלנו הלבנת מוצא ההתמרה. הלבנה זו פירושה ביטול הקורלציה בין מקדמי ההתמרה. בוקטורים אקראיים גאוסיים חוסר קורלציה שקול לחוסר תלות סטטיסטית (שני משתנים בלתי תלויים סטטיסטית אם $p(x)p(y)=p(x,y)$). בכל משתנה אקראי אחר אין זה כך. לכן, אם היחד הוא הורדת תלות סטטיסטית ולא רק ביטול קורלציה, הרי שהתהליך מורכב יותר לאין ערוך ושימוש מספר פונקציות h יכול להיות חיקוי לא רע כלל להשגת יעד חדש זה. קיימת התכונה הבאה לשני משתנים בלתי תלויים סטטיסטית:

$$\forall h, E\{h(x)h(y)\} = E\{h(x)\}E\{h(y)\}$$

עד כה התרכזנו בביטול הקורלציה ע"י שימוש ב- $h_1(x) = x$ (שהרי משתנים חסרי קורלציה מאופיינים בתכונה $E\{xy\} = E\{x\}E\{y\}$). לכן, אם נבנה פונקצית מחיר כמוצע, נתקרב יותר לחוסר תלות סטטיסטית. האלגוריתם במקרה זה גם הוא ישען על גרדיאנט סטוכסטי, אך יכלול פשוט יותר איברים בעדכון התוצאה.

אם אמנם רצוננו בביטול תלות סטטיסטית, הגישה הישירה ליעד זה היא שימוש במושג האנטרופיה המשותפת. אנו לא נציג נושא זה כאן, והספרות בתחום זה ענפה.

נספח ג' – הקצאת סיביות אופטימלית בדחיסה

נושא יפה שעולה בדחיסת אותות הוא נושא הקצאת סיביות אופטימלית. הרעיון הבסיסי הוא חלוקה חכמה של תקציב סיביות נתון לשם קבלת שגיאה מינימלית. מעניין לציין כי למרות יופיו של רעיון זה, הוא אינו מיושם בסטנדרט ה-JPEG שם יכול היה להתאים. בפועל מנגנונים אחרים בסטנדרט זה נותנים אפקט דומה. בנספח זה נתאר את הבעיה ואת דרך פתרונה.

1. הצגת הבעיה – דחיסה מבוססת התמרה

שיטת דחיסה יעילה בתמונות מציעה חלוקת התמונה לבלוקים בגודל קבוע (נניח 8 על 8), וביצוע התמרה על כל בלוק. רצויה התמרה קרובה ל-KLT על מנת לאפשר קיצוץ מקדמים עם נזק מינימלי למידע הבלוק. ראינו בפרק 7 כי שימוש ב-DCT מקרב היטב את ה-KLT, ומביא עימו יתרון חשוב – קיומו של אלגוריתם מהיר וקבוע (כלומר ללא צורך לשדר את תכונות ההתמרה) ליישום ההתמרה. אנו נניח כי בטרם בוצעה התמרת ה-DCT לכל בלוק, הוחסר ממוצע הבלוק וקודד בנפרד (כפי שבפועל יש לעשות). פירוש הדבר שבכל בלוק יש 63 מקדמים (כי מקדם ה-DC נוטרל).

כעת בידנו N בלוקים כשבכל אחד 63 ערכי התמרה, ורצוננו בהקצאה חכמה של תקציב הסיביות בין מקדמי ההתמרה לקבלת שגיאה מינימלית. אנו מניחים כי חוקיות הקצאת הסיביות תהיה זהה מבלוק לבלוק, דהיינו, מקדם [3,4] בכל הבלוקים יקבל כמות סיביות זהה. הנחה זו הגיונית למדי כיוון שמקדם במיקום זהה בבלוק הינו בעל חשיבות דומה. טיפול חכם יותר יניח כי תמונה מכילה בלוקים ממקורות שונים לגמרי (בלוקים הלקוחים ממרקס, אזורי שפה, אזורים חלקים, אזורי רעש) ונכון יהיה לבצע הקצאת סיביות שונה למקדמי ההתמרה לכל סוג בלוק בנפרד. במקרה כזה רצוי לבצע אישכול (Clustering) על בלוקי ההתמרה, ואז לתכנן את הקצאת הסיביות לכל קבוצה בנפרד.

נחזור למקרה בו בידנו קבוצת בלוקים עליה נתכנן הקצאת סיביות קבועה. כאמור, בידנו תקציב סיביות, ועל כל קביעה של חלוקתן בין המקדמים יש לדרוש שסך הסיביות מתאים לתקציב ממנו יצאנו – דבר זה יהווה אילוץ בבעיית האופטימיזציה אותה אנו עומדים להגדיר. בהינתן כמות סיביות מסויימת למקדם התמרה, ובידיעה של פילוגו (שהרי יש לנו N מקדמים כאלו וניתן להתייחס להיסטוגרמה שלהם, או לפיזורם), נוכל לקבוע את השגיאה הצפויה בקוונטיזציה סקלרית של מספרים אלו. לצורך זה נוכל לעשות שימוש בקשרים אותם קיבלנו בנספח א' למשל. לכן, נוכל לחשב את השגיאה הכוללת ולהביאה למינימום בכפוף לאילוץ שהזכרנו. מטרת תהליך זה היא לקבוע 63 מספרי קסם של כמות סיביות לכל מקדם התמרה, ואלו ינבעו כפתרון לבעיית האופטימיזציה. נציג כעת את הרכב הבעיה ופתרונה בצורה מתמטית מסודרת.

נתון לנו סט של וקטורים אקראיים - $\{\mathbf{x}_k\}_{k=1}^L$. כל וקטור כולל N איברים חסרי קורלציה ביניהם המפולגים גאוסיאנית עם ממוצע אפס ושונות σ_k (במילים אחרות, בדוגמה שהוזכרה קודם של קידוד תמונה, יש לנו $L=63$ קבוצות ובכל קבוצה N מספרים סקלאריים בעלי שונות ידועה אשר מייצגים את אותו מקדם התמרה בבלוקים שונים). נתונים לנו B סיביות לקידוד כלל כל NL המספרים הנתונים. נסמן ב- $\{n_k\}_{k=1}^L$ את כמות הסיביות שתוקצה למקדם ה- j . לכן, האילוץ שיש על אוסף הנעלמים לקיים הוא

$$N \cdot \sum_{k=1}^L n_k = B$$

כלומר אילוץ ליניארי פשוט שדורש שסך הסיביות יתאים לתקציב הסיביות הכולל.

אנו נניח כי שגיאת הקוונטיזציה (שגיאה ריבועית ממוצעת) של משתנה אקראי גאוסני בעל ממוצע

אפס ושונות σ_k אשר קודד ב- n_k סיביות נתונה ע"י הביטוי

$$E^2 = \text{Const} \cdot \sigma_k^2 \cdot 2^{-2n_k}$$

לכן, השגיאה הכוללת בבלוק תהיה

$$E_T^2 = \sum_{k=1}^L \text{Const} \cdot \sigma_k^2 \cdot 2^{-2n_k}$$

הבעיה שנרצה לפתור היא לפתור היא לפיכך

$$\text{Min}_{\{n_k\}_k} \sum_{k=1}^L \text{Const} \cdot \sigma_k^2 \cdot 2^{-2n_k} \quad \text{Subject to:} \quad N \cdot \sum_{k=1}^L n_k = B$$

פתרון בעיה זו בעזרת כופלי לגרנז' נותן

$$L\{n_1, n_2, \dots, n_L\} = \sum_{k=1}^L \text{Const} \cdot \sigma_k^2 \cdot 2^{-2n_k} + \lambda \left(\sum_{k=1}^L n_k - \frac{B}{N} \right)$$

$$\Rightarrow \frac{\partial L}{\partial n_k} = -2 \ln 2 \cdot \text{Const} \cdot \sigma_k^2 \cdot 2^{-2n_k} + \lambda = 0$$

$$\Rightarrow n_k = \frac{1}{2} \cdot \log_2 \left(\frac{2 \ln 2 \cdot \text{Const} \cdot \sigma_k^2}{\lambda} \right)$$

קיבלנו כי הקצאת הסיביות תגדל עם הגדלת השונות – דבר צפוי למדי. הצבה חזרה לאילוץ נתון

$$\frac{B}{N} = \frac{1}{2} \cdot \sum_{k=1}^L \log_2 \left(\frac{2 \ln 2 \cdot \text{Const} \cdot \sigma_k^2}{\lambda} \right) = L \cdot \log_2 \left(\frac{2 \ln 2 \cdot \text{Const}}{\lambda} \right) + \log_2 (\sigma_1^2 \cdot \sigma_2^2 \cdot \dots \cdot \sigma_L^2)$$

$$\Rightarrow \lambda = 2 \ln 2 \cdot \text{Const} \cdot 2^{\frac{1}{L} \log_2 (\sigma_1^2 \cdot \sigma_2^2 \cdot \dots \cdot \sigma_L^2) - \frac{B}{NL}}$$

והצבת תוצאה זו חזרה בביטוי של הקצאת הסיביות נותנת

$$\begin{aligned}
 n_k &= \frac{1}{2} \cdot \log_2 \left(\frac{2 \ln 2 \cdot \text{Const} \cdot \sigma_k^2}{\lambda} \right) = \frac{1}{2} \cdot \log_2 \left(\frac{\sigma_k^2}{2^{\frac{1}{L} \log_2 (\sigma_1^2 \cdot \sigma_2^2 \cdot \dots \cdot \sigma_L^2)} \cdot \frac{B}{NL}} \right) = \\
 &= \frac{1}{2} \cdot \log_2 (\sigma_k^2) - \frac{1}{2L} \log_2 (\sigma_1^2 \cdot \sigma_2^2 \cdot \dots \cdot \sigma_L^2) + \frac{B}{NL} = \\
 &= \frac{1}{2} \cdot \log_2 \left(\frac{\sigma_k^2}{\sqrt[L]{\sigma_1^2 \cdot \sigma_2^2 \cdot \dots \cdot \sigma_L^2}} \right) + \frac{B}{NL}.
 \end{aligned}$$

התוצאה שהתקבלה מעניינת ובעלת הרכב אינטואיטיבי. הגורם B/NL אינו אלא כמות הסיביות הממוצעת לכל מקדם – יש B סיביות שמתחלקות בין NL מקדמים. והיה וכל השונויות שוות, נצפה כי חלוקה אחידה זו היא שתתקבל. ואמנם, הגורם הראשון קובע את משקלה היחסי של השונות ה- k ית מול הממוצע הגיאומטרי של כל השונויות. אם כל השונויות זהות גורם זה יהיה 0, וההקצאה תהיה אחידה. לכל מבנה אחר נקבל הקצאה דינמית המעדיפה מקדמים גבוהי שונות, כצפוי.

בעיה טורדנית אחת עם תוצאה זו היא העובדה שלא כפינו על הסידרה לתת מספרים אי-שליליים. בהחלט יתכן מצב בו תתקבל הקצאה שלילית עבור מקדם התמרה בעל פיזור קטן מאוד ביחס לאחרים. במקרה כזה, במקום לשלב אילוף חיוביות לתוך כופלי הלגרנז' ניתן פשוט לפתור פעם ראשונה כמתואר, ופעם שנייה כשאנו משמיטים את המקדמים שזכו להקצאה שלילית, וקובעים את הקצאתם לאפס. על היתר נבצע הקצאה אופטימלית כבנוסחא הני"ל עם L קטן יותר (כי חלק מהמקדמים הוצאו). ייתכן כי נידרש למספר סבבי פתרון בצורה זו. לב העניין הינו פשוט קביעת סף על השונויות לפיו כל מקדם בעל שונות קטנה מסף מסויים מקבל 0 סיביות ומושמט מהתהליך החישובי. בעיה שניה בתהליך שהוצע כאן היא הצורך בערכי הקצאה שלמים. בפועל, תוצאת ההקצאה עוברת העגלה, ואז נבדק האילוף. במקרה של סטייה גדולה מידי ניתן לחזור על התהליך החישובי עם B שונה לשם קבלת תוצאה מתוקנת.

מה תהיה השגיאה המינימלית שתושג? נציב את נוסחאת ההקצאה האופטימלית לתוך ביטוי השגיאה ונקבל

$$\begin{aligned}
 E_T^2 &= \sum_{k=1}^L \text{Const} \cdot \sigma_k^2 \cdot 2^{-2n_k} = \text{Const} \cdot 2^{-2B/NL} \cdot \sum_{k=1}^L \sqrt[L]{\sigma_1^2 \cdot \sigma_2^2 \cdot \dots \cdot \sigma_L^2} \\
 &= \sqrt[L]{\sigma_1^2 \cdot \sigma_2^2 \cdot \dots \cdot \sigma_L^2} \cdot L \cdot \text{Const} \cdot 2^{-2B/NL}
 \end{aligned}$$

אנו רואים מספר תופעות בביטוי זה :

- א. סך השגיאה פרופורציוני לממוצע הגאומטרי של שונויות המקדמים.
- ב. סך השגיאה פרופורציוני באופן אקספוננציאלי הפוך לתקציב הסיביות למקדם בודד.
- ג. אם נניח כי כל השונויות שוות נקבל כי השגיאה מתלכדת עם הנוסחא המקורית של שגיאת קוונטיזציה לסקלר מקודד.

כנקודה אחרונה חשוב להבהיר כי הקוונטיזציה שתיעשה תפעל בהנחת ממוצע אפס על המשתנים האקראיים שאם לא כן, נידרש להעביר את הממוצע לצורך הפענוח. כמו כן, סידרת השונויות חייבת להיות מקודדת אף היא על מנת לאפשר קביעת רמות הייצוג במקלט.

נספח ד' – נושאים מתקדמים בשחזור תמונות

בנספח זה נזכיר מספר נושאים מעניינים בשחזור תמונות שלא נדונו בפרק 6. נתחיל עם שחזור מבוסס אנטרופיה מרבית אשר מהווה אלטרנטיבה לגישת ה-ML וה- MAP. גישה זו יעילה ופופולארית בטיפול בתמונות בחקר אסטרונומי בהן הרקע כהה מאוד ואחיד. נושא שני שייסקר הוא נושא השיחזור העיוור בו פונקציית הטשטוש אינה ידועה. התוצאות בתחום זה רבות ורובן מורכבות למדי (וקשורות באורח פלא ל-ICA שהוזכר בנספח ב'). אנו נתמקד ביסודות בלבד ונראה גישות שונות לתקיפת הבעיה. נושא שלישי שנזכיר הוא בעיית ה-Denoising – הורדת רעש מתמונה. אנו לכאורה יודעים הכל על נושא זה בהיותו מקרה פרטי של בעיית השחזור, אך נראה כי עם פישוט הבעיה ושימוש בהתמרת Wavelet נוצר מרווח מעניין לתוצאות מרתקות. נושא אחרון בו ניגע בקצרה הוא שילוב מעניין בין שחזור תמונות (או למעשה פתרון בעיות היפוך) ובין למידה חישובית. שילוב זה צומח מבניית Prior מבוסס דוגמאות.

1. שחזור מבוסס אנטרופיה מינימלית

אנו ראינו כבר דרכים לפתרון בעיית השיחזור של התמונה \underline{X} ממדידתה \underline{Y} הנתונה ע"י $\underline{Y} = H\underline{X} + \underline{N}$, כשאנו מניחים כי H מטריצה הפועלת כטשטוש התמונה ו- \underline{N} הוא רעש אקראי מפולג גאוסיאנית עם ממוצע אפס ומטריצת קווריאנס $\sigma_n^2 I$ (דהיינו, רעש לבן). ראינו בפרט כי הצלחה במשימת השחזור תלויה בבחירה נאותה של רגולריזציה ובלעדיה (שימוש ב-ML) יניב תוצאות עלובות.

בצילום אסטרונומי של גורמי שמיים התמונות הנרכשות בטלסקופ סובלות מהפרעות טשטוש ורעש ולפיכך הקשר $\underline{Y} = H\underline{X} + \underline{N}$ מתאים היטב. לעומת זאת, טיפול בתמונות אלו ע"י כלי הרגולריזציה שהוצעו בהם כפינו חלקות אינו מוצלח כיוון שבתמונות כאלו המבנה בדרך כלל יכול רקע כהה מאוד אשר מאכלס את רוב שיטחה של התמונה (רמות אפור קרובות לאפס), ושיאים בהירים קטני שטח המייצגים גורמי שמיים. ישנה סבירות רבה שטיפול קונוונציונלי שמתאים לתמונות בעלות מבנה Piece-wise-smooth מהסוג שמקובל בצילום דיגיטלי ביתי יביא למחיקת כוכבים בתמונה, וזה כמובן אינו רצוי. לכן עלינו להחליף את אסטרגיית הרגולריזציה לתמונות אלו.

אם ניקח את הוקטור הנעלם \underline{X} ונתייחס אליו לרגע כפונקציית הסתברות (אחרי נירמול ע"י חלוקה בסכום איבריו), נוכל לטעון שתי טענות חשובות אשר להן קשר להצלחת תהליך השחזור:

א. כמו כל פונקציית הסתברות, כל ערכי הפיקסלים חייבים להיות אי-שליליים! תכונה זו מתאימה לרצוננו לשחזר תמונה עם ערכי רמות אפור אי-שליליים. נוכל כמובן לחזור לכלי ה-ML וה- MAP ולהוסיף תכונה זו כאילוץ בבעיית האופטימיזציה שמובילה לשחזור. כפיית אילוץ זה קלה – בתהליך איטרטיבי מקובל לאחר כל עדכון לאפס את הערכים השליליים. בהתייחסות לתמונה כפונקציית הסתברות אנו פטורים מאילוץ זה – הוא משולב באופן טבעי כפי שנראה. במקרה של תמונות בהן הרקע קרוב לאפס, כפיית אילוץ זה תורמת רבות לאיכות התוצאה כיוון שאפילו ML אשר ייתן תופעות צלזולים יאלץ להעלימן אם מדובר בשפה בין רקע כהה מאוד שאינו אפס ובין שיא בהיר.

ב. ההתייחסות ל- \underline{X} כפונקציית הסתברות כמוה כמו לומר – מה ההסתברות למצוא בפיסקל מסויים כוכב – ערך בהיר פירושו הסתברות גבוהה, ורוב הרקע מתייחס להסתברות קלושה לקיום כוכב. על חוק פילוג זה שהוגדר נוכל לומר כי אנו מצפים שיהיה בעל אנטרופיה נמוכה מאוד. כזכור מפרק 8, אנטרופיה תגיע לשיאה עבור פילוג אחיד ותקטן ככל שחוק הפילוג מתקרב להתנהגות דטרמיניסטית. קיומם של מספר שיאים בודד בתמונה ייתן אנטרופיה מאוד נמוכה וזו תועדף על מספר שיאים זהה אך עם מריחה.

נסח אם כך את בעיית השחזור שבידנו כבעיית מינימיזציה של אנטרופיה בכפוף לאילוץ הנובע מהמידות.

$$\text{Min}_{\underline{X}} - \sum_{k,j} x[k,j] \cdot \ln x[k,j] \quad \text{Subject to: } \|\underline{HX} - \underline{Y}\|_2^2 = \sigma_2^2$$

או ביטוי מקביל בו האילוץ מוצג כתוספת לפונקציית המחיר עם מקדם 0 שהינו כופל לגרנז' של האילוץ)

$$\text{Min}_{\underline{X}} - \sum_{k,j} x[k,j] \cdot \ln x[k,j] + \frac{\lambda}{2} \|\underline{HX} - \underline{Y}\|_2^2$$

נשים לב לתכונה מעניינת שתסייע לנו – עניין נרמול \underline{X} . מקובל להניח כי אופרטור הטשטוש משמר ממוצע, כלומר סכום איברי הגרעין המטשטש הוא 1. לכן, בהינתן ממוצע התמונה \underline{X} יהיה שווה בקירוב טוב לממוצע התמונה \underline{Y} והרי זה ערך ידוע. לכן, אם נתחיל את התהליך ע"י חלוקת \underline{Y} בסכום איבריה נביא אותה גם כן למבנה של חוק הסתברות, ומכאן ואילך נוכל לכפות על \underline{X} להיות גם היא מנורמלת. מציאת המינימום תיעשה ע"י גזירת הפונקציה. גזירת ביטוי האנטרופיה לפי איבר יחיד ייתן

$$(x \log x)' = \log x + 1$$

$$\frac{\partial}{\partial X} \left[\sum_{k,j} -x[k,j] \cdot \ln x[k,j] + \frac{\lambda}{2} \|\underline{HX} - \underline{Y}\|_2^2 \right] = -\underline{1} - \log \underline{X} + \lambda H^T (\underline{HX} - \underline{Y}) = 0$$

משוואה זו סתומה וקשה לפתרון בדרך ישירה, אך נוכל הציע תהליכים איטרטיביים שונים לבניית הפתרון. לדוגמה, שימוש באלגוריתם ה-SD יביא לאיטרציה מהצורה

$$\underline{X}_{n+1} = \underline{X}_n - \mu \left[-\underline{1} - \log \underline{X}_n + \lambda H^T (\underline{HX}_n - \underline{Y}) \right]$$

גישה אחרת קרויה שיטת ההחלפה. שיטה נומרית קלאסית לפתור משוואות סתומות מהצורה

$$\underline{X} - \Phi(\underline{X}) = 0 \quad \underline{X}_{n+1} = \Phi(\underline{X}_n) \quad \text{ברוח זו נוכל להציע}$$

$$-\underline{1} - \log \underline{X}_{n+1} + \lambda H^T (\underline{HX}_n - \underline{Y}) = 0 \Rightarrow \underline{X}_{n+1} = \exp \left\{ \lambda H^T (\underline{HX}_n - \underline{Y}) - \underline{1} \right\}$$

הרעיון במשוואה זו הוא שאם נוכל להבטיח שתתקבל התכנסות של סידרת הפתרונות, נקבל כעבור מספר רב של איטרציות כי $\underline{X}_{n+1} = \underline{X}_n$ ואז יתקבל פתרון שמקיים את איפוס הגרדיאנט כרצוי.

מעניין לציין כי אי-שליליות הפתרון תנבע ישירות מפעולת האיטרציה, ולכן לא נדרש לכפותה באמצעים נוספים. לעומת זאת, ייתכן כי סכום איברי \underline{X} לאחר האיטרציה לא יהיה 1, ולכן לאחר כל איטרציה יש לנרמל את הוקטור.

לכל אורכו של פרק 6 הנחנו כי במודל הקלקול $\underline{Y} = H\underline{X} + \underline{N}$ אנו יודעים במדויק את פונקציית הטשטוש. באופן מעשי זה לרוב לא נכון, ועם קביעה שרירותית של המטריצה H אנו יכולים לגרום לשגיאות שחזור קשות. קיים היקף גדול של עבודות תחת השם "שחזור עיוור" שבאים להוריד הנחה זו, ומוצעות דרכים שונות (ומשונות, חייבים להודות) לשחזור בתנאים קשים אלו.

עבודות ראשונות בתחום בחרו לנסות ולקבוע את מידת הטשטוש ע"י מציאת מקומות בתמונה בהן יש נקודה בהירה על רקע כהה (או להיפך). במקרה של קיום מין "פונקציית דלתא" כזו, ההתנהגות באזורה תהיה המסנן המטשטש. באין נקודה בודדת ניתן לחפש שפה חדה ולעשות זאת במספר שפות ובזוויות שונות וכך לשחזר את הטשטוש בטרם מתחיל תהליך השחזור. שיטות אלו חלשות למדי בשל תלותן במאפייני תמונה שאינם זמינים בהכרח, בשל הפרעה שהרעש יוצר, ובשל הגבר השגיאות שנובע מהפרדת הבעיה לשתי בעיות – קביעת הטשטוש ולאחריה שחזור התמונה.

גישות מודרניות יותר ראו שיש לשלב את קביעת הטשטוש כחלק בתהליך השחזור. הוצעו מגוון כלים כאלה, כשהמובילים בהם הן הגישות הבאות:

1. אלגוריתם ה-EM – זוהי הרחבה של משעריך הסבירות המירבית (ML) לתנאים בהם חסר מידע בבעיית השערוך. המידע החסר אצלנו הוא הטשטוש ולכן מבנה זה טבעי לשימוש. אלגוריתם ה-EM מציע תהליך איטרטיבי של פתרון ה-ML שבו כל איטרציה בנויה משני נתחים – הראשון בו נקבע המידע החסר ע"י ממוצע (Expectation) של פתרון ה-ML, והשני בו מבוצע ה-ML עם הנחת ידיעת המידע החסר משלב קודם.
2. שיטות מבוססות ICA – ניתן להציג את בעיית השחזור העיוור כפירוק של עירוב ליניארי של אותות בלתי-תלויים-סטטיסטית. זוהי בעיית ה-ICA, ופתרונות עבורה קיימים אם כי אינם פשוטים.
3. שיטות מבוססות הנחות מודל על הטשטוש – ממש כשם שיכולנו לקבוע פונקציית הסתברות על התמונה הלא נודעת ולהיעזר בזה בבניית הפתרון, ניתן להציע גישה דומה לטובת הטשטוש. אם ידוע כי גרעין הטשטוש מהווה פונקציה חלקה, או אם ידוע כי הוא קיים לאורך קו (בשל טשטוש Motion Blur), מידע זה יכול לשמש בבניית השחזור.

גישה פשוטה למדי מהמשפחה השלישית (הנחת מודל על הטשטוש) שהוצעה ע"י You ו-Kave נותנת שילוב של מציאת הטשטוש תוך כדי מציאת התמונה המשוחזרת. אם נשתמש במונחים מפרק 6, לו ידענו את H היינו אמורים להביא למינימום את הפונקציה

$$\varepsilon\{\underline{X}\} = \|\underline{Y} - H\underline{X}\|_2^2 + \lambda\rho\{D\underline{X}\}$$

אם נניח כי אופרטור הטשטוש קבוע במקום ובעל מבנה חלק (כגון גאוסייאן), נוכל להתייחס לפונקציה הבאה כפונקציית המחיר למינימיזציה

$$\tilde{\varepsilon}\{\underline{X}, \underline{h}\} = \|\underline{Y} - H(\underline{h}) \cdot \underline{X}\|_2^2 + \lambda\rho\{D\underline{X}\} + \gamma\rho\{D\underline{h}\}$$

בביטוי זה אנו כופים על הפתרון (\underline{h}) להיות חלק ממש כמו תמונה, כשהמקדם למרכיב מחיר זה צריך להיבחר בקפידה כדי למשקל נכון את חשיבות חלקות זו. הסימון $H(\underline{h})$ אינו אלא בניית מטריצת

הטשטוש מתוך גרעין הטשטוש בהנחה של אופרטור קבוע במקום. גזירה של פונקציה זו ביחס לנעלמים תיתן

$$\frac{\partial \tilde{\varepsilon}\{\underline{X}, \underline{h}\}}{\partial \underline{X}} = \mathbf{H}^T [\mathbf{H}\underline{X} - \underline{Y}] + \lambda \mathbf{D}^T \rho'\{\mathbf{D}\underline{X}\}$$

$$\frac{\partial \tilde{\varepsilon}\{\underline{X}, \underline{h}\}}{\partial \underline{h}} = \mathbf{X}_M^T [\mathbf{X}_M \underline{h} - \underline{Y}] + \gamma \mathbf{D}^T \rho'\{\mathbf{D}\underline{h}\}$$

כיוון שישנה קונבולוציה בין \underline{X} ל- \underline{h} , ניתן להחליף את תפקידם במכפלה $\mathbf{H}\underline{X}$ ולבנות מטריצה מ- \underline{X} שתכפול את הוקטור \underline{h} – זה בדיוק מה שעומד מאחורי הסימון \mathbf{X}_M .

פתרון נומרי של בעיה זו ייעשה ע"י קביעת איתחול ל- \underline{h} ול- \underline{X} , והפעלת ירידה לפי הגרדיאנט (SD) לסירוגין בעזרת שתי המשוואות הנ"ל עד להתכנסות. כיוון שפונקציית היעד אינה קמורה, אין כל הבטחה למציאת הפתרון הטוב ביותר, ובמידה רבה בחירת האיתחול קובעת את תוצאת הסיום.

3. הורדת רעש (Denoising) ע"י התמרת Wavelet

הורדת רעש מתמונה אינה אלא מקרה פרטי של בעיית השיחזור הכללית. לכן, בשימוש בשערוך MAP עם טיפול נאות בשפות, תוצאת התהליך תהיה נקודת המינימום של הפונקציה

$$\varepsilon\{\underline{X}\} = \|\underline{Y} - \underline{X}\|_2^2 + \lambda \rho\{\mathbf{D}\underline{X}\}$$

שאלה מעניינת שנבחנה בספרות היא שאלת בחירת האופרטור \mathbf{D} . מסתבר כי עבור בחירת \mathbf{D} כמטריצה אורתונורמלית מתקבל פתרון סגור ונוח אשר קל במיוחד ליישום. עד כה פעלנו תחת ההנחה כי \mathbf{D} מיישם גזירה כלשהי. נוכל להציע התמרת Wavelet אורתונורמלית כאופרטור זה – הרכב המטריצה כולל נגזרות שונות (מסנני HPF במרווחי תדר שונים), ולכן בחירה זו נראית סבירה. נניח אם כך כי \mathbf{D} היא מטריצה אורתונורמלית ונסמן $\mathbf{D}\underline{X} = \underline{X}_D$. מתקבל כי מינימיזציה של פונקציית המחיר הנ"ל שקולה למינימיזציה של

$$\varepsilon\{\underline{X}\} = \|\underline{Y} - \underline{X}\|_2^2 + \lambda \rho\{\mathbf{D}\underline{X}\} = \|\mathbf{D}^T(\underline{Y}_D - \underline{X}_D)\|_2^2 + \lambda \rho\{\underline{X}_D\} =$$

$$= \|\underline{Y}_D - \underline{X}_D\|_2^2 + \lambda \rho\{\underline{X}_D\} = \sum_{k=1}^N (y_d[k] - x_d[k])^2 + \lambda \rho\{x_d[k]\}$$

בפיתוח הנ"ל עשינו שימוש בכך שנורמת L_2 אינה מושפעת מאופרטור אורתונורמלי, דהיינו $\|\mathbf{D}\underline{X}\|_2 = \|\underline{X}\|_2$. כמו כן השתמשנו בעובדה שהפונקציה מהווה סכום פונקציות סקלריות על ערכי הוקטור המוזן –

$$\rho\{\underline{V}\} = \sum_{k=1}^N \rho\{v_k\}$$

ברישום הנ"ל אנו מניחים כי התמונה כוללת N פיקסלים וכך גם מספרם של מקדמי ההתמרה.

יופיו של הביטוי שהתקבל נובע מכך שחישוב של כל מקדם $x_d[k]$ יכול להיעשות בנפרד כפתרון של בעיית אופטימיזציה סקלרית פשוטה. עלינו להביא למינימום את הפונקציה

$$\begin{aligned}
 f(x) &= (x - y)^2 + \lambda \rho(x) \\
 \Rightarrow f'(x) &= 2(x - y) + \lambda \rho'(x) = 0 \\
 \Rightarrow x &= y - \frac{\lambda}{2} \rho'(x)
 \end{aligned}$$

ובינתיים נראה כי יידרש פתרון נומרי לפתרון בעיה זו (אם כי הוא פשוט! בשל היות הנעלם חד-מימד).

אחת העבודות המוכרות ביותר בתחום ה-Denoising היא עבודתם של Donoho ו-Johnston. נקודת המוצא שלהם הייתה האבחנה שאם מסתכלים על תמונת רועשת דרך מקדמי ה-Wavelet שלהם, רואים כי ישנם הרבה מקדמים קטנים שנובעים אך ורק מרעש. לכן, דיכוי אלו יביא לאפקט ניקוי. במלים אחרות, קבעו השניים כי תמונה נקייה מרעש צריכה להיות בעלת מקדמי Wavelet שרובם אפסים. בתרגום לשפתנו, הם בחרו בפונקציה רובסטית מהצורה

$$\rho(x) = |x|^p \quad p \in (0,1]$$

עבור $p=0$ אנו סופרים בפועל את מספר האיברים בהתמרה ששונה מאפס, וכך חותרים ליעד של דלילות מקדמי ההתמרה. אפילו שזו בחירה לא קמורה, ניתן לפתור בצורה סגורה בעיית אופטימיזציה זו. במקום להתייחס לנגזרות (שתעוררנה בעיות בשל אי-חלקות הפונקציה) נחזור לפונקציית המחיר המלאה שקובעת במקרה זה כי

$$\varepsilon\{\underline{X}\} = \sum_{k=1}^N (y_d[k] - x_d[k])^2 + \sum_{k=1}^N \lambda |x_d[k]|^p$$

הגורם השני סופר מספר איברים שונה מאפס בוקטור \underline{X}_D והאיבר הראשון רוצה לראות קירבה בין וקטור זה לוקטור המדידות \underline{Y}_D (שאינו אלא \underline{Y} לאחר התמרת Wavelet). נניח כי מקדמי ההתמרה ב- \underline{Y}_D סודרו בסדר יורד בערך מוחלט כך שהקרובים שבהם לאפס מצויים בתחתית הוקטור. אם עלינו לאפס איברים, ברור כי יש לעשות זאת בזנב הוקטורים שם השגיאה הריבועית הנגרמת קטנה. לכן, נתחיל עם הבחירה $\underline{X}_D = \underline{Y}_D$ ונאפס את האיבר האחרון, זה שלפניו וחוזר חלילה. עם איפוס האיברים נבחן את ערך הפונקציה $\varepsilon\{\underline{X}\}$ ונעצור את תהליך האיפוס במקום בו תתחיל עליה בשגיאה. אם נבחרו להישאר M איברים ראשונים שונים מאפס יהיה מחיר השגיאה

$$\varepsilon_M\{\underline{X}\} = \sum_{k=M+1}^N y_d[k]^2 + \lambda M$$

במעבר ל- $M-1$ איברים שונים מאפס (בשל איפוס איבר נוסף) השגיאה תהיה

$$\varepsilon_{M-1}\{\underline{X}\} = \sum_{k=M}^N y_d[k]^2 + \lambda(M-1)$$

ואם $\varepsilon_{M-1}\{\underline{X}\} > \varepsilon_M\{\underline{X}\}$ יש להספיק את איפוס האיברים. דרישה זו מיתרגמת ל-

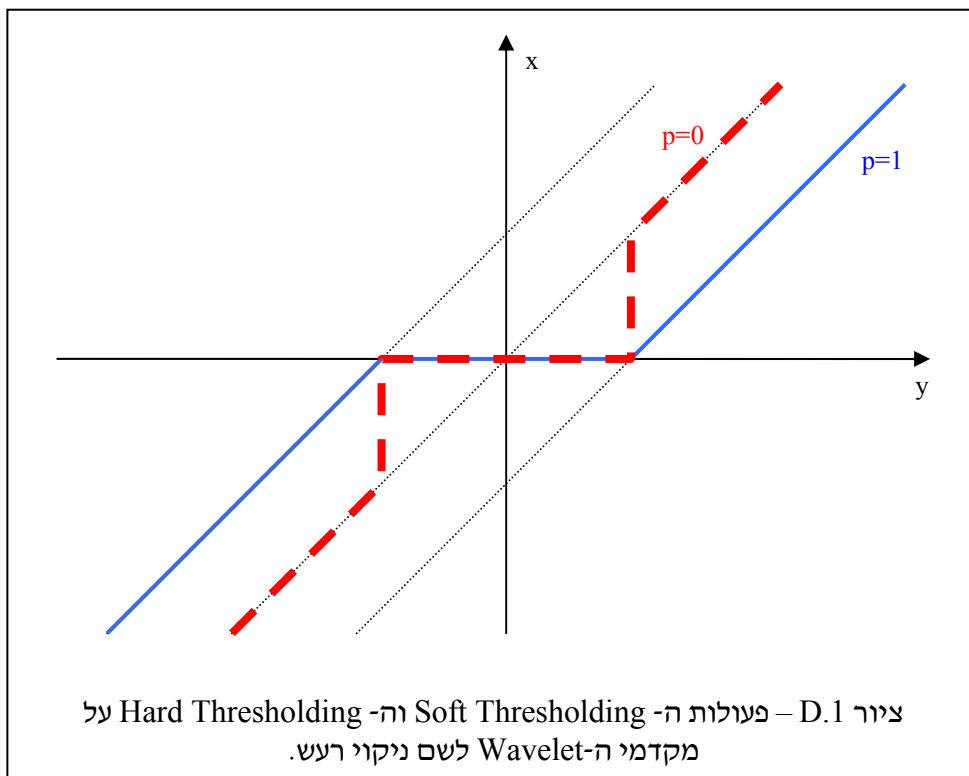
$$\begin{aligned}
 \varepsilon_{M-1}\{\underline{X}\} &= \sum_{k=M}^N y_d[k]^2 + \lambda(M-1) > \sum_{k=M+1}^N y_d[k]^2 + \lambda M = \varepsilon_M\{\underline{X}\} \\
 \Rightarrow y_d[M]^2 > \lambda &\quad \Rightarrow |y_d[M]| > \frac{\sqrt{\lambda}}{2}
 \end{aligned}$$

התקבל אלגוריתם לניקוי רעש שמציע את השלבים הבאים – יש להתחיל בלקיחת וקטור המדידות \underline{Y} ולבצע עליו התמרת Wavelet. את כל מקדמיו שערכם המוחלט מתחת לסף $\sqrt{\lambda}/2$ יש לאפס. את הווקטור הנוצר יש להמיר חזרה ע"י התמרת Wavelet הפוכה ומתקבלת תוצאת ניקוי הרעש.

החלפת p ב-1 ייתן קירוב לדלילות הרצויה ויביא לריכוך בדרך פעולת האלגוריתם. יש לשים לב לכך שבחירת $p=2$ תעבוד נגד דלילות ותעדיף ערכים רבים קטנים על פני ערך אחד גדול בהתמרה, כפי שכבר ראינו בפרק 6, ולכן אינה רצויה. חמושים בבחירת הפונקציה הרובסטית עם $p=1$ נקבל שפתרון בעיית הרעש בתמונה יוצא כפתרון של N בעיות בלתי תלויות מהצורה

$$x = y - \frac{\lambda}{2} \rho'(x) = y - \frac{\lambda}{2} \text{sign}(x)$$

אם y גדול מ- $\lambda/2$ אזי פתרון למשוואה זו הוא $x = y - \lambda/2$. באופן דומה, אם y קטן מ- $-\lambda/2$, הפתרון יהיה $x = y + \lambda/2$. לכל ערך באינטרוול שבין שני ספים אלה לא קיים פתרון למשוואה זו בשל אי-הרציפות של פעולת הסימן, ואז נציע כפתרון אפס לשם קבלת קרבה מרבית בין שני האגפים, ולשם קבלת עקביות עם הפתרון עבור $y=0$. ציור D.1 מראה את פעולת השינוי על מקדמי ה-Wavelet בשני המקרים שנתחו.



נושא חשוב שצובר תאוצה ונמצא ברובו בזירה מחקרית עדיין הוא נושא השימוש בדוגמאות כתחליף לפונקציית הסתברות אנליטית לתמונה. אם נחשוב על כך לרגע, יש נאיביות גדולה ברישום מהצורה

$$P(\underline{X}) = \text{Const} \cdot \exp\{-\lambda \rho\{D\underline{X}\}\}$$

כיוון שהבחירה בביטוי אנליטי שנוח לגזירה מחייב פשרות אכזריות עם המציאות – תמונה לרוב נוטה להיות חלקה (חוץ מפעמים שהיא לא כל כך חלקה, כמו במרקמים), שפות נוטות להיות חדות (חוץ מבמקרים שבהם שפה שאינה חדה קורית, כמו במקרה של תנאי הארה לא נוחים), ועוד ועוד. כך יוצא שעל מזבח הנוחות המתמטית אנו מקריבים את הצלחת פתרון בעיית ההיפוך בה נעשה שימוש ב-Prior מעין זה.

ומהי האלטרנטיבה? דוגמאות! הרבה מאוד דוגמאות! מאגר עצום של תמונות שתישמרנה כניסיון לייצג את מרבית התופעות שאנו צופים לראות בתמונה. בהינתן דוגמאות אלה נוכל תיאורתית לנסות ולבנות את פונקציית ההסתברות, ואז לבצע את השחזור, אלא שגישה זו מאוד לא יעילה. ניסיון מעין זה הוצע ע"י Mumford ו-Zhu – הם בחרו פונקציית הסתברות מעט יותר כללית מהצורה

$$P(\underline{X}) = \text{Const} \cdot \exp\left\{-\sum_k \lambda_k \rho_k\{D_k \underline{X}\}\right\}$$

בפונקציה זו לא פועל אופרטור גזירה יחד אלא מספר אופרטורים כאלה, ועל כל אחת מתוצאות הגזירה מופעלת פונקציה רובסטית אחרת. עיקר עבודתם של Mumford ו-Zhu היתה פתרון השאלה – בהינתן בנק של תמונות דוגמה, האם נוכל למצוא את הבחירה הטובה ביותר לאופרטורים D_k – הבחירת הפונקציות הרובסטיות שנלוות להן? גישת הפתרון אותה בחרו מורכבת למדי, וקרוב לודאי שניתן לפשטה ללא אובדן ביצועים. הנקודה היא שהם יצרו את הקשר בין פונקציית ה-Prior ובין אוסף דוגמאות. גישתם היא גישת השערוך הפרמטרי ואין זו בהכרח הגישה המוצלחת כיוון שהיא כופה סד של מבנה הפונקציה ממנו מתחילים.

האם נוכל לעשות שימוש בדוגמאות בצורה ישירה יותר? תשובה שהולכת ונרקמת בימים אלה. פאן מסויים אותו נזכיר נובע מעבודות על סינטזת מרקמים של Effros ואחרים. עבודה של העת האחרונה (פורסמה חלקית ב-2003) היא עבודתם של Weissman ו-Sarousi (בעלי עבר מהטכניון) והיא נראית כמי שתביא למהפיכה בתחום זה של שימוש בדוגמאות לשם פתרון בעיות היפוך, כשאצלם הנחות היסוד מורכבות עוד יותר – הדוגמאות הן נעשה שימוש נובעות מהמידע הפגום הנחוץ לשחזור.

נציע את הגישה הבאה לשימוש במאגר התמונות, ונתייחס לצורך התיאור של השימוש במאגר בבעיית ניקוי הרעש מתמונה. נבחר חלון של $(2L+1) \times (2L+1)$ פיקסלים סביב כל פיקסל במאגר, ונבנה בצורה כזו מילון עצום של וקטורים באורך $(2L+1)^2$. לכל פיקסל בתמונה המקולקלת הנתונה נתייחס לחלון בגודל $(2L+1) \times (2L+1)$ סביבו, נתעלם מערך הפיקסל עצמו המצוי במרכז ונביא מהמילון שנבנה את כל הוקטורים בהם ישנו דימיון גדול לתוכן הבלוק (למעט איבר האמצע). דימיון זה יכול להימדד כ-

MSE, או בדרכים אחרות שלוקחות בחשבון שפות, כיוונים ועוד. כעת שבידנו K דוגמאות של בלוקים בעלי רקע דומה נסתכל על פילוג הערך המצוי במרכז וזה יאמר לנו את חוק פילוגו של הפיקסל המטופל. נניח כי מצאנו פילוג גאוסי בעל ממוצע m ושונות s . נניח כי הערך שנמדד בפועל הוא y ואנו שואלים במה להחליפו. שימוש במשעך MAP יאמר כי את x יש לקבוע לפי

$$\hat{x}_{\text{opt}} = \underset{x}{\text{Arg min}} \frac{1}{\sigma_n^2} (x - y)^2 + \frac{1}{s^2} (x - m)^2 = \frac{\frac{y}{\sigma_n^2} + \frac{m}{s^2}}{\frac{1}{\sigma_n^2} + \frac{1}{s^2}}$$

(אנו מניחים ידיעת שונות הרעש בתמונה הנתונה). ברור כי אין אנו מחוייבים לייצוג ע"י פילוג גאוסי וכל פילוג אחר יאפשר מציאת x אופטימלי, כשמה שנדרש באמת הוא היסטוגרמת ערכי הפיקסל האמצאי בבלוקים שהובאו מהמילון כבלוקים קרובים.

אם נחזור על תהליך זה לכל פיקסל ונעשה זאת למספר סיבובים על פני כל הפיקסלים, תתקבל תמונה ששואבת את ערכיה מתוך מאגר הנתונים.

לדוגמה, נניח כי בתמונה המדודה בחלון מסוים שכל תוכנו בעל רמת אפור קרוב ל-100 הפיקסל המרכזי הוא 255. נניח כי לקיחת הקונטקסט (כל הבלוק למעט מרכזו) והבאת בלוקים דומים מהמאגר הניבה 10000 בלוקים רלוונטיים ובהם פילוג הערך המרכזי מצביע בסבירות גבוהה על 100, ייצא כי s (פיזור הערכים) קטן מאוד ולכן הערך במרכז הבלוק יוחלף ב-100, וכך ניקינו רעש נקודתי. לרעש זה אין זכות להתקיים כי במאגר התמונות הוא נדיר עד כדי בלתי אפשרי. לא די בסיבוב אחד על הפיקסלים כיוון שבסיבוב ראשון הזנו קונטקסט פגום בשל הרעש שנכלל בו, ומטבע הדברים ההחלטה שנבעה משימוש בקונטקסט זה שגויה במקצת. סיבובים נוספים בהם הופחת הרעש ישפרו את ניקוי התמונה.

הגישה שהוצעה כאן בנויה בצורה אינטואיטיבית, והיא דומה מאוד לגישה בה מילא Effros חורים המרקמים באופן מוצלח למדי. ניתן לאמץ ראייה הסתברותית יותר תיאורתית לתהליך זה, אך אנו נעצור את הדיון כאן.