



# עיבוד ספרתי של תמונות

# Digital Image Processing

נמרוד פלג

המעבדה לעיבוד אותות ותמונות  
הפקולטה להנדסת חשמל, הטכניון

עדכון : אפריל 2010

# פרק 1 : מבוא לעיבוד תמונה

- מהי תמונה?  
האור המוחזר מ"אובייקט", לאחר מעבר במערכת אופטית, ונקלט במערך חיישנים (סרט צילום, CCD ...).
- התמונה היא דו-ממדית, ובה עוצמת האור (בהיקות) וצבעו (אורך גל) תלויים במקום:

פילוג העצמה והצבע של הקרינה

# “תמונות” אחרות

- ישנם סוגים רבים של תמונות שאינן נקלטות דווקא במערכת אופטית:

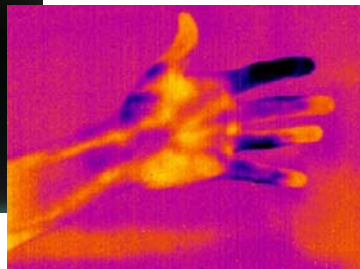
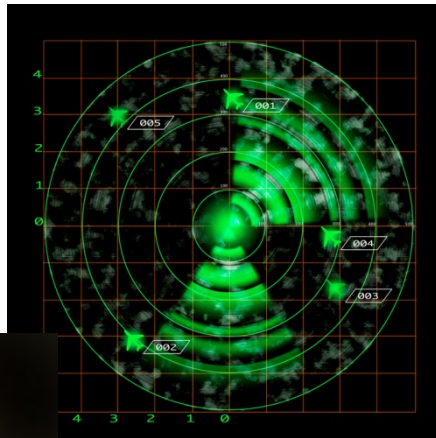
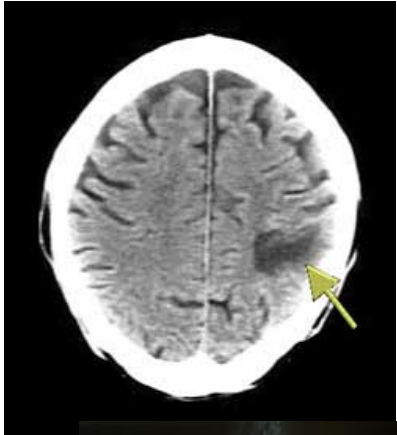
– ראדאר

– רנטגן

– CT

– מערך מיקרופונים

– אור בלתי נראה (IR) ועוד....



אבל מה שאנחנו אוהבים לראות זה...

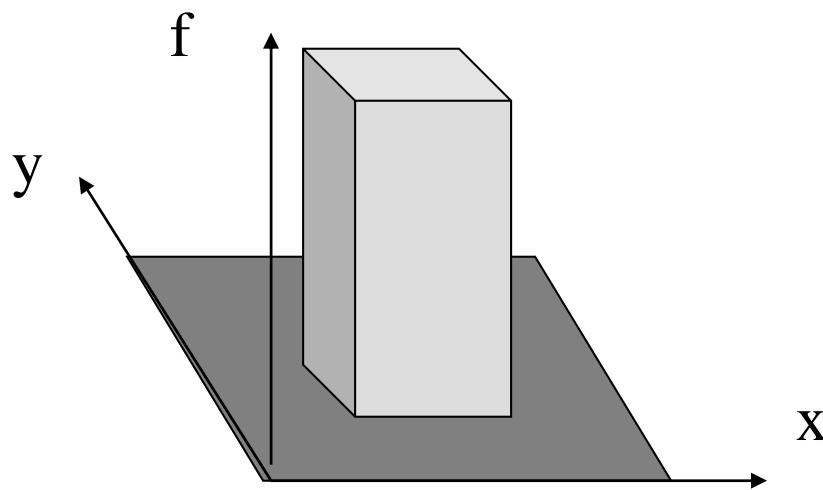
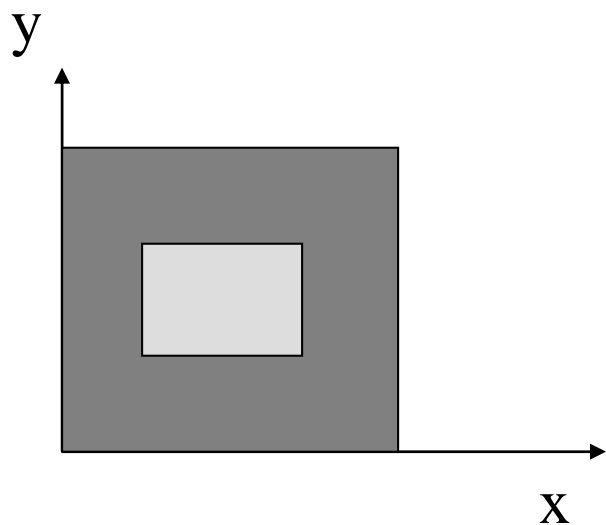


צינזורה!



# ייצוג מתמטי של תמונה

- נתאר תמונה כפונקציה של שני משתנים:  $f(x, y)$
- $x, y$  הן הקורדינטות במישור דו-ממדי  $(x, y)$ , והערך של  $f$  מבטא (בתמונה אופטית) את הבהיקות (עצמת אור) בנקודה.



## ומה בתמונת צבע?

- בתמונת צבע מורכבת כל נקודה (pixel) משלשה צבעי יסוד: אדום, ירוק וכחול (מסומן RGB) ולכן נתאר כל פיקסל ע"י שלש פונקציות:

$$r(x, y), g(x, y), b(x, y)$$

בדור"כ נעסוק בתמונות רמות-אפור, לשם הנוחות.

# רמות האפור ורציפותן

- במקור, האובייקט אותו מצלמים הינו "רציף" הן בשטחו ("מרחב") והן בצבעו ("בהיקות"), אך במעבר למחשב חייבים, משיקולים מעשיים לדגום את התמונה בנקודות מסוימות (פיקסלים בדידים) ובצבעים מסוימים (רמות אפור בדידות)
- מתקבל מערך (או מטריצה) של נקודות מיצגות

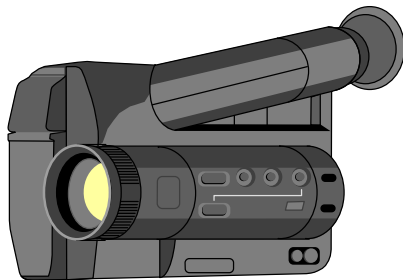
הערה: בדיד = דיסקרטי , Discrete

# אמצעי קלט לתמונות

• אמצעי הרכשה (Grabbing) לתמונה:

– מצלמה ספרתית (בדר"כ מבוססת: CCD)

- מדידת מתח/זרם יחסי לכמות פוטונים, עבור כל פיקסל.
- רזולוציה בסיסית: 640 על 480 - (VGA) רחוק מסרט צלום.
- אלטרנטיבה: שורת חישנים עם סריקה מכנית (איטי)
- לצבע: שלשה מערכים + מסננים אופטיים מתאימים.
- קיימים חישנים מתאימים גם ל IR-רנטגן וכד.

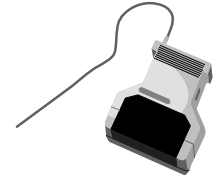




# אמצעי קלט לתמונות (המשך)

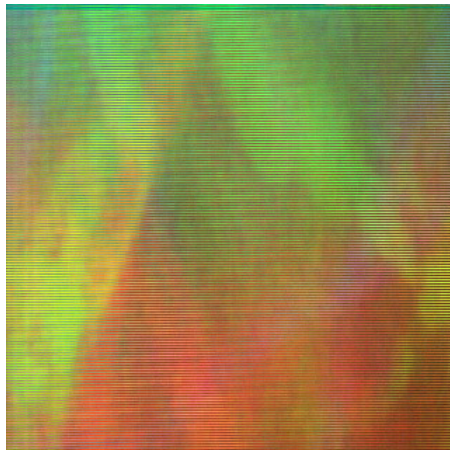
– סורק אופטי:

- עקרון דומה עם הקלות: כיול נוח, רקע אחיד, תאורה מבוקרת.
- רזולוציית הסריקה נמדדת בדר"כ ב , dpi - ולא עולה על 600 נקודות לאינץ' (כ- 24 נקודות למ"מ)
- רזולוציות גבוהות יותר מושגות ע"י אינטרפולציה!

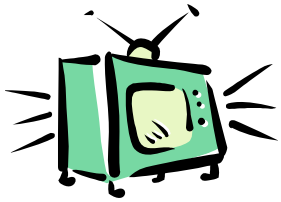


– תמונות סינטטיות:

- תמונות מחושבות בשיטות הדמייה: מכ"מ... US, MRI, CT,
- מדידת תופעה פיזיקלית ותרגום שלה
- למערך דו-ממדי.



# אמצעי פלט לתמונות



• המרת תמונה ספרתית לתצוגה: (Display)

– צגי שפופרת קרן-קתודית (שק"ק: CRT)

• קרן סורקת המאירה נקודות של חומר פלואורסנטי. (RGB)

• פרמטרים חשובים: מהירות הסריקה, רזולוציית המסך.

• אפשרויות אחרות, לדוגמא: LCD :

– מדפסות:

• Continuous Tone יקרות ודומות לפיתוח סרט צילום אמיתי.

• בדר"כ שימוש במגוון צבעים מוגבל וטכניקת Half-Tone

ליצירת אשליה של רמות אפור/צבע. (Laser, Ink, Wax)

# תמונת lna עם אפקט Half-Tone



# עיבוד תמונה ותחומים סמוכים

- “עיבוד תמונות” דן ב“משימות” בהן הקלט והפלט הינם תמונות (ספרתיות או אנלוגיות).
- “ראיה ממוחשבת” : הקלט תמונה, והפלט אינו תמונה אלא מידע עליה. (Computer Vision)
- “גרפיקה ממוחשבת” : הקלט הינו מידע כלשהו והפלט הינו תמונה. (Computer Graphics).
- לדוגמא : מציאת המיפוי התלת-מימדי של גופים בתמונה, לעומת בנית תמונה ריאליסטית ממידע על גופים תלת-מימדיים במרחב. (Rendering)

# מסלול אופיני לתמונה

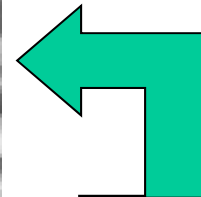


# תחומים שונים בעיבוד תמונה

- שיפור (Enhancement): סילוק "קלקולים" שנוצרו בדגימה.
- שחזור (Restoration): סילוק קלקולים שטיבם ידוע.
- ניתוח (Analysis): זיהוי מרכיבים בדר"כ לצורך מע' אוטומטיות.
- דחיסה (Compression): ייצוג חסכוני ע"י סילוק מידע לא חיוני.
- בנייה (Reconstruction): הרכבה על-סמך מידע חלקי (הדמיה ע"י "חתכים")

# שיפור תמונה

קלקולים אופייניים המתקבלים במהלך דגימת או  
צילום התמונה,  
מפריעים להבחין  
בפרטים, מטשטשים  
את האובייקט  
וכדומה.



תמונת מקור : תצ"א  
(Aerial)



# ניגודיות נמוכה (Low Contrast)





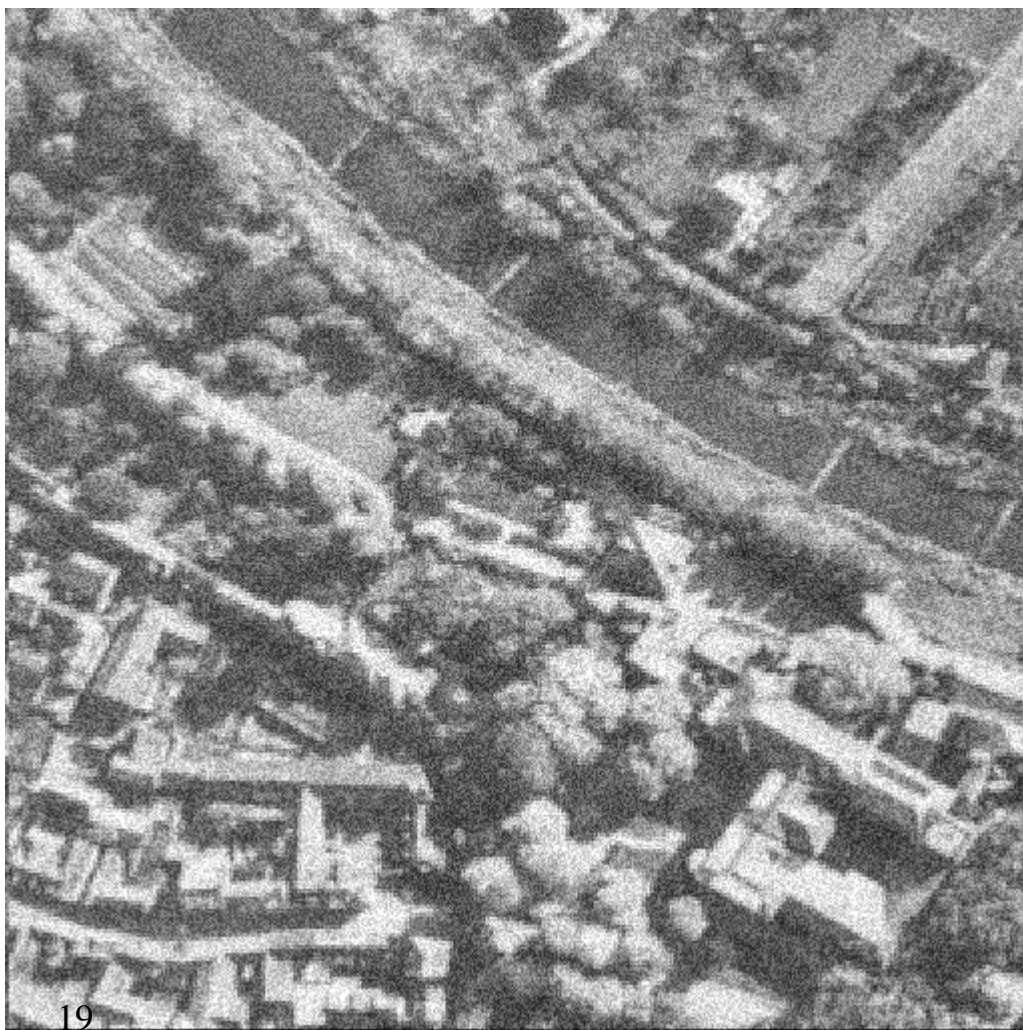
# טשטוש (Blur)



# מריחה בתנועה (Motion Blur)



# רעש אחיד (Uniform Noise)



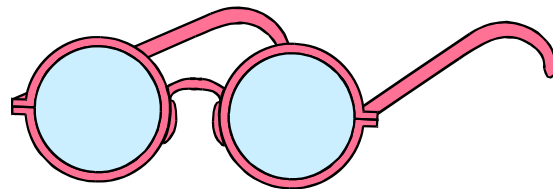
# קלקולים נוספים

- מיקוד לא נכון. (De-Focus).
- תזוזה או רעידה של המצלמה או הגוף.
- תנועות תרמיות של אויר (טורבולנציה)
- ניתן לתקן חלק מהבעיות ע"י עיבוד מתאים אך  
חייבים לזכור:

לא ניתן להוסיף מידע שלא היה קיים בתמונה המקורית !

# שיטות לשיפור התמונה

- השיטות רבות ומגוונות ותלויות באופי הקלקול ובמטרת השיפור!
- מדידת השיפור, פעמים רבות סובייקטיבית (לפי העין), ולעיתים ניתנת להערכה מתמטית.



# ניתוח תמונה

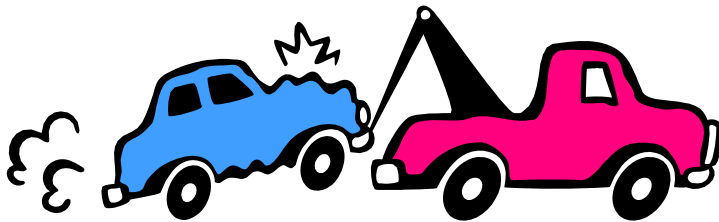
- הכוונה לשיטות המאפשרות שליפת מידע מתמונה באמצעים ממוחשבים, ללא צורך המתבונן בנושא.
- דוגמאות:

– מיון עצמים

– איתור תקלות במעגלים מודפסים

– זיהוי כתב

– ניהוג רכב אוטומטי



# דחיסת תמונה



33,252B דחוס 8:1



262,144B = 512 x 512 מקור

# טכניקות של עיבוד תמונה

- עיבוד אופטי
- עיבוד אלקטרוני - אנלוגי
- עיבוד אלקטרוני - ספרתי



# עיבוד אופטי

- אין המרה של **התמונה הרציפה** לנקודות בדידות
- מספר פעולות ניתנות לביצוע:
  - התמרת פורייה
  - סינון מרחבי
  - חישובי קורלציה, קונבולוציה
- יתרונות: **מהירות** (האור), **מקביליות** (כל הנקודות מחושבות במקביל), **דיוק** ועוד.

## עיבוד אופטי (המשך)

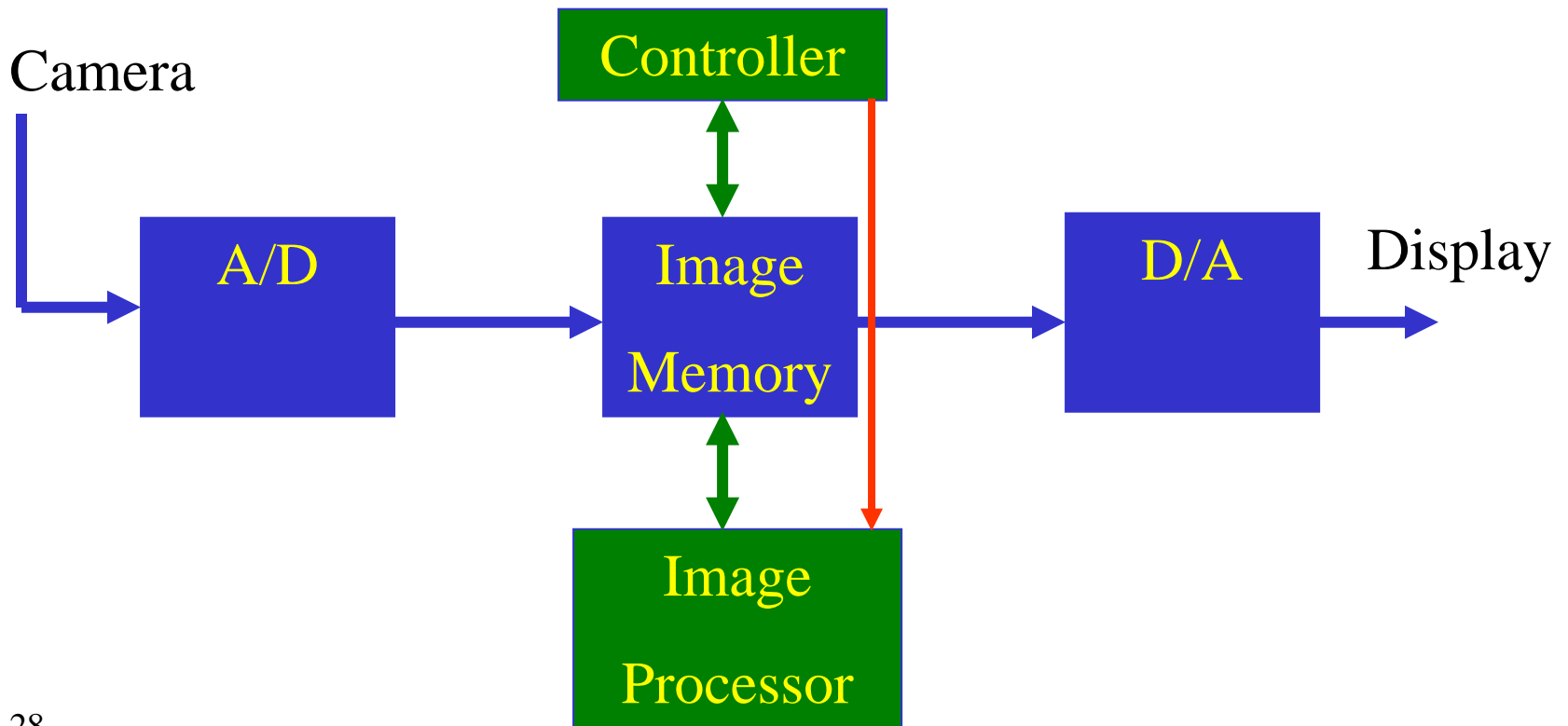
- מגוון פעולות מוגבל (חסרות אפילו 4 פעולות היסוד)
- חוסר גמישות מחייב מערך אופטי ייחודי לכל פעולה.
- דרוש אור קוהרנטי (לייזר)
- העברה ושמירת המידע הינם בעייתיים מאד.

# עיבוד אלקטרוני - אנלוגי

- המרה למידע חשמלי, בייצוג ע"י אות אנלוגי.
- ישנם מספר עיבודים אפשריים: שיפור קונטרסט, הורדת רעשים ועוד.
- העיבודים בזמן-אמת, שומר על דיוק גבוה וזול.
- גם כאן יש בעיות מעשיות של מערכות חמרה: חוסר גמישות, רגישות להתיישנות רכיבים, רגישות לשינויי טמפ', בעיית איכסון ועוד.

# עיבוד אלקטרוני - ספרתי

- תהליך שמחייב המרת התמונה ע"י דגימה והמרה בחזרה לאות רציף ע"י שיחזור:



# עיבוד אלקטרוני - ספרתי (המשך)

- יתרונות:

- כל פעולה אריתמטית אפשרית

- אמצעי איכסון והעברה פשוטים וזולים

- גמישות רבה

- יציבות מערכת

- מחיר זול

- עדיין קיימת בעיית מהירות עיבוד אך הולכת ויורדת עם התפתחות המיחשוב (שעון, מיקבול וכו')

# אפיון של תמונות

ננסה לאפיין תמונות לפי מדדים אובייקטיביים:

1. צבעוניות
2. בהירות (Brightness) ובהירות (Luminance)
3. ניגודיות (Contrast)
4. כושר הפרדה (Resolution)
5. תדרים מרחביים
6. רעש (SNR, MSE)

# איפיון: צבעוניות

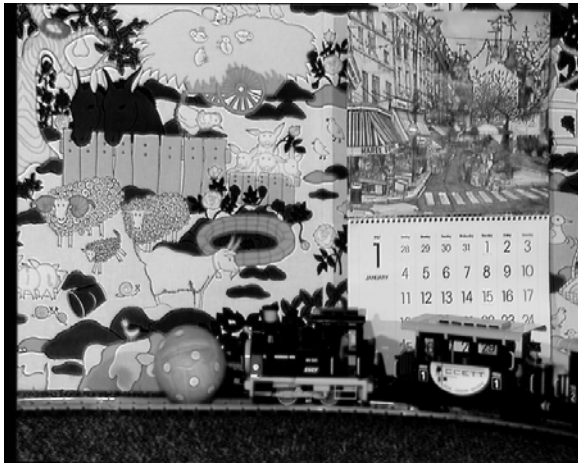
- הצבע מוסיף **מידע וטבעיות**, נעים יותר לעין.
- בדר"כ ייצוג הצבע במרחב תלת-מימדי: בטלויזיה ע"י אדום, כחול וירוק, (RGB) ובדפוס ע"י ארגמן, תכול וצהוב, (YCM) + שחור, מסיבות של איכות ומחיר.
- המשמעות היא הכפלה פי שלשה של כמות המידע. בתמונת צבע לעומת תמונת רמות-אפור.

# RGB Color Space

Original (color) image



R



G



B

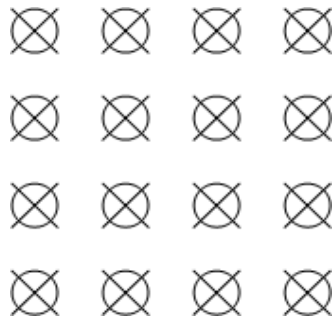




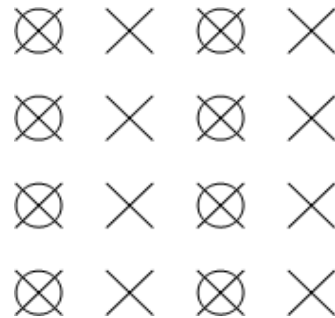
# YCrCb Color Space

- Linear transformation of RGB color components
- Y = luminance component (brightness)  
Cr,Cb = chrominance components (color)

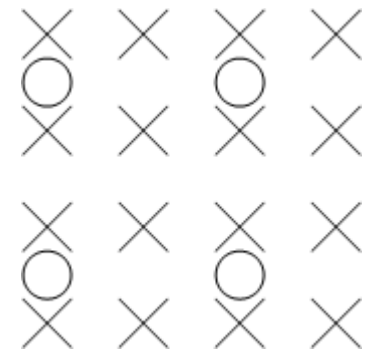
YCrCb 4:4:4



YCrCb 4:2:2



YCrCb 4:2:0



x – luma samples

o – chroma samples

# 4:2:0 Format - Example

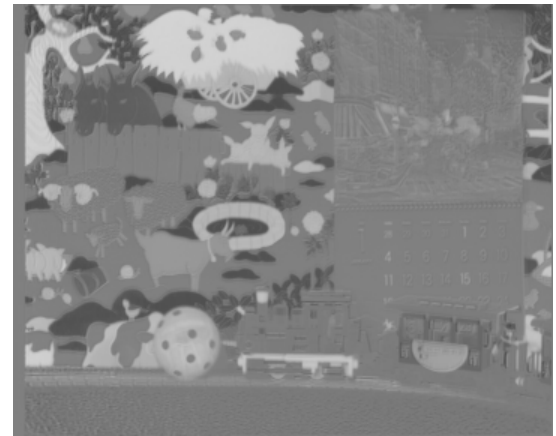
**Luminance - Y**



**Chrominance**



**Cb**



**Cr**

## איפיון II : בהיקות ובהירות

- המידע בתמונה טמון **בפילוג הבהירות** בה. הבהירות הינה סובייקטיבית ומתוארת ע"י הגודל האובייקטיבי **בהיקות** : עצמת האור הנפלטת מיחידת שטח, לכוון ההסתכלות.
- **הבהיקות** מתארת גם את הצג ותנאי התאורה, וניתן לשנותה כרצוננו. הגודל הכמותי המתאר את הבהיקות הוא רמת האפור, (Gray-Level) הנותן מיקום מיספרי בין רמת-השחור לרמת-הלבן המרביות.

## איפיון II : בהיקות ובהירות (המשד)

- רמות-אפור : בדר"כ רמת-האפור מתוארת ע"י 8 סיביות ולכן תנוע בתחום 0-255 או לחלופין מנורמלת בין 0 ל-1.



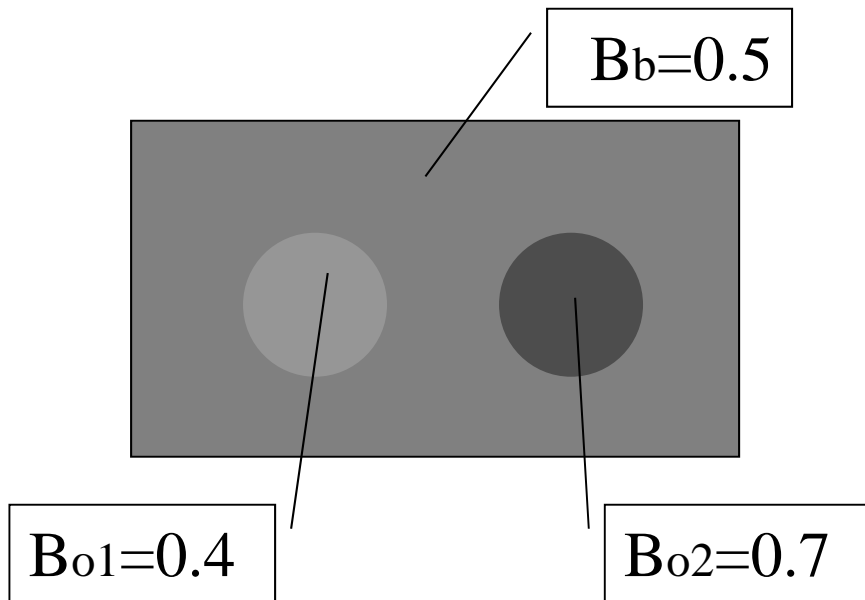
תמונה בה רמות  
האפור נעות  
בהדרגתיות בין לבן  
לשחור

## איפיון III : ניגודיות

- הניגודיות מציינת את ההבדל בין רמות האפור של אובייקט לאלו של סביבתו.
- יכולת ההבחנה של העין תלויה מאד בניגודיות והרבה פחות בבהירותו !
- ניגודיות מקומית מוגדרת ע"י :

$$Contrast = \left| \frac{B_{object} - B_{background}}{B_{background}} \right|$$

## איפיון III : ניגודיות (המשך)



$$C_1 = \left| \frac{B_{o1} - B_b}{B_b} \right| = \frac{0.5 - 0.4}{0.5} = 0.2$$

$$C_2 = \frac{0.7 - 0.5}{0.5} = 0.4$$

- במקרה זה לעיגול הימני ניגודיות חזקה יותר כלפי הרקע מאשר לימני.

## איפיון IV : כושר הפרדה

- כל תמונה ספרתית מתוארת ע"י מערך של נקודות (פיקסלים), המהוות כל אחת **דגימה מרחבית** של התמונה האמיתית.
- ככל שמספר הפיקסלים וצפיפותם יהיו גדולים יותר, יהיה ניתן להבחין בפרטים "עדינים" יותר.
- מספר הפיקסלים בכל כיוון קובע את "**גבול הפרדה**" של התמונה.
- בפועל, מספר **הפיקסלים האקטיביים** בתמונה קטן מהגבול התאורטי.

# איפיון IV: כושר הפרדה (המשך)

- מדידת ההפרדה נעשית ע"י תמונת קווים מחזוריים:



למדידת ההפרדה האפקית נשתמש בקווים אנכיים בצפיפות משתנה, בפילוג סינוסי (כמו כאן) או, בדר"כ, פשוט בקווי שחור-לבן.

למדידת ההפרדה בתמונה מספיק לתאר כמות קווי הפרדה אפקית ואנכית .

במכשיר תצוגה (צג) או הדפסה מתייחסים גם לגודל הפיסי :  
צפיפות נקודות ליחידת אורך. (DPI).

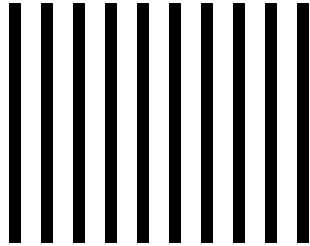


# איפיון $V$ : תדרים מרחביים

- השימוש בייצוג ע"י תדר זמני מוכר היטב באותות חד-מימדיים, ומקובל גם בדו-מימד, בו הוא מייצג תדר מרחבי.
- טכניקת המעבר היא ע"י התמרת פורייה (Fourier Transform) ודומותיה (הרבות), בצורה הדו-ממדית שלהן.
- תכונות התמרות דו-ממדיות דומות בדר"כ לחד-ממדיות (אם הן ספרביליות) ויוכרו בהמשך.

# איפיון V: תדרים מרחביים (המשך)

- דוגמאות של תדר מרחבי:



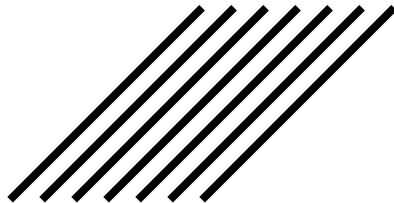
$$f_x = 30_{cycles / width}$$

$$f_y = 0_{cycles / height}$$



$$f_x = 0_{cycles / width}$$

$$f_y = 60_{cycles / height}$$



$$f_x = 60_{cycles / width}$$

$$f_y = 30_{cycles / height}$$

המספרים  
מקורבים בלבד!

ניתן לחשב לפי  
רוחב הדף, לפי  
ס"מ, אינץ' וכו'.

# איפיון VI : רעש (לכלוך ?)

- ה"רעש" (מונח הלקוח מעיבוד אות דיבור...) הינו ערך אקראי ובלתי רצוי, הנוסף בכל אחד משלבי יצירת התמונה ונוסף למידע הרצוי.



תמונת Lena  
המקורית, בתוספת  
רעש גאוסני לבן

$$f(x, y)$$

$$g(x, y) = f(x, y) + n(x, y)$$

## איפיון VI: רעש (המשד)

- איפיון הרעש מתמקד בדר"כ בתלותו במקום בתמונה, ובעצמתו. (מניחים **אי תלות בתמונה** עצמה).
- מניחים (אם אין מידע אחר) אי-תלות בגורמים ידועים ולכן הרעש הוא אקראי.
- שיערוך סטטיסטי של הרעש אפשרי, באמצעות פונקצית פילוג ההסתברות, כשהפילוג הגאומטרי (או ה"נורמלי") הוא אחד המקובלים.

# רעש : פילוג גאوسی (I)

- לפי הפילוג הגאוסי, ההסתברות שערכו המקומי של הרעש יהיה בין שני ערכים סמוכים  $n, n + \Delta n$

הנו :

$$p(n) = \frac{1}{\sqrt{2\pi\sigma_n^2}} e^{-\frac{n^2}{2\sigma_n^2}} \cdot \Delta n$$

Standard Deviation : סטיית התקן :  $\sigma_n$

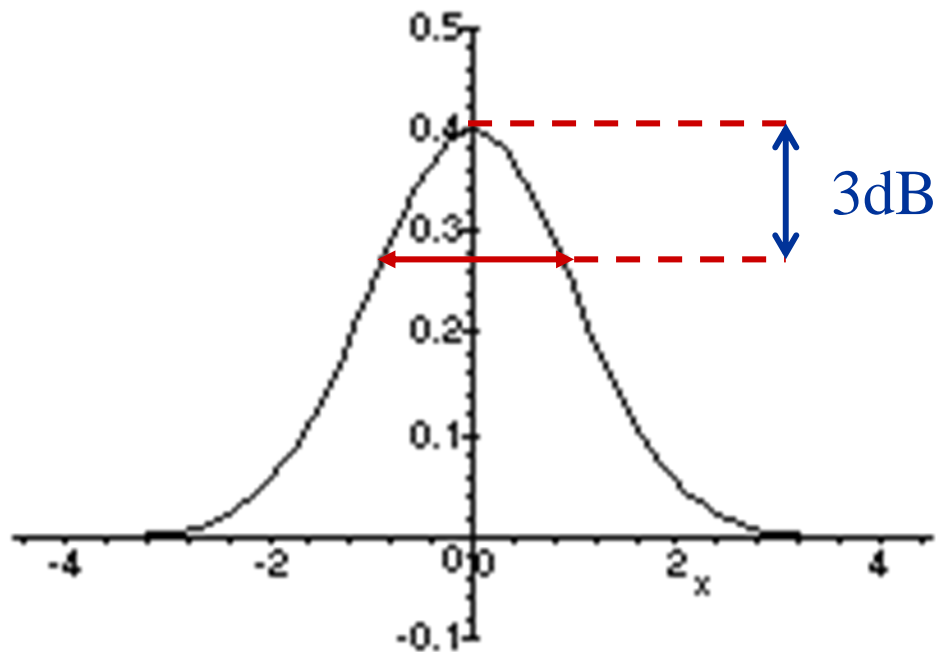
Variance : שונות :  $\sigma_n^2$



Carl Friedrich Gauss, 1777-1855

# רעש : פילוג גאוסִי - עקומת פעמון

## (II)



## רעש : פילוג גאוס (III)

• תכונות של רעש גאוס:

– יכול לקבל ערכים חיוביים ושליילים

– ההסתברות יורדת ככל שערך הרעש עולה

– הערך הממוצע של משתנה אקראי נקרא **תוחלת**

ומחושב לפי:

$N_v, N_h$   
מס' הפיקסלים  
בשני הצירים

$$E_n = \frac{1}{N_v N_h} \sum_i \sum_j n_{ij}$$

$$\sigma_n^2 = \frac{1}{N_v N_h} \sum_i \sum_j (n_{ij} - E_n)^2 \quad \text{ואת השונות לפי:}$$

## רעש : פילוג גאוס (IV)

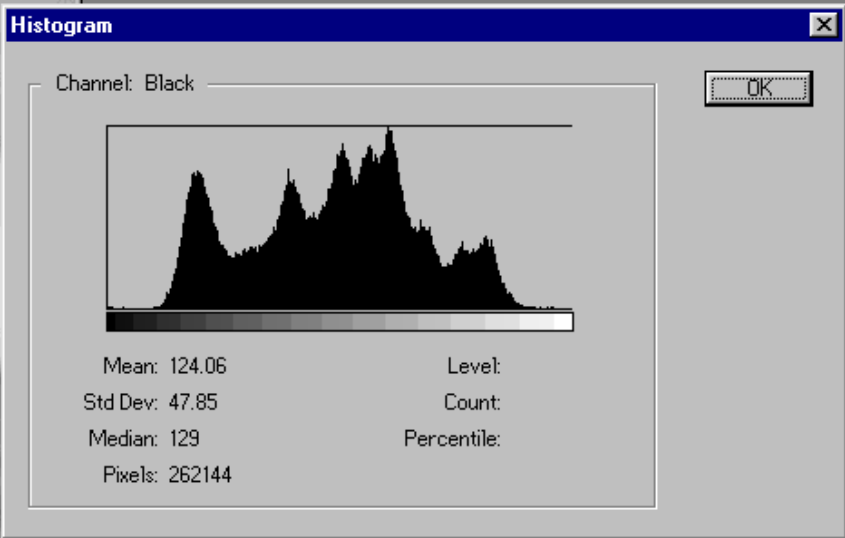
- תוחלת הרעש (Expected Value) היא ערך הרעש במרכז הפעמון (הסימטרי)
- השונות מייצגת את ממוצע ריבוע המרחק בין ערכי הרעש השונים לבין התוחלת.
- תוחלת: עבור משתנה אקראי, ממוצע: עבור משתנה דטרמיניסטי

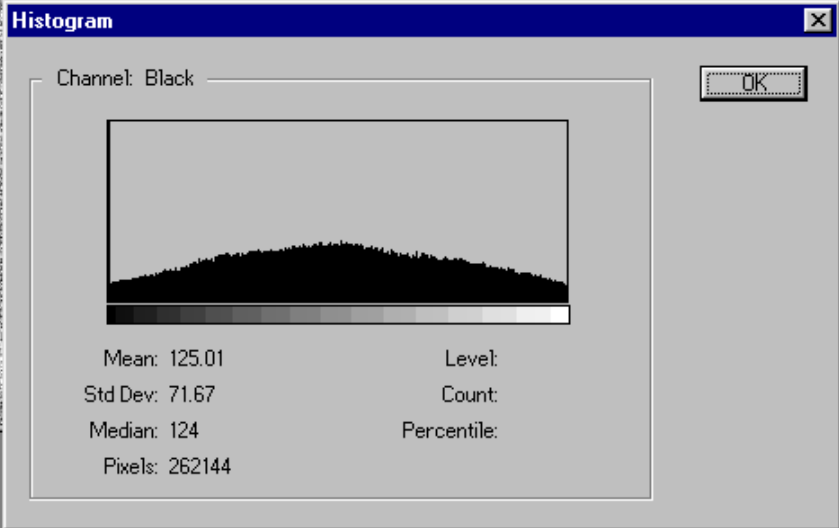
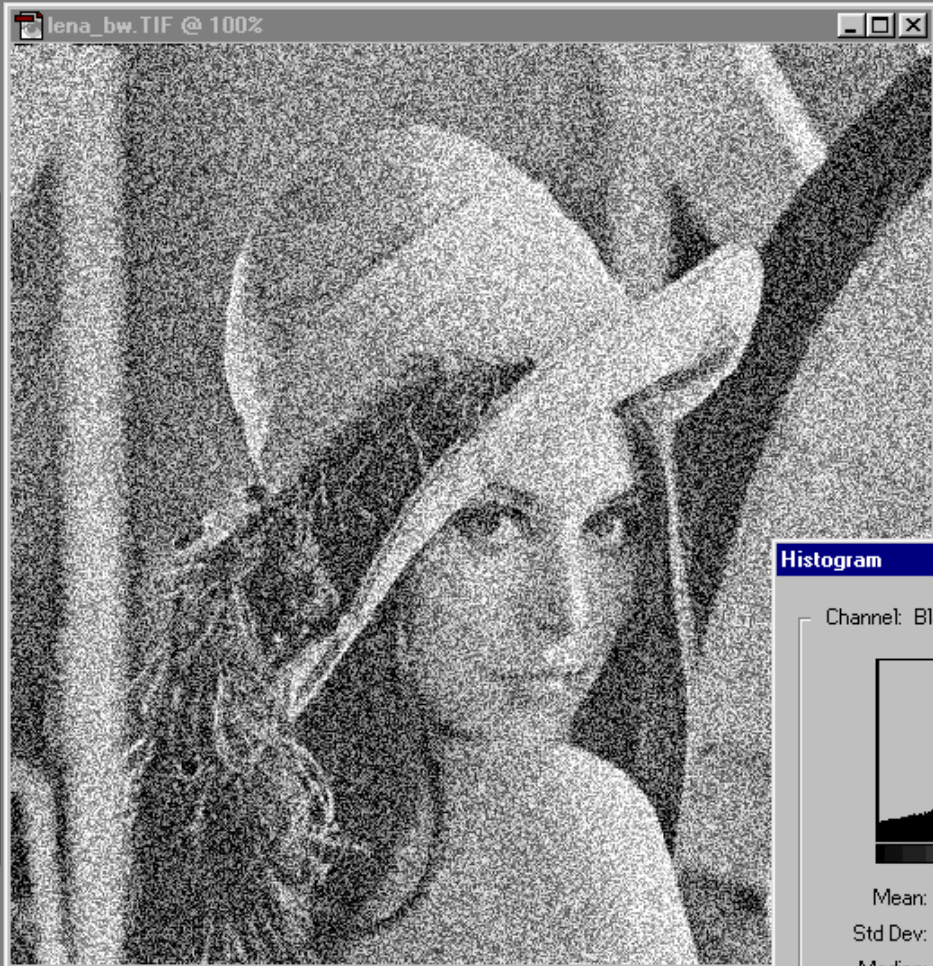


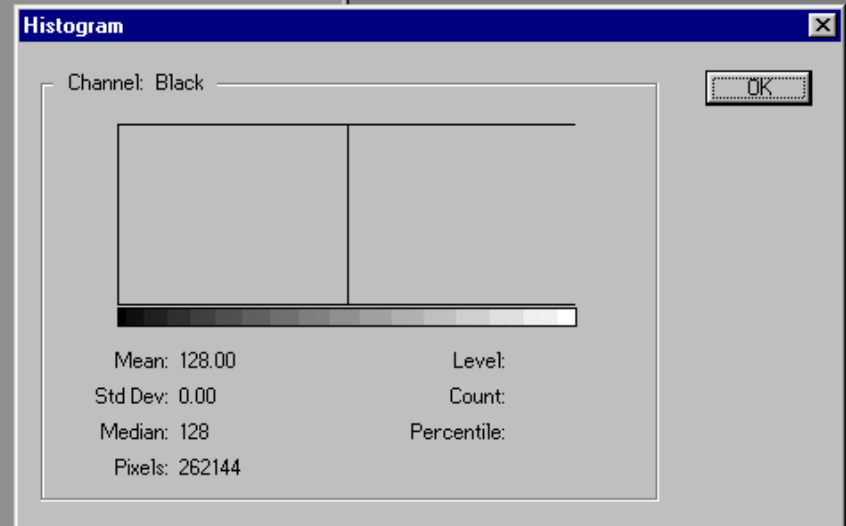
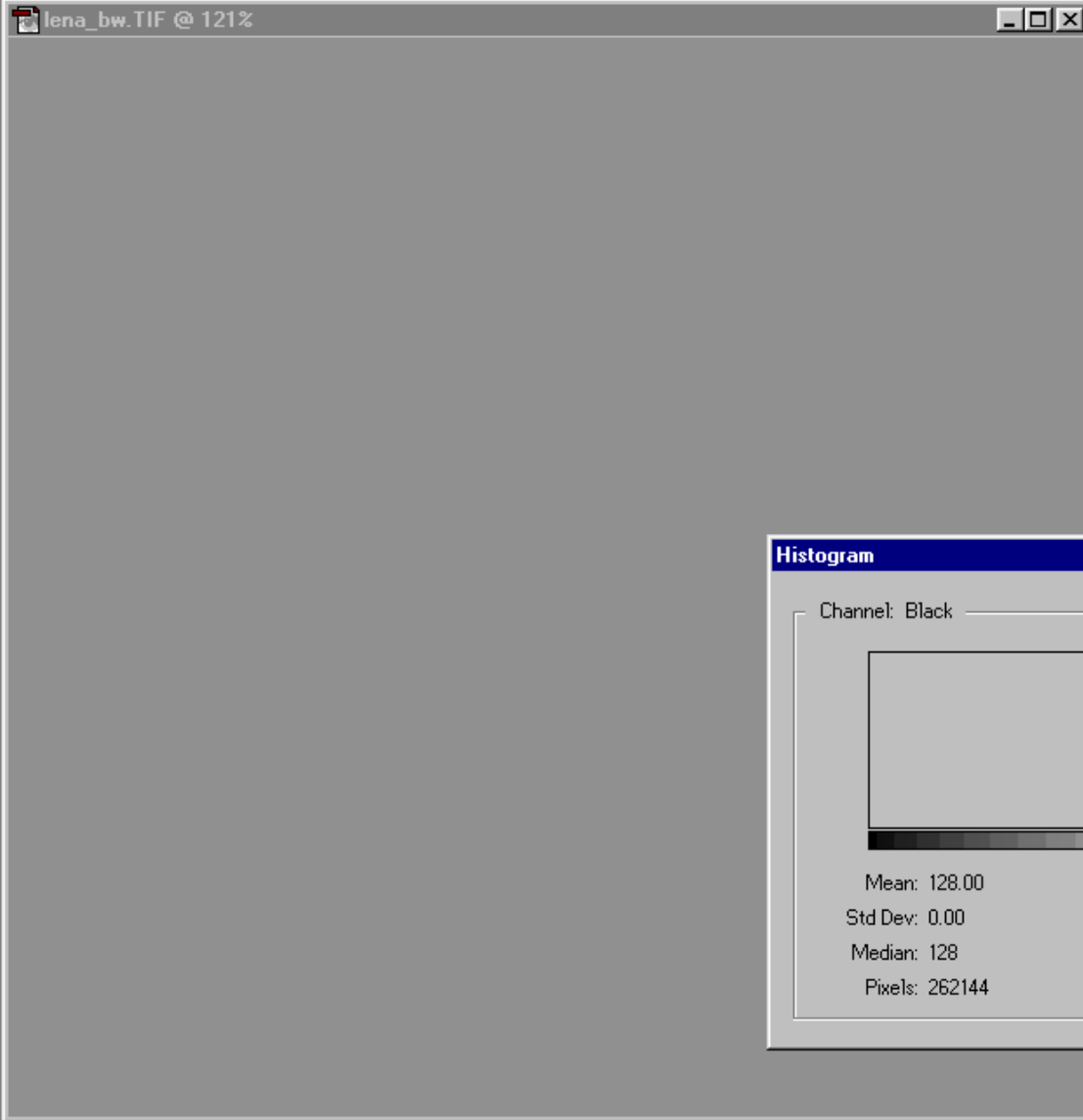
# היסטוגרמה של תמונה

- התפלגות רמות האפור בתמונה מתוארת ע"י היסטוגרמה: צבירת כמות הפיקסלים בכל רמת אפור, גובה כל עמודה מתאר את כמות הפיקסלים.
- בדוגמאות הבאות:
  - תמונת Lena עם ההיסטוגרמה שלה.
  - תמונת Lena בתוספת רעש יוניפורמי + היסטוגרמה.
  - תמונה חלקה
  - בתוספת רעש גאوسي

lena\_bw.TIF @ 100%

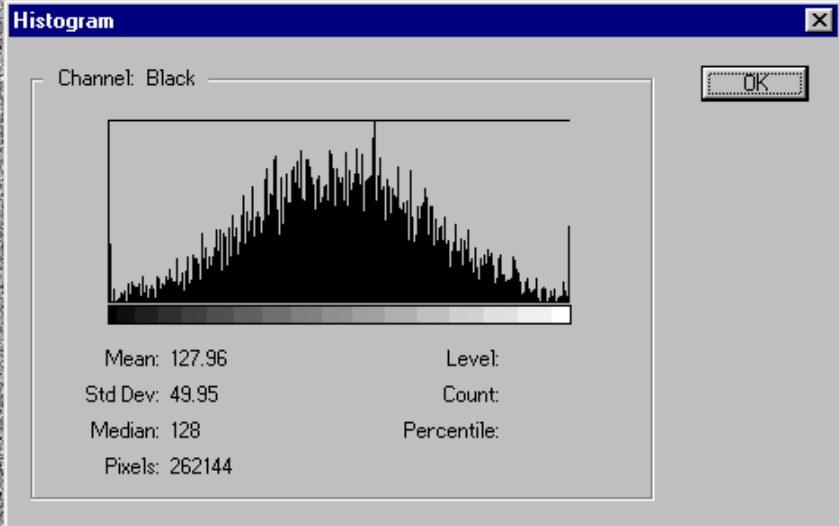
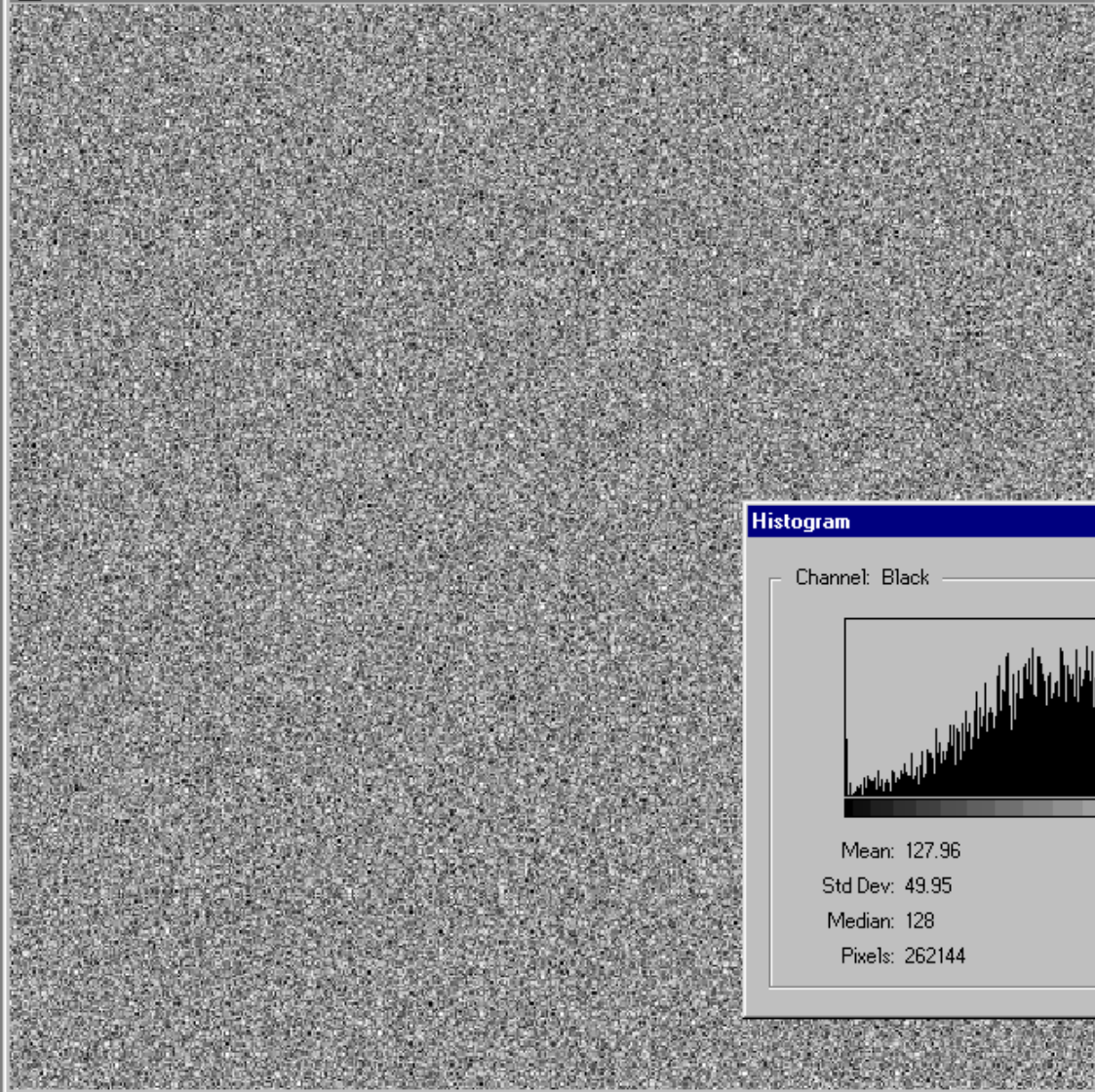








lena\_bw.TIF @ 121%



# MSE : מדידת רעש בתמונה

- שגיאה ריבועית ממוצעת – Mean Squared Error
- הבסיס הוא השוואה בין שתי תמונות (או אותות אחרים כמובן), כאשר אחד "מקור" והשני "תוצאה"
- נמדוד את השוני ביניהם, ולפי השוני ניתן איפיון של **דמיון** (=איכות) או לחלופין **שגיאה** (=עיוות).
- $\mathbf{x} = \{X_i | i = 1, 2, \dots, N\}$  and  $\mathbf{y} = \{Y_i | i = 1, 2, \dots, N\}$   
are two finite-length, discrete signals (e.g. Images)

$$MSE_{(X, Y)} = \frac{1}{N} \sum_{i=1}^N (X_i - Y_i)^2$$

$e_i = X_i - Y_i$  is the "error" signal

# נורמה כללית

- ה"נורמה" היא צורה כללית יותר לבטא זאת:

$$d_p(X, Y) = \left( \sum_{i=1}^N |e_i|^p \right)^{1/p} \text{ is the norm } l_p$$

- בעיבוד תמונות בדרך-כלל נבטא זאת ע"י יחס האות לרעש – **PSNR**

# יחס אות לרעש : SNR

- כמות ההפרעה שנוספת לתמונה נמדדת ע"י היחס בין האות המקורי לרעש הנלווה. Signal to Noise Ratio :
- נהוג להגדיר את כמות המידע ("אנרגיה") באות כלשהו, כשונות (Variance) שלו. לפיכך, "גודל" האות יוגדר:

$$\sigma_s^2 = \frac{1}{N_v N_h} \sum_i \sum_j (s_{ij} - E_s)^2$$

$S_{i,j}$  : ערך האות בנקודה  $(i,j)$

$E_s$  : תוחלת בהירות האות



## יחס אות לרעש SNR: (המשך)

- ואילו יחס האות לרעש SNR יוגדר בסקלה לוגריתמית:

$$SNR = 10 \log_{10} \left( \frac{\sigma_s^2}{\sigma_n^2} \right) = 10 \log_{10} \left( \frac{\sum_i \sum_j (s_{ij} - E_s)^2}{\sum_i \sum_j (n_{ij} - E_n)^2} \right)$$

- יש לשים לב כי עבור רעש גאוסי לבן בעל ממוצע אפס, בתמונה בעלת מס' סופי של נקודות  $N$  השאיפה לאפס היא יחסית ל-  $\sqrt{N}$ .

# יחס אות לרעש : PSNR

- מדידה המתאימה יותר לתמונות ולמערכת הראיה נקראת Peak-SNR ומחושבת לפי:

$$PSNR = 10 \log_{10} \frac{(2^n - 1)^2}{MSE}$$

כאשר  $n$  הוא מספר הסיביות של הדגימות – כלומר במונה נמצא רבוע הערך הגבוה ביותר האפשרי באות.

# מגבלות השימוש ב- PSNR

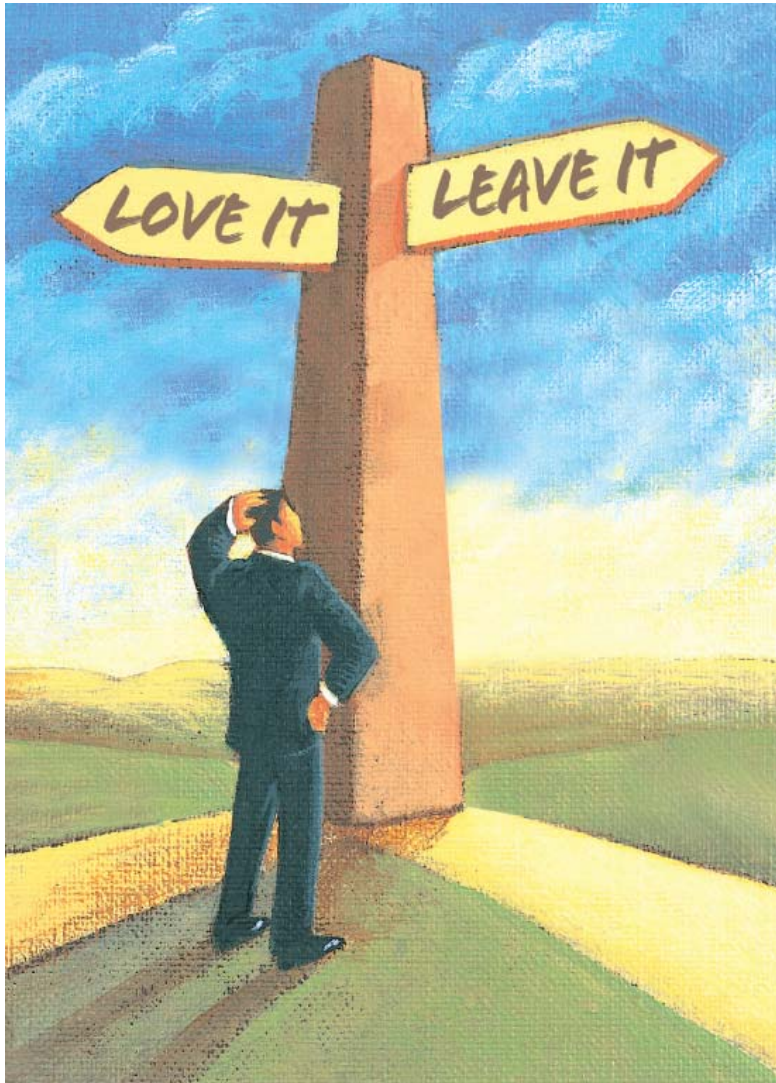
- המדידה דורשת ידיעת תמונת המקור !
- המדידה הינה יחסית למדידות מקבילות ואין משמעות למספרים עצמם .
- מאמצים רבים מושקעים בפיתוח שיטות מדידת איכות אובייקטיביות כדוגמת התקנים :
  - ITU-R BT.500-10
  - P.900
  - ITU-T Video Quality Expert Group (VQEG)

# תחום דינמי של אות

- התחום הדינמי (Dynamic Range) הוא הערך המרבי של SNR שניתן לקבל בתמונה עבור רמת רעש נתונה. כלומר: מספר הפעמים שהרעש "נכנס" באות המרבי.



# Appendix : MSE and SSIM



Based on:

Zhou Wang and Alan C. Bovik,

**Mean Squared Error:**

**Love It or Leave It?**

IEEE Signal Processing Magazine,

**January 2009**

# Why MSE ?

1. It is **simple**. It is parameter free and inexpensive to compute, *only one multiply and two additions per sample*.
2. It is also **memoryless**: evaluated at each sample, independent of other samples.
3. All  $l_p$  norms are **valid distance metrics** in  $\mathbf{R}^N$ ,
  - Nonnegativity:  $d_p(\mathbf{x}, \mathbf{y}) \geq 0$
  - Identity:  $d_p(\mathbf{x}, \mathbf{y}) = 0$  *if and only if*  $\mathbf{x} = \mathbf{y}$
  - Symmetry:  $d_p(\mathbf{x}, \mathbf{y}) = d_p(\mathbf{y}, \mathbf{x})$
  - Triangular inequality:  $d_p(\mathbf{x}, \mathbf{z}) \leq d_p(\mathbf{x}, \mathbf{y}) + d_p(\mathbf{y}, \mathbf{z})$ .
4. In particular, the  $p = 2$  case (proportional to the square root of the MSE) is the **ordinary distance metric** in  $N$ -dimensional Euclidean space.
5. It has a **clear physical meaning**: it is the natural way to **define the energy** of the error signal. Such an energy measure is preserved after any orthogonal linear transformation (Fourier etc)

The energy preserving property guarantees that the energy of a signal distortion in the transform domain is the same as in the signal domain. This property distinguishes  $d_2$  from the other  $l_p$  energy measures, which are not energy preserving.

# Why MSE ? Cont'd

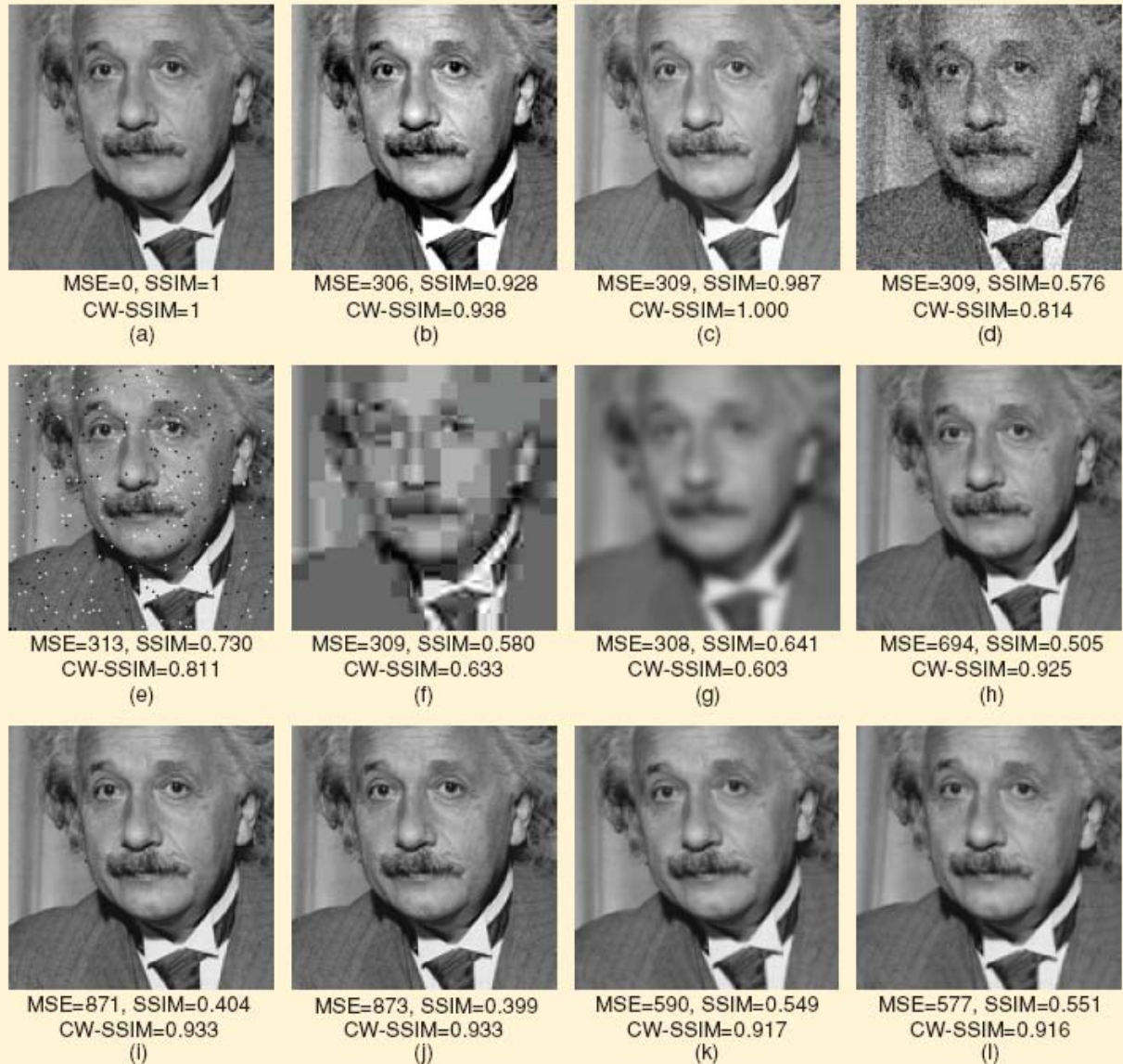
6. The MSE is an excellent metric in the context of **optimization**. The MSE possesses the very satisfying properties of convexity, symmetry, and differentiability. Minimum-MSE (MMSE) optimization problems often have **closed-form analytical solutions**.
  - when they don't, iterative numerical optimization procedures are often easy to formulate, since the gradient of the MSE is easy to compute.
  
7. MSE is **widely used** simply because it is a **convention**.
  - Historically, it has been employed extensively for optimizing and assessing a wide variety of signal processing applications, including filter design, signal compression, restoration, de-noising, reconstruction, and classification.

# Why **NOT** MSE ?

- The more fundamental issue has been missing: does the MSE really measure signal fidelity?
- Given all of its above-mentioned attractive features, a signal processing practitioner might choose the MSE if it proved to be a reasonable signal fidelity measure. **But is that the case?**



# Comparison of Image Fidelity Measures



[FIG2] Comparison of image fidelity measures for “Einstein” image altered with different types of distortions. (a) Reference image. (b) Mean contrast stretch. (c) Luminance shift. (d) Gaussian noise contamination. (e) Impulsive noise contamination. (f) JPEG compression. (g) Blurring. (h) Spatial scaling (zooming out). (i) Spatial shift (to the right). (j) Spatial shift (to the left). (k) Rotation (counter-clockwise). (l) Rotation (clockwise).

Zhou Wang and Alan C. Bovik,  
**Mean Squared Error:  
Love It or Leave It?**  
IEEE Signal Processing Magazine ,  
January 2009

# An alternative: **Structural Similarity**

- A framework of **image fidelity measurement** as an image communication problem
- The SSIM proves high effectiveness for measuring the **fidelity** of signals
- The principle underlying the SSIM approach is that the human visual system is highly adapted to **extract structural information** from visual scenes

# Basic Form of SSIM

- Suppose that  $\mathbf{x}$  and  $\mathbf{y}$  are local image patches taken from the same location of two images that are being compared.
- The local SSIM index measures the similarities of three elements of the image patches:
  - $l(\mathbf{x}, \mathbf{y})$  of the local patch **luminances** (brightness values)
  - $c(\mathbf{x}, \mathbf{y})$  of the local patch **contrasts**
  - $s(\mathbf{x}, \mathbf{y})$  of the local patch **structures**
- These local similarities are expressed using simple, easily computed statistics, and combined together to form **local SSIM**

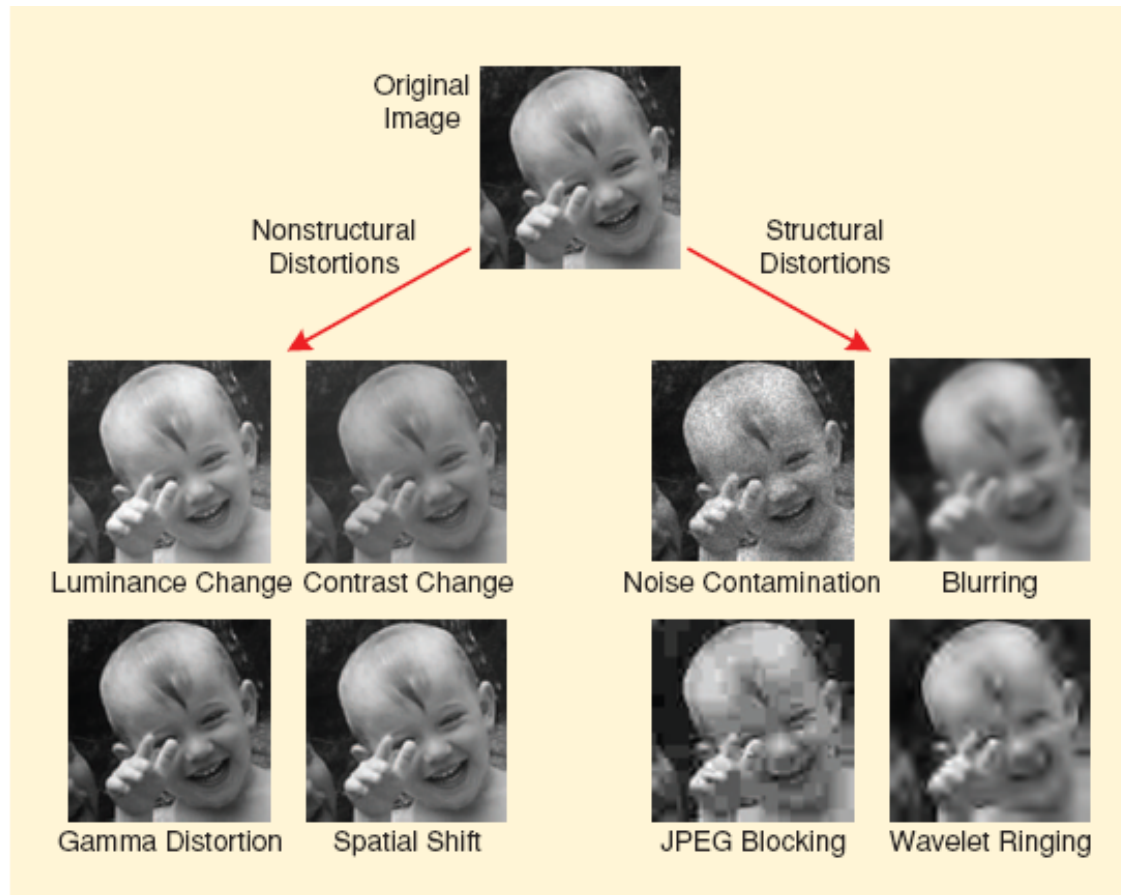
# SIMM Calculation

$$\begin{aligned} S(x, y) &= l(x, y) \cdot c(x, y) \cdot s(x, y) = \\ &= \left( \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \right) \cdot \left( \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \right) \cdot \left( \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \right) \end{aligned}$$

Where:

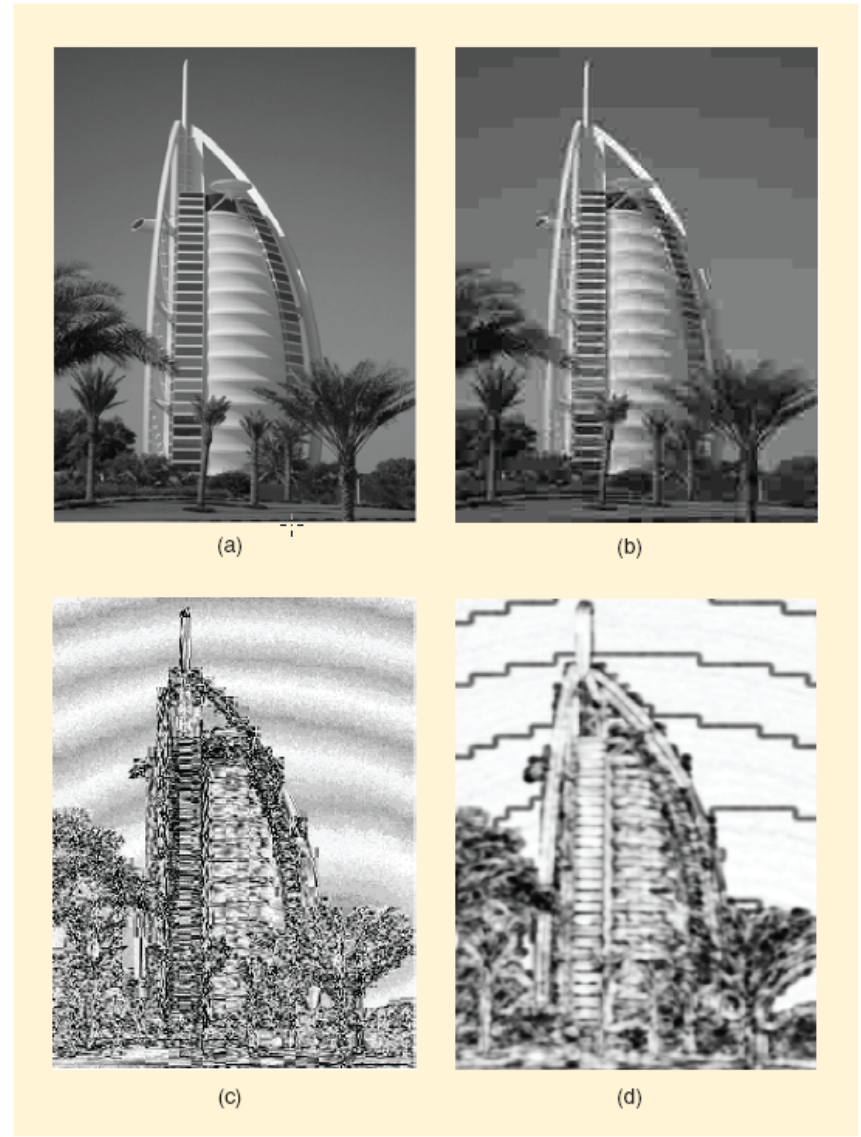
- $\mu_x$  and  $\mu_y$  are (respectively) the local sample **means of x and y**
- $\sigma_x$  and  $\sigma_y$  are (respectively) the local sample **standard deviations of x and y**
- $\sigma_{xy}$  is the sample **cross correlation of x and y** after removing their means.
- The items  $C_1, C_2, C_3$  are **small positive constants that stabilize each term**, so that near-zero sample means, variances, or correlations do not lead to numerical instability

# Structural vs. Nonstructural Distortions



# Example

- JPEG induced annoying **pseudo-contouring effects** (in the sky region) and **blocking artifacts** (along the boundaries of the building) that are successfully captured by the SSIM index, yet poorly predicted by the absolute error map.



[FIG8] Comparison of image fidelity/distortion maps. (a) Reference image. (b) JPEG compressed image. (c) Absolute error map of the distorted image (enhanced for visibility). (d) SSIM index map of the distorted images (enhanced for visibility).