2014 Israel Computer Vision Day
Sunday, December 21, 2014

Sponsored by:

Vision Day Schedule

Time	Speaker and Collaborators	Affiliation	Title
08:50-09:20	Gathering		
09:20-09:40	Gil Ben Artzi Michael Werman Shmuel Peleg	HUJI	Recognizing that two videos show the same event even from disparate viewpoints
9:45-10:05	Alon Faktor Michal Irani	Weizmann	Video Segmentation by Non-Local Consensus Voting
10:10-10:30	Rotem Mairon Ohad Ben Shahar	BGU	A Closer Look at Context: From Coxels to the Contextual Emergence of Object Saliency
10:35-10:55	Tal Hassner Moria Tau	Open U	Towards Dense Correspondences Between Any Two Images
11:00-11:30	Coffee Break		
11:30-11:50	Ran Margolin Lihi Zelnik-Manor Ayelet Tal	Technion	OTC: A Novel Local Descriptor for Scene Classification
11:55-12:15	Michael Baltaxe Peter Meer Michael Lindenbaum	Technion Rutgers U.	Local Variation as a Statistical Hypothesis Test
12:20-12:40	Gilad Michael Nahum Kiryati	TAU	Example Based Demosaicing
12:45-13:05	Lior Wolf Ofir Levy	TAU	Online Repetition Counting via a Convolutional Deep Neural Network Trained on Unrealistic Synthetic Data
13:05-14:10	Lunch		
14:10-14:20	"Intermezzo"		
14:20-14:40	Tali Treibitz	Haifa	How Can Computer Vision Advance Ocean Science?

14:45-15:05	Yael Moses Lior Talker Ilan Shimshoni	IDC Haifa	Have a look at what I see
15:10-15:30	Raanan Fattal	HUJI	Dehazing using Color-Lines
15:35-15:55	Anat Levin Daniel Glasner Todd Zickler	Weizmann Harvard	A Reflectance Display
16:00-16:20	Coffee Break		
16:20-16:40	Amit Aides Yoav Shechner	Technion	Clouds in the Cloud
16:45-17:05	Tammy Riklin Raviv C. Wählby V. Ljosa A.L. Conery P.Golland F.M. Ausubel A.E.Carpente	BGU Uppsala MIT Harvard	Morphology-Guided Graph Search for Untangling Objects
17:10-17:30	Matan Sela Yonathan Aflalo Ron Kimmel	Technion	Shape Exaggeration and Auto-Caricaturization

Abstracts

Recognizing that two videos show the same event even from disparate viewpoints

Gil Ben-Artzi, Michael Werman, Shmuel Peleg – Hebrew University

"Motion barcodes" are robust motion-based features that can be used to determine if two videos are of the same event even when they are taken from significantly different viewing directions. They works well even in cases where the videos are so different that appearance based methods (e.g. SIFT) are not sufficient.
The "motion barcode" records the existence/non-existence of motion as a function of time. While appearance, motion magnitude, and motion direction can vary between viewpoints, the existence of motion is viewpoint invariant. Based on the motion barcode, a similarity measure is developed for videos from very different viewpoints.

This measure is robust to occlusions common under different viewpoints and can be computed efficiently.

Video Segmentation by Non-Local Consensus Voting

Alon Faktor and Michal Irani - Weizmann

We address the problem of Foreground/Background segmentation of "unconstrained" video. By "unconstrained" we mean that the moving objects and the background scene may be highly non-rigid (e.g., waves in the sea); the camera may undergo a complex motion with 3D parallax; moving objects may suffer from motion blur, large scale and illumination changes, etc. Most existing segmentation methods fail on such unconstrained videos, especially in the presence of highly non-rigid motion and low resolution. We propose a computationally efficient algorithm which is able to produce accurate results on a large variety of unconstrained videos. This is obtained by casting the video segmentation problem as a voting scheme on the graph of similar ('re-occurring') regions in the video sequence. We start from crude saliency votes at each pixel, and iteratively correct those votes by 'consensus voting' of re-occurring regions across the video sequence. The power of our consensus voting comes from the *non-locality* of region re-occurrence, *both in space and in time* – enabling propagation of diverse and rich information across the entire video sequence. Qualitative and quantitative experiments indicate that our approach outperforms current state-of-the-art methods.

A Closer Look at Context: From Coxels to the Contextual Emergence of Object Saliency

Rotem Mairon and Ohad Ben Shahar – Ben Gurion University

Visual context is used in different forms for saliency computation. While its use in saliency models for fixations prediction is often reasoned, this is less so the case for approaches that aim to compute saliency at the object level.We argue that the types of context employed by these methods lack clear justification and may in fact interfere with the purpose of capturing the saliency of whole visual objects. In this paper we discuss the constraints that different types of context impose and suggest a new interpretation of visual context that allows the emergence of saliency for more complex, abstract, or multiple visual objects. Despite shying away from an explicit attempt to capture "objectness" (e.g., via segmentation), our results are qualitatively superior and quantitatively better than the state-of-the-art.

Towards Dense Correspondences Between Any Two Images

Tal Hassner, Moria Tau - Open University

We present a practical method for establishing dense correspondences between two images with similar content, but possibly different 3D scenes. One of the challenges in designing such a system is the local scale differences of objects appearing in the two images. Previous methods often considered only small subsets of image pixels; matching only pixels for which stable scales may be reliably estimated. More recently, others have considered dense correspondences, but with substantial costs associated with generating, storing and matching scale invariant descriptors. Our work here is motivated by the observation that pixels in the image have contexts -- the pixels around them -- which may be exploited in order to estimate local scales reliably and repeatably. In practice, we demonstrate that scales estimated in sparse interest points may be propagated to neighboring pixels where this information cannot be reliably determined. Doing so allows scale invariant descriptors to be extracted anywhere in the image, not just in detected interest points. As a consequence, accurate dense correspondences are obtained even between very different images, with little computational costs beyond those required by existing methods.

OTC: A Novel Local Descriptor for Scene Classification

Ran Margolin, Lihi Zelnik-Manor and Ayelet - Technion

Scene classification is the task of determining the scene type in which a photograph was taken. In this paper we present a novel local descriptor suited for such a task: Oriented Texture Curves (OTC). Our descriptor captures the texture of a patch along multiple orientations, while maintaining robustness to illumination changes, geometric distortions and local contrast differences. We show that our descriptor outperforms all state-of-the-art descriptors for scene classification algorithms on the most extensive scene classification benchmark to-date.

Local Variation as a Statistical Hypothesis Test

Michael Baltaxe (Technion), Peter Meer (Rutgers U), Michael Lindenbaum (Technion)

Over-segmentation procedures divide an image into many pieces, each of which should ideally be fully contained in a single object. One of the most successful oversegmentation algorithms is known as local variation (LV, by Felzenswzalb & Huttenlocher). In this work, we study this algorithm and show that algorithms similar to LV can be devised by applying different statistical considerations and decisions, thus providing further theoretical justification to the LV approach. One of these algorithms, based on statistics of natural images and on a hypothesis testing decision, is denoted probabilistic local variation, and provides state-of-the-art results while keeping the same computational complexity of the LV algorithm.

Example Based Demosaicing

Gilad Michael and Nahum Kiryati – Tel-Aviv University

Demosaicing is an algorithm used to reconstruct a color image from the incomplete color samples of a color filter array (CFA). Most demosaicing algorithms can be broadly classified into spatial-domain and frequency-domain approaches. Despite significant progress in the past decade, current state of the art demosaicing

algorithms still tend to produce artifacts at high-saturation edges. In this paper we propose a new approach to demosaicing - example based. Comparative experimental evaluation shows that example-based demosaicing (EBD) produces visually superior, artifact-free results..

Online Repetition Counting via a Convolutional Deep Neural Network Trained on Unrealistic Synthetic Data

Lior Wolf and Ofir Levy – Tel-Aviv University

The task of counting the number of repetitions of approximately the same action in an input video sequence is addressed using convolutional neural networks. The proposed system runs online and not on the complete video. It analyzes sequentially blocks of 20 non-consecutive frames. The cycle length within each block is evaluated using a deep network architecture and the information is then integrated over time. A unique property of our method is that it is shown to successfully train on entirely synthetic data, created by synthesizing moving random patches. It therefore effectively exploits the high generalization capability of deep neural networks. Coupled with a region of interest detection mechanism and a suitable mechanism to identify the time scale of the video, the system is robust enough to handle real world videos collected from youtube and elsewhere. In addition, the same neural network is used not only to estimate the number of cycles, but also to infer the ``motion-start'' and ``motion-end'' states.

How Can Computer Vision Advance Ocean Science?

Tali Treibitz – University of Haifa

The ocean is a complex, vast, foreign environment that is hard to explore and therefore much about it is still unknown. Interestingly, only 5% of the ocean floor has been seen so far. As human access to most of the ocean is very limited, optical imaging systems can serve as our eyes in those remote areas. However, optical imaging underwater is challenging due to intense pressures at depth, strong color and distance dependent attenuation, refraction at the interface air/water, and the ever-changing and rugged conditions of the natural ocean. Thus, imaging underwater pushes optical imaging to its limits. This is where advanced computer vision methods may overcome some of these obstacles post-acquisition and enable large-scale operations using machine learning.
As a result, imaging systems for the ocean require a dedicated effort throughout all the development steps: design, optical, electrical and mechanical engineering and computer vision algorithms. In this talk I will demonstrate that new imaging systems can help solve acute scientific problems through an underwater in situ high-resolution microscope I developed and talk about future challenges.

Have a look at what I see

Yael Moses (IDC), Lior Talker (Haifa), Ilan Shimshoni (Haifa)

Many events today are characterized by an abundance of cameras. The performance of many applications depends on the available overlapping views of the scene. We

propose a method for guiding a photographer to rotate her/his camera to obtain an image that overlaps with an image of another photographer viewing the same scene. Such overlapping views may be used for better analysis of the region of interest. In addition, they allow the direction of translation between the two photographers to be computed, in order to help them meet. Our solution avoids 3D scene reconstruction it relies instead on a rough representation of the scene points. In addition since in our setup humans rather than robots are users of the method, we propose a novel interface for guiding the photographer to rotate the camera to the desired view.

Dehazing using Color-Lines

Raanan Fattal – Hebrew University

Photographs of hazy scenes typically have low-contrast and offer a limited scene visibility. We describe a new method for single-image dehazing that relies on a generic regularity in natural images in which pixels of small image patches exhibit one-dimensional distributions in RGB space, known as color-lines. We derive a local formation model that explains the color-lines in the context of hazy scenes and use it for recovering the scene transmission based on the lines' offset from the origin. Moreover, this model allows us to identify and dismiss pixels that do not follow the color-line model and hence, unlike existing approaches that follow their assumptions across the entire image, our algorithm validates its hypotheses and obtains more reliable estimates where possible. We also describe a Markov random field model that is dedicated for producing complete and regularized transmission maps. Unlike traditional field models that consist of local coupling, the new model is augmented with long-range connections between pixels of similar color. This allows our algorithm to properly resolve the transmission in isolated regions where nearby pixels do not offer relevant information. An extensive evaluation of our method over different types of images and its comparison to state-of-the-art methods over established benchmark images show a consistent improvement in the accuracy of the estimated scene transmission and recovered haze-free radiances.

A Reflectance Display

Anat Levin (Weizmann), Daniel Glasner (Harvard), and Todd Zickler (Harvard)

We introduce a reflectance display: a dynamic digital array of dots, each of which can independently display a custom, time-varying reflectance function. The display passively reacts to illumination and viewpoint changes in real-time, without any illumination-recording sensors, head tracking, or on-the-fly rendering. In this example the time-varying reflectance functions create a "reflectance video" that gives the illusion of a dynamic 3D model being physically-shaded by the room's ambient lighting. The top row shows a time-sequence of photographs of the dynamic display from a stationary viewpoint under fixed ambient lighting, and the bottom row shows how the display reacts to changes in ambient lighting by passively inducing the appropriate 3D shading effects.

Clouds in The Cloud

Amit Aides and Yoav Shechner, (Technion)

Lightfield imaging can be scaled up to a very large area, to map the Earth's atmosphere in 3D. Multiview spaceborne instruments suffer from low spatio-temporal-angular resolution, and are very expensive and unscalable. We develop sky lightfield imaging, by a wide, scalable network of wide-angle cameras looking upwards, which upload their data to the cloud. This new type of imaging-system poses new computational vision and photography problems, some of which generalize prior monocular tasks. These include radiometric self-calibration across a network, overcoming are by a network, and background estimation. On the other hand, network redundancy offers solutions to these problems, which we derive. Based on such solutions, the lightfield network enables unprecedented ways to measure nature. We demonstrate this experimentally by 3D recovery of clouds, in high spatio-temporal resolution. It is achieved by space carving of the volumetric distribution of semi-transparent clouds. Such sensing can complement satellite imagery, be useful to meteorology, make aerosol tomography realizable, and give new, powerful tools to atmospheric and avian wildlife scientists

Morphology-Guided Graph Search for Untangling Objects

Tammy Riklin Raviv (BGU), C. Wählby (Broad Institute of MIT and Harvard, Uppsala University), V. Ljosa, (Broad Institute of MIT and Harvard), A.L. Conery (Mass. General Hospital), P.Golland, (MIT), F.M. Ausubel (Mass. General Hospital), A.E.Carpente (Broad Institute of MIT and Harvard)

In the talk I will present a computational approach for extracting cluttered objects based on their morphological properties. Specifically, the problem of untangling C. elegant worms clusters in high-throughput microscopy images, is addressed. The key idea is based on the representation of each worm cluster by a sparse directed graph whose vertices and edges correspond to worm segments and their adjacencies, respectively. A combinatorial approach is then used to search and select paths in the graph that are most likely to represent individual worms while minimizing overlap. The worm likelihood measure is defined on a low-dimensional feature space that captures different worm poses, obtained from a training set of isolated worms. The work, which was done in collaboration with the imaging platform of the Broad institute of MIT and Harvard, is the foundation of one of its Cell Profiler toolboxes for high-throughput microscopy imaging analysis.

Shape Exaggeration and Auto-Caricaturization

Matan Sela, Yonathan Aflalo and Ron Kimmel (Technion)

We propose a computational approach for automatic caricaturization. The idea is to rely on intrinsic geometric properties of a given model that are invariant to poses, articulations, and gestures. A property of a surface that is preserved while it undergoes such deformations is self-isometry. In other words, while smiling, running, and posing, we do not change much the intrinsic geometry of our facial surface, the area of our body, or the size of our hands.
The proposed method locally amplifies the area of a given surface based on its

Gaussian curvature. It is shown to produce a natural comic exaggeration effect which can be efficiently computed as a solution of a Poisson equation. We demonstrate the power of the proposed method by applying it to a variety of meshes such as human faces, statues, and animals. The results demonstrate enhancement and exaggeration of the shape's features into an artistic caricature. As most poses and postures are almost isometries, the use of the Gaussian curvature as the scaling factor allows the proposed method to handle animated sequences while preserving the identity of the animated creature.