2016 Israel Computer Vision Day
Sunday, December 25, 2016

Sponsored this year by:

Vision Day Schedule

Time	Speaker and Collaborators	Affiliation	Title
08:50-09:20	Gathering		
09:20-09:40	Gerard Medioni Tal Hassner	USC OpenU	Faces, deep learning and the pursuit of training data
9:45-10:05	Elad Richardson Matan Sela Roy Or-El Ron Kimmel	Technion	Learning Detailed Face Reconstruction from a Single Image
10:10-10:30	Margarita Osadchy Julio Hernandez-Castro Stuart Gibson Orr Dunkelman Daniel Perez-Cabo	Haifa U of Kent U of Vigo	No Bot Expects the DeepCAPTCHA! Introducing Immutable Adversarial Examples, with Applications to CAPTCHA Generation
10:35-10:55	Nadav Cohen Amnon Shashua	HUJI	Inductive Bias of Deep Convolutional Networks through Pooling Geometry
11:00-11:30	Coffee Break		
11:30-11:50	Dan Feldman Soliman Nasser Ibrahim Jubran	Haifa	Low-cost and Faster Tracking Systems Using Core-sets for Pose-Estimation
11:55-12:15	Ronen Basri Soumyadip Sengupta Tal Amir Meirav Galun Tom Goldstein David Jacobs Amit Singer	Weizmann UMD Princeton	A New Rank Constraint on Multi-view Fundamental Matrices and its Application to Camera Location Recovery

12:20-12:40	Mike Werman Shmuel Peleg Ben Arzi Yoni Kesten Tavi Halperin	HUJI	Epipolar Geometry From Epipolar Lines
12:45-13:05	Ehud Barnea Ohad Ben-Shahar	BGU	High-Order Contextual Object Detection with a Few Relevant Neighbors
13:05-14:10	Lunch		
14:10-14:20	"Intermezzo"		
14:20-14:40	Yaniv Taigman Adam Polyak Lior Wolf	Facebook AI Research (FAIR), TLV	Unsupervised Cross-Domain Image Generation
14:45-15:05	Micha Lindenbaum Avi Kaplan Tamar Avraham	Technion	Interpreting the Ratio Criterion for Matching SIFT Descriptors
15:10-15:30	Or Litany Alex Bornstein Emanuele Rodolà Michael Bronstein	TAU USI Lugano	Fully spectral partial shape matching
15:35-15:55	Itamar Talmi Roey Mechrez Lihi Zelnik-Manor	Technion	Template Matching with Deformable Diversity Similarity
16:00-16:20	Coffee Break		
16:20-16:40	Netalee Efrat Piotr Didyk Mike Foshey Wojciech Matusik Anat Levin	Weizmann Max Planck Saarland U/MMCI MIT	Cinema 3D: Large Scale Automultiscopic Display
16:45-17:05	David Avidar David Malah Meir Barzohar	Technion	Point cloud registration using a viewpoint dictionary
17:10-17:30	Yael Moses Shachaf Melman Gerard Medioni Yinghao Cai	IDC USC	The Multi-Strand Graph for a PTZ Tracker

Abstracts

Faces, deep learning and the pursuit of training data

Gerard Medioni and Tal Hassner – USC, OPENU

The abilities of machines to detect and recognize faces improved remarkably over the last few years. This progress can at least partially be explained by the sizes of the training sets used to train deep learning models: huge numbers of face images downloaded and manually labeled. It is not clear, however, if the formidable task of collecting and labeling so many images is truly necessary or even effective. I will discuss the problems of data collection and describe a number of effective techniques for maximizing deep learning capabilities when collecting additional data is not an option. Importantly, though this talk will focus on face processing related tasks, these techniques can be applied in other image understanding problems where obtaining enough labeled examples for training deep learning systems is hard.

Learning Detailed Face Reconstruction from a Single Image

Elad Richardson, Matan Sela, Roy Or-El and Ron Kimmel - Technion

Reconstructing the detailed geometric structure of a face from a given image is a key to many computer vision and graphics applications, such as motion capture and reenactment. The reconstruction task is challenging as human faces vary extensively when considering expressions, poses, textures, and intrinsic geometry. While many approaches tackle this complexity by using additional data to reconstruct the face of a single subject, extracting facial surface from a single image remains a difficult problem. As a result, single-image based methods can usually provide only a rough estimate of the facial geometry. In contrast, we propose to leverage the power of convolutional neural networks to produce a highly detailed face reconstruction from a single image. For this purpose, we introduce a CNN framework which derives the shape in a coarse-to-fine fashion. The proposed architecture is composed of two main blocks, a network that recovers the coarse facial geometry (CoarseNet), followed by a CNN that refines the facial features of that geometry (FineNet). The proposed networks are connected by a novel layer which renders a depth image given a mesh in 3D. Unlike object recognition and detection problems, there are no suitable datasets for training CNNs to perform face geometry reconstruction. Therefore, our training regime begins with a supervised phase, based on synthetic images, followed by an unsupervised phase that uses only unconstrained facial images. The accuracy and robustness of the proposed model is demonstrated by both qualitative and quantitative evaluation tests.

No Bot Expects the DeepCAPTCHA! Introducing Immutable Adversarial Examples, with Applications to CAPTCHA Generation

Margarita Osadchy, Julio Hernandez-Castro, Stuart Gibson, Orr Dunkelman

Daniel Perez-Cabo – Haifa, U of Kent, U of Vigo

Recent advances in Deep Learning (DL) allow for solving complex AI problems that used to be considered very hard. While this progress has advanced many fields, it is considered to be bad news for CAPTCHAs (Completely Automated Public Turing tests to tell Computers and Humans Apart), the security of which rests on the hardness of some learning problems.

In this work we introduce DeepCAPTCHA, a new and secure CAPTCHA scheme based on adversarial examples, an inherit limitation of the current Deep Learning networks . These adversarial examples are constructed visual inputs, either synthesized from scratch or computed by adding a small and specific perturbation called adversarial noise to correctly classified items, causing the targeted DL network to misclassify them. We show that plain adversarial noise is insufficient to achieve secure visual CAPTCHA schemes, which leads us to introduce immutable adversarial noise --- an adversarial noise resistant to removal attempts. We implement a proof of concept system, and its analysis shows that the scheme offers high security and good usability compared to the best previously existing CAPTCHAs.

Inductive Bias of Deep Convolutional Networks through Pooling Geometry

Nadav Cohen and Amnon Shashua - HUJI

Our formal understanding of the inductive bias that drives the success of convolutional networks on computer vision tasks is limited. In particular, it is unclear what makes hypotheses spaces born from convolution and pooling operations so suitable for natural images. In this paper we study the ability of convolutional networks to model correlations among regions of their input. We theoretically analyze convolutional arithmetic circuits, and empirically validate our findings on other types of convolutional networks as well. Correlations are formalized through the notion of separation rank, which for a given partition of the input, measures how far a function is from being separable. We show that a polynomially sized deep network supports exponentially high separation ranks for certain input partitions, while being limited to polynomial separation ranks for others. The network's pooling geometry effectively determines which input partitions are favored, thus serves as a means for controlling the inductive bias. Contiguous pooling windows as commonly employed in practice favor interleaved partitions over coarse ones, orienting the inductive bias towards the statistics of natural images. Other pooling schemes lead to different preferences, and this allows tailoring the network to data that departs from the usual domain of natural imagery. In addition to analyzing deep networks, we show that shallow ones support only linear separation ranks, and by this gain insight into the benefit of functions brought forth by depth - they are able to efficiently model strong correlation under favored partitions of the input.

Low-cost and Faster Tracking Systems Using Core-sets for Pose-Estimation

Dan Feldman, Soliman Nasser and Ibrahim Jubran - Haifa

How can a $20 toy quadcopter navigate using a weak "Internet of Things" mini-computer and a web-cam? In the pose-estimation problem we need to align

(rotate+translate) a set of n marker (points) and choose one of their n! permutations, so that the sum of squared corresponding distances to another ordered set of markers is minimize.
We prove that every set has a weighted subset (core-set) of constant size (independent of n), such that computing the optimal orientation of the small core-set would yield exactly the same result as using the full set of n markers. A deterministic algorithm for computing this core-set in O(n) time is provided, using the Caratheodory Theorem from computational geometry.
We then developed a $50 tracking system based on this algorithm that turns a toy drone into an autonomous drone. The experimental results are almost identical to those obtained via a commercial $10,000 tracking system (OptiTrack).

A New Rank Constraint on Multi-view Fundamental Matrices and its Application to Camera Location Recovery

Ronen Basri , Soumyadip Sengupta, Tal Amir, Meirav Galun, Tom Goldstein, David Jacobs, Amit Singer – Weizmann, UMD, Princeton

Accurate estimation of camera matrices is an important step in structure from motion algorithms. In this paper we introduce a novel rank constraint on collections of fundamental matrices in multi-view settings. We show that in general, with the selection of proper scale factors, a matrix formed by stacking fundamental matrices between pairs of images has rank 6. Moreover, this matrix forms the symmetric part of a rank 3 matrix whose factors relate directly to the corresponding camera matrices. We use this new characterization to produce better estimations of fundamental matrices by optimizing an L1-cost function using Iterative Re-weighted Least Squares and Alternate Direction Method of Multiplier. We further show that this procedure can improve the recovery of camera locations, particularly in multi-view settings in which fewer images are available.

Epipolar Geometry From Epipolar Lines

Mike Werman, Shmuel Peleg, Ben Arzi, Yoni Kasten and Tavi Halperin - HUJI

The fundamental matrix is the basic building block of multiple view geometry and its computation is the first step in many vision tasks. Its computation is usually based on pairs of corresponding points. It is known that the fundamental matrix can also be computed from three matching epipolar lines. This was rarely used as there were no good methods to find these correspondences. Here we present 3 practical methods of finding such epipolar line correspondences resulting in superior fundamental matrices.

High-Order Contextual Object Detection with a Few Relevant Neighbors

Ehud Barnea and Ohad Ben-Shahar - BGU

A natural way to improve the detection of objects is to consider contextual constraints imposed by the detections of other objects in the scene. In this work we exploit the spatial relations between objects to improve a given set of detections and analyze the different properties of the problem in an exact probabilistic setting. In contrast to previous methods that are based on various complicated assumptions but typically focus on pairwise interactions only, here we employ a single realistic assumption that the

existence of an object at any given location is influenced by just few other relevant locations in space in order to facilitate a more exact calculation of object probability while using higher order interactions as well. We suggest a method for identifying these relevant locations and integrate them into an exact calculation of probability based on the raw detector responses. Among other insights, we argue that while it is generally difficult but possible to learn when an object reduces the probability of another, in many cases it is practically impossible to do for the task of improving the results of an object detector. We show that this also applies to some cases where an object greatly increases the probability of another, but that generally this occurs less than the former case. Finally, we demonstrate that the suggested approach improves detection results more than previous approaches over the challenging KITTI dataset.

Unsupervised Cross-Domain Image Generation

Yaniv Taigman, Adam Polyak and Lior Wolf – Facebook AI Research (FAIR), TLV

We study the problem of transferring a sample in one domain to an analog sample in another domain. Given two related domains, S and T, we would like to learn a generative function G that maps an input sample from S to the domain T, such that the output of a given function f, which accepts inputs in either domains, would remain unchanged. Other than the function f, the training data is unsupervised and consist of a set of samples from each domain. The Domain Transfer Network (DTN) we present employs a compound loss function that includes a multiclass GAN loss, an f-constancy component, and a regularizing component that encourages G to map samples from T to themselves. We apply our method to visual domains including digits and face images and demonstrate its ability to generate convincing novel images of previously unseen entities, while preserving their identity.

Interpreting the Ratio Criterion for Matching SIFT Descriptors

---

Micha Lindenbaum, Avi Kaplan, Tamar Avraham - Technion

Matching keypoints by minimizing the Euclidean distance between their SIFT descriptors is an effective and extremely popular technique. Using the ratio between distances, as suggested by Lowe, is even more effective and leads to excellent matching accuracy. Probabilistic approaches that model the distribution of the distances were found effective as well. This work focuses on analyzing Lowe's ratio criterion using a probabilistic approach.
We provide two alternative interpretations of this criterion, which show that it is not only an effective heuristic but can also be formally justified. The first interpretation shows that Lowe's ratio corresponds to a conditional probability that the match is incorrect. The second shows that the ratio corresponds to the Markov bound on this probability. The interpretations make it possible to slightly increase the effectiveness of the ratio criterion, and to obtain matching performance that exceeds all previous (non-learning based) results.

Fully spectral partial shape matching

Or Litany and Alex Bornstein, Emanuele Rodolà, Michael Bronstein –
Tau, USI Lugano

We propose an efficient procedure for calculating partial dense intrinsic correspondence between deformable shapes performed entirely in the spectral domain. Our technique relies on the recently introduced partial functional maps formalism and on the joint approximate diagonalization (JAD) of the Laplace-Beltrami operators previously introduced for matching non-isometric shapes. We show that a variant of the JAD problem with an appropriately modified coupling term (surprisingly) allows to construct quasi-harmonic bases localized on the latent corresponding parts. This circumvents the need to explicitly compute the unknown parts by means of the cumbersome alternating minimization used in the previous approaches, and allows performing all the calculations in the spectral domain with constant complexity independent of the number of shape vertices. We provide an extensive evaluation of the proposed technique on standard non-rigid correspondence benchmarks and show state-of-the-art performance in various settings, including partiality and the presence of topological noise.

Template Matching with Deformable Diversity Similarity

Itamar Talmi, Roey Mechrez, Lihi Zelnik-Manor - Technion

We propose a novel measure for template matching named Deformable Diversity Similarity – based on the diversity of feature matches between a target image window and the template. We rely on both local appearance and geometric information that jointly lead to a powerful approach for matching. Our key contribution is a similarity measure that is robust to complex deformations, significant background clutter, and occlusions. Empirical evaluation on the most up-to-date benchmark shows that our method outperforms the current state-of-the-art in its detection accuracy while improving computational complexity.

Cinema 3D: Large Scale Automultiscopic Display

Netalee Efrat, Piotr Didyk, Mike Foshey, Wojciech Matusik, Anat Levin – Weizmann
Max Planck, Saarland U/MMCI, MIT

3D movies are gaining popularity, viewers in a 3D cinema still need to wear cumbersome glasses in order to enjoy them. Automultiscopic displays provide a better alternative to the display of 3D content, as they present multiple angular images of the same scene without the need for special eyewear. However, automultiscopic displays cannot be directly implemented in a wide cinema setting due to variants of two main problems: (i) The range of angles at which the screen is observed in a large cinema is usually very wide, and there is an unavoidable tradeoff between the range of angular images supported by the display and its spatial or angular resolutions. (ii) Parallax is usually observed only when a viewer is positioned at a limited range of distances from the screen. This work proposes a new display concept, which supports automultiscopic content in a wide cinema setting. It builds on the typical structure of cinemas, such as the fixed seat positions and the fact that different rows are located on a slope at different heights. Rather than attempting to display many angular images spanning the full range of viewing angles in a wide cinema, our design only displays the narrow angular range observed within the limited width of a single seat. The same narrow range content is then replicated to all rows and seats in the cinema. To achieve this, it uses an optical construction based on two sets of parallax barriers, or lenslets, placed in front of a standard screen. This paper derives the geometry of such a display, analyses its limitations, and demonstrates a proof-of-concept prototype.

Point cloud registration using a viewpoint dictionary

David Avidar, David Malah and Meir Barzohar - Technion

The use of 3D point clouds is currently of much interest. One of the cornerstones of 3D point cloud research and applications is point cloud registration. Given two point clouds, the goal of registration is aligning them in a common coordinate system. In particular, we seek in this work to align a sparse and noisy local point cloud, created from a single stereo pair of images, to a dense and large-scale global point cloud, representing an urban outdoors environment. The common approach of keypoint-based registration, tends to fail due to the sparsity and low quality of the stereo local cloud. We propose here a new approach. It consists of the creation of a dictionary of much smaller clouds using a grid of synthetic viewpoints over the dense global cloud. We then perform registration via an efficient dictionary search. Our approach shows promising results on data acquired in an urban environment.

The Multi-Strand Graph for a PTZ Tracker

Yael Moses, Shachaf Melman, Gerard Medioni and Yinghao Cai – IDC, USC

We present a joint Deep Convolutional Neural Network and Support Vector Regression approach for estimating a person's age from a face. We start by leaning a robust face representation using deep network, followed by kernel-based support vector regression. We then show the age estimation accuracy can be further improved that by learning an age-related dimensionality reduction metric. The proposed schemes were successfully applied to the MORPH-II and FG-Net datasets outperforming contemporary state-of-the-art approaches.