2015 Israel Computer Vision Day
Sunday, April 17, 2016

Sponsored by:

Vision Day Schedule

Time	Speaker and Collaborators	Affiliation	Title
08:50-09:20	Gathering		
09:20-09:40	Yedid Hoshen Shmuel Peleg	HUJI	Visual Learning of Arithmetic Operations
9:45-10:05	Yaniv Romano Michael Elad	Technion	Con-Patch: When a Patch Meets its Context
10:10-10:30	Mor Dar Yael Moses	IDC	Temporal Epipolar Regions
10:35-10:55	Maria Kushnir Ilan Shimshoni	Haifa	A General Preprocessing Method for Improved Performance of Epipolar Geometry Estimation Algorithms
11:00-11:30	Coffee Break		
11:30-11:50	Nadav Cohen Or Sharir Amnon Shashua	HUJI	On the Expressive Power of Deep Learning: A Tensor Analysis
11:55-12:15	Yuval Bahat Michal Irani	Weizmann	Blind Dehazing Using Internal Patch Recurrence
12:20-12:40	Dana Berman Tali Treibitz Shai Avidan	Haifa + TAU	Non-Local Image Dehazing

12:45-13:05	Lior Wolf David Gadot	TAU	PatchBatch: a Batch Augmented Loss for Optical Flow
13:05-14:10	Lunch		
14:10-14:20	"Intermezzo"		
14:20-14:40	Aviad Levis Yoav Schechner Amit Aides	Technion	Airborne Three-Dimensional Cloud Tomography
14:45-15:05	Inbar Huberman Raanan Fattal	HUJI	Detecting Repeating Objects using Patch Correlation Analysis
15:10-15:30	Aaron Wetzler Ron Slossberg Ron Kimmel	Technion	Rule Of Thumb: Deep derotation for improved fingertip detection
15:35-15:55	Dan Levi Noa Garnett Ethan Fetaya	GM	StixelNet: A Deep Convolutional Network for Obstacle Detection and Road Segmentation
16:00-16:20	Coffee Break		
16:20-16:40	Leah Bar Nir Sochen Nahum Kiryati	TAU	Blind Restoration of Images with Piecewise Space-Variant Blur
16:45-17:05	Harel Haim Emanuel Marom Alex Bronstein	TAU	Computational multi-focus imaging combining sparse model with color dependent phase mask
17:10-17:30	Omry Sendik Yosi Keller	Bar Ilan	How old are you? DeepAge to the rescue

The Computer Vision Day is sponsored by:

Visual Learning of Arithmetic Operations

Yedid Hoshen and Shmuel Peleg - HUJI

A simple Neural Network model is presented for end-to-end visual learning of arithmetic operations from pictures of numbers. The input consists of two pictures, each showing a 7-digit number. The output, also a picture, displays the number showing the result of an arithmetic operation (e.g., addition or subtraction) on the two input numbers. The concepts of a number, or of an operator, are not explicitly introduced. This indicates that addition is a simple cognitive task, which can be learned visually using a very small number of neurons.

Other operations, e.g., multiplication, were not learnable using this architecture. Some tasks were not learnable end-to-end (e.g., addition with Roman numerals), but were easily learnable once broken into two separate sub-tasks: a perceptual Character Recognition and cognitive Arithmetic sub-tasks. This indicates that while some tasks may be easily learnable end-to-end, other may need to be broken into sub-tasks.

Con-Patch: When a Patch Meets its Context

Yaniv Romano and Michael Elad - Technion

Measuring the similarity between patches in images is a fundamental building block in various tasks. Naturally, the patch-size has a major impact on the matching quality, and on the consequent application performance. Under the assumption that our patch database is sufficiently sampled, using large patches (e.g. 21-by-21) should be preferred over small l ones (e.g. 7-by-7). However, this "dense-sampling" assumption is rarely true; in most cases large patches cannot find relevant nearby examples. This phenomenon is a consequence of the curse of dimensionality, stating that the database-size should grow exponentially with the patch-size to ensure proper matches. This explains the favored choice of small patch-size in most applications.
Is there a way to keep the simplicity and work with small patches while getting some of the benefits that large patches provide? In this work we offer such an approach. We propose to concatenate the regular content of a conventional (small) patch with a compact representation of its (large) surroundings -- its context. Therefore, with a minor increase of the dimensions (e.g. with additional 10 values to the patch representation), we implicitly/softly describe the information of a large patch. The additional descriptors are computed based on a self-similarity behavior of the patch surrounding.
We show that this approach achieves better matches, compared to the use of conventional-size patches, without the need to increase the database-size. Also, the effectiveness of the proposed method is tested on three distinct problems: (i) External natural image denoising, (ii) Depth image super-resolution, and (iii) Motion-compensated frame-rate up-conversion.

Temporal Epipolar Regions

Mor dar and Yael Moses – IDC

Dynamic events are often photographed by a number of people from different viewpoints at different times, resulting in an unconstrained set of images. Finding the corresponding moving features in each of the images allows us to extract information about objects of interest in the scene. Computing correspondence of moving features in such a set of images is considerably more challenging than computing correspondence in video due to possible significant differences in viewpoints and inconsistent timing between image captures. The prediction methods used in video for improving robustness and efficiency are not applicable to a set of still images. In this paper we propose a novel method to predict locations of an approximately linear moving feature point, given a small subset of correspondences and the temporal order of image captures. Our method extends the use of epipolar geometry to divide images into valid and invalid regions, termed Temporal Epipolar Regions (TERs). We formally prove that the location of a feature in a new image is restricted to valid TERs. We demonstrate the effectiveness of our method in reducing the search space for correspondence on both synthetic and challenging real world data, and show the improved matching.

A General Preprocessing Method for Improved Performance of Epipolar Geometry Estimation Algorithms

Maria Kushnir and Ilan Shimshoni - Haifa

In this paper a deterministic preprocessing algorithm is presented. It is especially designed to deal with repeated structures and wide baseline image pairs. It generates putative matches and their probabilities. They are then given as input to state-of-the-art epipolar geometry estimation algorithms, improving their results considerably, succeeding on hard cases on which they failed before. The algorithm consists of three steps, whose scope changes from local to global. In the local step, it extracts from a pair of images local features (e.g. SIFT), clustering similar features from each image. The clusters are matched yielding a large number of matches. Then pairs of spatially close features (2keypoint) are matched and ranked by a classifier. The highest ranked 2keypoint-matches are selected. In the global step, fundamental matrices are computed from each two 2keypoint-matches. A match's score is the number of fundamental matrices, which it supports. This number combined with scores generated by standard methods is given to a classifier to estimate its probability. The ranked matches are given as input to state-of-the-art algorithms such as BEEM, BLOGS and USAC yielding much better results than the original algorithms. Extensive testing was performed on almost 900 image pairs from six publicly available datasets.

On the Expressive Power of Deep Learning: A Tensor Analysis

Nadav Cohen, Or Sharir, and Amnon Shashua - HUJI

It has long been conjectured that hypotheses spaces suitable for data that is compositional in nature, such as images or text, may be more efficiently represented with deep hierarchical networks than with shallow ones. Despite the vast empirical evidence supporting this belief, theoretical analyses to date are limited. In particular, they do not account for the locality, sharing and pooling constructs of convolutional networks, the most successful deep learning architecture to date. In this work we

derive an equivalence between convolutional networks and hierarchical tensor decompositions. The type of decomposition corresponds to the structure of a network (depth, breadth, receptive fields), and the underlying algebraic operations correspond to the choice of activation and pooling operators.

Using tools from measure theory and tensor analysis, we show that linear activation and product pooling, corresponding to the SimNet architecture, lead to "complete depth efficiency", meaning that besides a negligible set, all functions that can be implemented by a deep network of polynomial size require exponential size in order to be implemented (or even approximated) by a shallow network. We then show that with rectified linear activation and max or average pooling, corresponding to standard convolutional neural networks, the expressive power deteriorates: average pooling leads to loss of universality, whereas max pooling brings forth incomplete depth efficiency. This leads us to believe that developing effective methods for training SimNets, thereby fulfilling their expressive potential, may give rise to a deep learning architecture that is provably superior to convolutional neural networks but has so far been overlooked by practitioners.

Blind Dehazing Using Internal Patch Recurrence

Yuval Bahat and Michal Irani - Weizmann

Images of outdoor scenes are often degraded by haze, fog and other scattering phenomena. In this work we show how such images can be dehazed using internal patch recurrence. Small image patches tend to repeat abundantly inside a natural image, both within the same scale, as well as across different scales. This behavior has been used as a strong prior for image denoising, super-resolution, image completion and more. Nevertheless, this strong recurrence property significantly diminishes when the imaging conditions are not ideal, as is the case in images taken under bad weather conditions (haze, fog, underwater scattering, etc.). In this work we show how we can exploit the deviations from the ideal patch recurrence for ``Blind Dehazing'' - namely, recovering the unknown haze parameters and reconstructing a haze-free image. We seek the haze parameters that, when used for dehazing the input image, will maximize the patch recurrence in the dehazed output image. More specifically, pairs of co-occurring patches at different depths (hence undergoing different degrees of haze) allow recovery of the Airlight color, as well as the relative-transmission of each such pair of patches. This in turn leads to dense recovery of the scene structure, and to full image dehazing.

Non-Local Image Dehazing

Dana Berman, Tali Treibitz and Shai Avidan – Haifa + TAU

Haze limits visibility and reduces image contrast in outdoor images. The degradation is different for every pixel and depends on the distance of the scene point from the camera. This dependency is expressed in the transmission coefficients, that control the scene attenuation and amount of haze in every pixel. Previous methods solve the single image dehazing problem using various patch-based priors. We, on the other hand, propose an algorithm based on a new, non-local prior. The algorithm relies on the assumption that colors of a haze-free image are well approximated by a few hundred distinct colors, that form tight clusters in RGB space. Our key observation is that pixels in a given cluster are often non-local, i.e., they are spread over the entire image plane and are located at different distances from the camera. In the presence of haze these

varying distances translate to different transmission coefficients. Therefore, each color cluster in the clear image becomes a line in RGB space, that we term a haze-line. Using these haze-lines, our algorithm recovers both the distance map and the haze-free image. The algorithm is linear in the size of the image, deterministic and requires no training. It performs well on a wide variety of images and is competitive with other state-of-the-art methods.

PatchBatch: a Batch Augmented Loss for Optical Flow

Lior Wolf and David Gadot - TAU

We propose new loss functions for learning patch based descriptors via deep Convolutional Neural Networks. The learned descriptors are compared using the L2 norm and do not require network processing of pairs of patches. The success of the method is based on a few technical novelties, including an innovative loss function that, for each training batch, computes higher moments of the score distributions. Combined with an Approximate Nearest Neighbor patch matching method and a flow interpolating method, state of the art performance is obtained on the most challenging and competitive optical flow benchmarks.

Airborne Three-Dimensional Cloud Tomography

Aviad Levis, Yoav Schechner, and Amit Aides - Technion

We seek to sense the three dimensional (3D) volumetric distribution of scatterers in a heterogenous medium. An important case study for such a medium is the atmosphere. Atmospheric contents and their role in Earth's radiation balance have significant uncertainties with regards to scattering components: aerosols and clouds. Clouds, made of water droplets, also lead to local effects as precipitation and shadows. Our sensing approach is computational tomography using passive multi-angular imagery. For light-matter interaction that accounts for multiple scattering, we use the 3D radiative transfer equation as a forward model. Volumetric recovery by inverting this model suffers from a computational bottleneck on large scales, which include many unknowns. Steps taken make this tomography tractable, without approximating the scattering order or angle range.

Detecting Repeating Objects using Patch Correlation Analysis

Inbar Huberman and Raanan Fattal - HUJI

In this paper we describe a new method for detecting and counting a repeating object in an image. While the method relies on a fairly sophisticated deformable part model, unlike existing techniques it estimates the model parameters in an unsupervised fashion thus alleviating the need for a user-annotated training data and avoiding the associated specificity. This automatic fitting process is carried out by exploiting the recurrence of small image patches associated with the repeating object and analyzing their spatial correlation. The analysis allows us to reject outlier patches, recover the visual and shape parameters of the part model, and detect the object instances efficiently.

In order to achieve a practical system which is able to cope with diverse images, we describe a simple and intuitive active-learning procedure that updates the object classification by querying the user on very few carefully chosen marginal classifications. Evaluation of the new method against the state-of-the-art techniques demonstrates its ability to achieve higher accuracy through a better user experience.

Rule Of Thumb: Deep derotation for improved fingertip detection

Aaron Wetzler, Ron Slossberg, and Ron Kimmel – Technion

We investigate a novel global orientation regression approach for articulated objects using a deep convolutional neural network. This is integrated with an in-plane image derotation scheme, DeROT, to tackle the problem of per-frame fingertip detection in depth images. The method reduces the complexity of learning in the space of articulated poses which is demonstrated by using two distinct state-of-the-art learning based hand pose estimation methods applied to fingertip detection. Significant classification improvements are shown over the baseline implementation. Our framework involves no tracking, kinematic constraints or explicit prior model of the articulated object in hand. To support our approach we also describe a new pipeline for high accuracy magnetic annotation and labeling of objects imaged by a depth camera.

StixelNet: A Deep Convolutional Network for Obstacle Detection and Road Segmentation

Dan Levi, Noa Garnett, Ethan Fetaya – General Motors

General obstacle detection is a key enabler for obstacle avoidance in mobile robotics and autonomous driving. We address the task of detecting the closest obstacle in each direction from a driving vehicle. As opposed to existing methods based on 3D sensing we use a single color camera. In our approach the task is reduced to a column-wise regression problem. The regression is then solved using a deep convolutional neural network (CNN). In addition, we introduce a new loss function based on a semi-discrete representation of the obstacle position probability to train the network. The network is trained using ground truth automatically generated from a laser-scanner point cloud. Using the KITTI dataset, we show that the our monocular-based approach outperforms existing camera-based methods including ones using stereo. We also apply the network on the related task of road segmentation achieving among the best results on the KITTI road segmentation challenge.

Blind Restoration of Images with Piecewise Space-Variant Blur

Leah Bar, Nir Sochen and Nahum Kiryati – TAU

We address the problem of a single image blind space-variant deblurring, where different parts of the image are blurred by different blur kernels. Assuming a region-wise space variant point spread function, a blur measure is defined followed by an evolving level set based segmentation procedure which extracts the regions. Then a blind kernel identification is carried out for each blur domain. We define a global space-variant deconvolution process which is stabilized by a unified common regularizer, thus preserving discontinuities between the differently restored image regions. Promising experimental results are presented for real images of two phase shift variant out of focus blur.

Computational multi-focus imaging combining sparse model with color dependent phase mask

Harel Haim, Emanuel Marom and Alex Bronstein - TAU

A method for extended depth of field imaging based on image acquisition through a thin binary phase plate followed by fast automatic computational post-processing is presented. By placing a wavelength dependent optical mask inside the pupil of a conventional camera lens, one acquires a unique response for each of the three main color channels, which adds valuable information that allows blind reconstruction of blurred images without the need of an iterative search process for estimating the blurring kernel. The presented simulation as well as capture of a real life scene show how acquiring a one-shot image focused at a single plane, enable generating a de-blurred scene over an extended range in space.

How old are you? DeepAge to the rescue

Omry Sendik and Yosi Keller – Bar-Ilan University

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.