

Publishable summary

Project objectives

Human actions can be categorized in many ways, one of which splits actions into a hierarchy ranging from basic atomic actions, such as “reach”, “grab” or “walk”, to complex activities such as “dining”, “exercising” or “having a meeting”. In this work I was mostly interested in basic motions. However, basic motions can change their semantic meaning depending on their context. For example, a person grabbing a cup is most likely about to drink, while if the grabbed object is a wrench, the person is more likely to be fixing something. To distinguish between drinking, eating and smelling flowers one must consider the context of the action, which changes our understanding of the picture content. Our research in this project aimed at providing the required building blocks for recognizing actions and their context. We have developed both advanced high-level models, as well as basic tool required for pre-processing.

Work performed and its main results

In this period our research continues to develop methods for consistent image segmentation. Additionally, we have developed methods for capturing context in human motion analysis in video. Furthermore, we have made progress in our methods for fast search in large databases of features and compact representations for image features. We elaborate next.

Consistent image segmentation of image collections and video:

In this work we propose a probabilistic framework for carrying out segmentation and recognition simultaneously. The framework combines an LDA ‘bag of visual words’ model for recognition, and a hybrid parametric-nonparametric model for segmentation. If applied to a collection of images, our framework can simultaneously discover the segments of each image, and the correspondence between such segments. Such segments may be thought of as the ‘parts’ of corresponding objects that appear in the image collection. Thus, the model may be used for learning new categories, detecting/classifying objects, and segmenting images. In this period we have published the following papers:

- M. Andreetto, L. Zelnik-Manor and P. Perona, “Unsupervised Learning of Categorical Segments in Image Collections”, Accepted for publication in PAMI, 2012.

Human motion analysis and incorporating context

In the second period we have worked on advanced methods for human action recognition. For human action recognition we chose to follow the Bag-of-Words approach which is popular also in other applications in computer vision. In this approach the underlying assumption is that every video clip showing an action can be viewed as an unordered collection of “words”. These “words” are typically features capturing local appearance and motion patterns of pixels in the video clip. These models have been previously shown to be effective for recognizing action such as walking or waving. Interactions with objects typically consist of an ordered set of atomic motions. To extend the applicability of Bag-Of-Words methods to recognition of object guided actions, we further proposed to incorporate the temporal order into the model. We have developed a model which combines the underlying ideas of bag-of-words models with temporal context. Our model captures the temporal order of sub-actions in multiple temporal scales. Our experiments show this leads to improved action recognition results. This work was published in:

- T. Glaser and L. Zelnik-Manor, “Incorporating Temporal Context in Bag-of-Words Models”, The 3rd IEEE Workshop on Video Event Categorization, Tagging and Retrieval for Real-World Applications, 2011.

Feature and feature databases in video

Subspaces offer convenient means of representing information in many Pattern Recognition, Machine Vision, and Statistical Learning applications. Contrary to the growing popularity of subspace representations, the problem of efficiently searching through large subspace databases has received little attention in the past. Hence we have presented a general solution to the Approximate Nearest Subspace search problem. Our solution uniformly handles cases where both query and database elements may differ in dimensionality, where the database contains subspaces of different dimensions, and where the queries themselves may be subspaces. Our tests indicate that an approximate nearest subspace can be located significantly faster than the nearest subspace, with little loss of accuracy. In this period we have published the following papers:

- R. Basri, T. Hassner and L. Zelnik-Manor, "Approximate Nearest Subspace Search", IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), Vol. 33, No. 2, pp. 266-278, 2011.

In addition, we have further developed novel methods for learning a dictionary that leads to sparse representations of signals which are known to reside on a union of subspaces. Two separate algorithms have been proposed. One that trains a sparsifying dictionary and another that learns a sensing matrix for obtaining compact sparse representations. This work was published in:

- K. Rosenblum, L. Zelnik-Manor and Y. Eldar, "Dictionary Optimization for Block-Sparse Representations", AAAI Fall 2010 Symposium on Manifold Learning.
- L. Zelnik-Manor, K. Rosenblum and Y. C. Eldar, "Sensing Matrix Optimization for Block-Sparse Decoding", IEEE Transactions on Signal Processing, Vol. 59, No. 9, 4300-4312, Sep. 2011.