Objects, Actions, Places*

Stephen J. McKenna
School of Computing, University of Dundee
Dundee DD1 4HN, UK

Jesse Hoey
David R. Cheriton School of Computer Science
University of Waterloo, 200 University Avenue West
Waterloo, ON, Canada N2L 3G1

Emanuele Trucco
School of Computing, University of Dundee
Dundee DD1 4HN, UK

January 2014

Computer vision has evolved dramatically in the last 20 years or so, especially in the domains of classification and recognition. With the advent of a usable internet, algorithms dealing with hundreds or thousands of images had to be transformed or replaced by ones dealing with millions or tens of millions. The internet has made crowd sourcing possible, reducing dramatically the time and cost needed for some semi-automatic tasks such as annotating large numbers of images. However, annotation remains a bottleneck for many applications, a topic addressed by a paper in this issue using active learning [2]. Simultaneously, higher computer power has become available at lower and lower prices via hardware solutions (GPUs, for example) and access to low-cost or even free high-throughput computing facilities (cloud computing, for example). There has never been such a convergence of huge opportunities and steep challenges for vision algorithm development. It is safe to say that the computer vision community has risen to these challenges and embraced the opportunities. The eight papers in this issue offer an intriguing cross-section of the state-of-the-art. Each of them has been carefully selected after multiple review cycles. Between them they deal with important issues in learning and feature representation, propose novel methods for recognition of objects, actions and places, and cast new light on time-honoured algorithms such as the Hough transform [8].

In *Object and Action Classification with Latent Window Parameters*, Bilen

*This is an author-created version of an editorial accepted for publication in the International Journal of Computer Vision. The final publication is available at http://link.springer.com. DOI: 10.1007/s11263-014-0699-3

et al. [1] propose a technique using latent support vector machines to learn sub-optimal, adaptive spatial divisions for object categorisation and action recognition. Several splitting models are considered. The technique does not need bounding boxes in the training data set and experiments demonstrate good performance when compared with spatial pyramid matching. The goal of efficient object classification is also addressed by Lehmann *et al.* [4] who propose a *branch and rank* method. Efficiency is achieved by partitioning the search space; the method learns a ranking function that compares candidate sets of windows and locates the most promising set to explore first. As the method requires only a few classifier runs, the authors can use non-linear kernels to improve performance and robustness.

Recognition under occlusion is tackled by Ren *et al.* [7] in *Regressing Local to Global Shape Properties for Online Segmentation and Tracking.* They propose a method for shape recovery under occlusion using a training set without occlusion. A 2D discrete cosine transform is used to estimate occluded low-frequency shapes from high-frequency harmonics that are not occluded. A locally weighted projection regression learns the mapping and has the advantages of being online and incremental.

In *Detecting People Looking at each other in Videos*, Marin-Jimenez *et al.* [6] use estimates of head pose to derive gaze volumes and thus determine whether peoples' eyelines match. They use Gaussian process regression models based on histograms of oriented gradients (HOG) to infer pitch and yaw estimates along with their uncertainty. Three measures based on these estimates are compared on a TV human interactions dataset. Annotations specifying which shots contain people looking at each other, as well as the trained head detector used in their experiments, are made available for other researchers to use.

The problem of recognizing outdoor places under changing conditions is tackled by Johns and Yang in their paper *Generative Methods for Long-Term Place Recognition in Dynamic Scenes* [3]. A certain building may look significantly different after a renovation whilst a green space in summer may look different in winter, for example. Spatio-temporal properties of each landmark are incrementally learned over time, making scene models robust to local changes. A new bag-of-words filtering approached is used along with a geometric verification scheme.

When building classifiers for recognition, obtaining class labels for training and validation is often the most labour intensive step. Active learning, in which examples are automatically selected for labelling during learning, offers one way to ease this bottleneck. In *Active Rare Class Discovery and Classification using Dirichlet Processes*, Haines and Xiang [2] consider the use of active learning with datasets that are highly imbalanced and contain examples of as yet unknown, rare classes. Their contribution is a criterion for active learning that balances the goals of obtaining good classification and discovering these rare classes. This is achieved using a Dirichlet process assumption to enable the probability of class membership for known classes to be estimated as well as the probability of belonging to a new, unknown class. The probability that an example will be misclassified is computed and used to select the next example for labelling.

They test their method, which is relatively simple to implement, on a wide range of machine learning and computer vision data sets.

Woodford *et al.* [8] approach one of the evergreens of computer vision, the Hough transform. Their paper *Demisting the Hough Transform for 3D Shape Recognition and Registration* proposes some new and interesting extensions leading to a competitive algorithm. They achieve linear complexity assuming that the Hough space is sparse in which case only some regions need sampling. They also observe that only one vote per feature is actually correct which allows them to minimize the entropy of the Hough space.

Finally, Liu *et al.* [5] take a principled, analytical approach to rotation-invariant feature extraction using HOG-like features. The main idea is to treat gradient histograms as continuous functions using polar coordinates (in 2D) or spherical harmonics (in 3D) and to represent them using a Fourier basis. The formulation for 3D volumetric images is of particular importance because alternative approaches to rotation invariance, based on pose normalisation or sampling, become unattractive in 3D where three angles are needed to specify pose. The practical utility of the approach is evidenced by experiments on three applications: car detection in aerial images, 3D shape retrieval, and voxel labelling in plant root images.

References

[1] H. Bilen, V. P. Namboodiri, and L. J. van Gool. Object and action classification with latent window parameters. *International Journal of Computer Vision*, 2014.

[2] T. S. F. Haines and T. Xiang. Active rare class discovery and classification using dirichlet processes. *International Journal of Computer Vision*, 2014.

[3] E. D. Johns and G.-Z. Yang. Generative methods for long-term place recognition in dynamic scenes. *International Journal of Computer Vision*, 2014.

[4] A. Lehmann, P. Gehler, and L. J. van Gool. Branch and rank for efficient object detection. *International Journal of Computer Vision*, 2014.

[5] K. Liu, H. Skibbe, T. Schmidt, T. Blein, K. Palme, T. Brox, and O. Ronneberger. Rotation-invariant HOG descriptors using fourier analysis in polar and spherical coordinates. *International Journal of Computer Vision*, 2014.

[6] M. J. Marin-Jimenez, A. Zisserman, M. Eichner, and V. Ferrari. Detecting people looking at each other in videos. *International Journal of Computer Vision*, 2014.

[7] C. Y. Ren, V. Prisacariu, and I. Reid. Regressing local to global shape properties for online segmentation and tracking. *International Journal of Computer Vision*, 2014.

[8] O. J. Woodford, M.-T. Pham, A. Maki, F. Perbet, and B. Stenger. Demisting the hough transform for 3D shape recognition and registration. *International Journal of Computer Vision*, 2014.