

Special Issue on Large-Scale Computer Vision: Geometry, Inference, and Learning

Roberto Cipolla · Carlo Colombo · Alberto Del Bimbo

Published online: 19 October 2014
© Springer Science+Business Media New York 2014

Computer vision is one of the most exciting areas of all information science and technology. It appeals both to the scientist looking for challenging research topics, and to the industrialist aiming at developing successful new products.

In the last few years, the proliferation of vision-related material (tutorials, publications, software, datasets, etc.) on the Internet has further developed the natural multidisciplinary call of our field, and created new occasions for cross fertilization. Today's research on computer vision is an original mix of mathematics, computer science, engineering, and physics, often taking inspiration from neighboring fields, such as the brain and behavioral sciences.

Technological advancements are also playing a crucial role in the rapid ripening of computer vision. The ever increasing performance of microprocessors and GPUs can support more and more complex software to run even in mobile, real-life scenarios. On the other hand, new generation data acquisition, storage and transmission devices can easily produce huge amounts of visual data such as high resolution images, videos, and 3D maps: Dealing with them effectively is a tremendous yet rewarding challenge, that must be met with careful data representations, powerful computational models and robust estimation techniques.

R. Cipolla
Department of Engineering, University of Cambridge, Cambridge
CB2 1PZ, England

C. Colombo (✉)
Computational Vision Group, Dipartimento di Sistemi e
Informatica, Università di Firenze, Via Santa Marta 3,
50139 Firenze, Italy
e-mail: carlo.colombo@unifi.it

A. Del Bimbo
Dipartimento di Sistemi e Informatica, Facoltà di Ingegneria –
Università degli Studi di Firenze, Via Santa Marta 3, 50137 Firenze,
FI, Italy

This special issue includes six carefully selected examples of current trends in pure and applied research in large-scale computer vision. The papers are all from renowned academic and industrial research groups scattered around the world—USA, Europe, Middle and Far East. The contributions cover different themes, from early vision to geometry and tracking, through visual recognition, learning and semantic segmentation.

In “Reconstructing the World’s Museums,” J. Xiao and Y. Furukawa offer a modern treatment of the problem of 3D reconstruction and visualization. They introduce a Constructive Solid Geometry representation consisting of volume primitives, in order to obtain well-regularized, texture-mapped three-dimensional maps of large-scale indoor environments. Although constructed from ground-level photographs and 3D laser points, the maps can be fully rendered from aerial viewpoints to improve fruition effectiveness.

The paper “People Watching: Human Actions as a Cue for Single View Geometry” by D.F. Fouhey, V. Delaitre, A. Gupta, A.A. Efros, I. Laptev and J. Sivic, combines in an original way the two traditional areas of scene reconstruction and visual recognition. The authors show that observing people performing different actions and suitably estimating body poses can be a powerful cue to the understanding of a 3D scene even when just a single view is available.

“Photo Sequencing” by T. Dekel, Y. Moses and S. Avidan addresses the difficult problem of temporally ordering a collection of still images taken asynchronously by a set of uncalibrated smartphone cameras. To this aim, static and dynamic features are first extracted from the images, which are used respectively to determine the relative geometry and produce a partial ordering for camera pairs. Rank aggregation is then used to combine the pairwise ordering into a globally consistent estimate of temporal order.

An important extension to the theory of MRFs and their use in early vision is proposed in “Filter-based Mean-Field Inference for Random Fields with Higher-Order Terms and Product Label-Spaces” by V. Vineet, J. Warrell and P.H.S. Torr. The authors show how to include higher-order terms in random field models in such a way that filter-based inference remains possible, and also extend their formulation to product label-space models. They demonstrate the efficiency of their approach on joint object-stereo labeling and object class segmentation.

A hot topic in learning and large-scale recognition is the optimization of classifiers for improving generalization performance while keeping low the computational cost. In “Low-Rank Bilinear Classification: Efficient Convex Optimization and Extensions,” T. Kobayashi proposes a convex optimization framework for bilinear classifiers based on trace norm minimization, which reduces the rank of the matrix with no approximations nor hard constraints on it. In addition, the paper proposes two novel extensions of the bilinear classifier in terms of multiple kernel learning and cross-modal learning.

The last paper of the issue, “ImageNet Auto-annotation with Segmentation Propagation” by M. Guillaumin, D. Küttel and V. Ferrari, focuses on the stimulating topic of semantic segmentation. The authors introduce ImageNet, a large-scale hierarchical database, and propose to automatically populate it starting from existing manual annotations, in the form of class labels and bounding boxes. The idea is to employ the images segmented so far to help segmenting new, unsegmented images. Segmentation propagation is based on semantic relationships, and is done both at the image level and at the class level.

We heartily wish all readers to enjoy the papers of this special issue, which we hope can be a source of inspiration for their future work and for the general progress of our fascinating discipline.

Acknowledgments Our heartfelt thanks go to Ms Courtney Clark of Springer USA for her kind help and assistance during the preparation of this special issue.