Int J Comput Vis (2015) 113:161–162
DOI 10.1007/s11263-015-0827-8

# Correspondence, Matching and Recognition

**Tilo Burghardt[1] · Dima Damen[1] · Walterio Mayol-Cuevas[1] · Majid Mirmehdi[1]**

Published online: 14 May 2015


The association of corresponding content across different visual representations and models is a fundamental task in many areas of computer vision. This special issue contains seven timely papers, all of which are concerned with solving a correspondence challenge, an associated matching or recognition task. The presented topics showcase some of the diversity in current computer vision research; they range widely from text recognition, motion segmentation, and cross-modal matching techniques to invariant descriptor construction, and aesthetic image analysis.

Pons-Moll et al. present work, which aims at inferring dense data-to-model correspondences. In their paper "Metric Regression Forests for Correspondence Estimation" (doi:10.1007/s11263-015-0818-9), they introduce a new decision forest training objective named Metric Space Information Gain (MSIG). They show that their methodology is a principled generalization of the proxy classification objective, which does not require an extrinsic isometric embedding of the model surface in Euclidean space. Backed by extensive experiments, the authors demonstrate that this leads to highly accurate associations, using few training images.

Matching structures in cases where no one-to-one correspondences, but only relative pairing information is available is addressed in the paper "Relatively-Paired Space Analysis: Learning A Latent Common Space from Relatively-Paired Observations" (doi:10.1007/s11263-014-0783-8) by Kuang et al. In their work they describe constructing a latent common space between different modalities for cross-modality pattern recognition using relatively-paired observations. To evaluate performance, they apply their framework to feature fusion, cross-pose face recognition, and text-image retrieval concluding superior performance above other state-of-the-art approaches.

In their paper "Label Embedding: A Frugal Baseline for Text Recognition" (doi:10.1007/s11263-014-0793-6), Rodriguez et al. suggest creating a single space of embedded word images and word labels to establish multimodal correspondences directly. Departing from traditional bottom-up approaches in text recognition, they propose to embed word labels and word images into a Euclidean space learned using a Structured SVM. With this in hand, they cast the text recognition problem as one of retrieval: given an image find the closest word label in the built space. The authors conclude that their approach can obtain results comparable to standard bottom-up approaches, establishing label embedding as an interesting and simple to compute baseline for text recognition.

"A Spline-Based Trajectory Representation for Sensor Fusion and Rolling Shutter Cameras" (doi:10.1007/s11263-015-0811-3), authored by Perez et al., establishes a continuous-time B-Spline trajectory model that corresponds to given camera measurements. Instead of a traditional, discrete-time pose representation, they present a formulation parameterized in the Lie Algebra of the group SE3. Experiments with visual-inertial SLAM show that the approach can also be used to calibrate entire camera systems.

In "Morphologically Invariant Matching of Structures with the Complete Rank Transform" (doi:10.1007/s11263-015-0800-6), Demetz et al. introduce two novel descriptors:

✉ Tilo Burghardt
tb2935@bristol.ac.uk

Dima Damen
dima.damen@bristol.ac.uk

Walterio Mayol-Cuevas
wmayol@cs.bris.ac.uk

Majid Mirmehdi
M.Mirmehdi@cs.bris.ac.uk

[1] Department of Computer Science, University of Bristol, MVB, Woodland Road, Bristol BS8 1UB, UK

the complete rank transform and the complete census transform. These are shown to provide robust structural correspondence information due to invariance under monotonically increasing intensity scaling. Experiments focus on the KITTI benchmark. The authors showcase robustness in relation to illumination changes and state-of-the-art performance.

Looking to extract pixel-to-region correspondences, Stückler et al. propose an efficient expectation maximization (EM) framework for dense 3D segmentation of moving rigid parts in RGB-D video in their paper "Efficient Dense Rigid-Body Motion Segmentation and Estimation in RGB-D Video" (doi:10.1007/s11263-014-0796-3). Their approach segments images into pixel regions that undergo coherent 3D rigid-body motion. The authors experimentally demonstrate that the approach recovers segmentation and 3D motion at good precision.

Finally, Murray et al. present their work in "Discovering Beautiful Attributes for Aesthetic Image Analysis" (doi:10.1007/s11263-014-0789-2) where they extract aesthetic properties of images by learning corresponding mid-level attributes nameable by humans. They propose to discover and learn the visual appearance of attributes from a recently introduced database, AVA, which contains more than a quarter of a million images together with their aesthetic scores and textual comments given by photography enthusiasts. The authors show that the learned mid-level attributes can be successfully used in aesthetic quality prediction, image classification and retrieval.

We extend our sincere thanks to all reviewers of the papers in this special issue, and hope that readers will find the papers put forward here inspiring and informative.