



## Editorial: Special Issue on Machine Vision

Edwin Hancock<sup>1</sup> · Richard Wilson<sup>1</sup> · William A. P. Smith<sup>1</sup> · Adrian Bors<sup>1</sup> · Nick Pears<sup>1</sup>

Published online: 26 October 2018  
© Springer Science+Business Media, LLC, part of Springer Nature 2018

This special edition of the *International Journal of Computer Vision* is devoted to a selection of the best papers submitted to the 2016 British Machine Vision Conference held in York, together with an account of the invited talk given by Katsushi Ikeuchi.

In total we received 365 submissions to the conference, of which 38 were accepted for oral presentation and 106 for poster presentation. Based on the reviewers' reports for the papers and feedback obtained at the conference, we invited the authors of 15 papers to submit extended versions to this Special Issue. These were subjected to the usual review process of IJCV, and resulted in nine selected papers.

The papers present a snapshot of some the best work in computer vision which reached fruition in the Spring of 2016, and cover a diverse set of topics. Several topics deserve note. As with most recent conferences in the area, there was keen interest in Deep Learning and Convolutional Neural Networks. Analysing people and their faces remains a pressing topic, with application in interfaces and biometrics. Finally, event-cameras which detect changes in pixel intensity with potentially very high equivalent frame rates proved topical too.

Below is a brief summary of the papers accepted.

Statistical face models have long been a popular topic at BMVC and there has been renewed interest and progress in the use of 3D models for 2D image analysis. Bernhard Egger,

Sandro Schönborn, Andreas Schneider, Adam Kortylewski, Andreas Morel-Forster, Clemens Blumer and Thomas Vetter present “Occlusion-aware 3D Morphable Models and an Illumination Prior for Face Image Analysis” which represents the state-of-the-art approach for analysis-by-synthesis. They introduce semantics by using a segmented model including a prototype-based beard model and explicitly model the background. They propose a robust and efficient method for illumination estimation, apply this to a large dataset and make the resulting learnt illumination prior publicly available as a resource to the community.

In “Person Re-Identification in Identity Regression Space” Hanxiao Wang, Xiatian Zhu, Shaogang Gong and Tao Xiang tackle the problem of re-identifying people from whole-body images in challenging realworld conditions. In particular, they address the problem of appearance change over time and scalability to large numbers of identities. This is done by formulating the task as an efficient regression problem in a space that allows incremental learning with human assistance.

Convolutional neural networks (CNN's) have had a major impact on computer vision over the past half decade or so. In “Deep Sign: Enabling Robust Statistical Continuous Sign Language Recognition via Hybrid CNN-HMMs”, Oscar Koller, Sepehr Zargaran, Hermann Ney and Richard Bowden present an end-to-end system where the output of a CNN can be used for statistical sign language recognition. While CNN's are often used as black-boxes, in this paper they are unpacked and their output used to construct a hidden Markov Model, which allows Bayesian classifiers to be constructed.

In the same vein, Isinsu Katircioglu, Bugra Tekin, Mathieu Salzmann, Vincent Lepetit and Pascal Fua seek to combine explicit modeling with deep learning in “Learning Latent Representations of 3D Human Pose with Deep Neural Networks.” Here, a method for estimating 3D human pose from 2D data is presented. The novelty lies in learning a latent pose representation that captures joint dependencies. To extend the method to video, they use a recurrent architecture to ensure temporal consistency.

---

Communicated by Edwin Hancock, Richard Wilson, William A. P. Smith, Adrian Bors, Nick Pears.

---

✉ William A. P. Smith  
William.Smith@york.ac.uk

Edwin Hancock  
erh@cs.york.ac.uk

Richard Wilson  
richard.wilson@york.ac.uk

Adrian Bors  
adrian.bors@york.ac.uk

Nick Pears  
nick.pears@york.ac.uk

<sup>1</sup> University of York, York, UK

The paper “Combining Shape from Shading and Stereo: A Joint Variational Method for Estimating Depth, Illumination and Albedo” by Daniel Maurer, Yong Chul Ju, Michael Breuß and Andrés Bruhn offer a new approach to depth estimation by fusing stereo and shading information. Conventional work in this area usually commences from an estimate of depth obtained from stereo pairs, and then refines this using shading information. In this work on the other hand, the process of surface height recovery is posed a joint variational problem, which allows for the recovery of depth, illumination and albedo maps.

Biologically inspired edge detection is a classical problem in computer vision that has attracted interest since the seminal work of Marr and Hildreth. Arash Akbarinia and Alejandro Parraga return to the problem in “Feedback and Surround Modulated Boundary Detection”. They develop models based on surround modulation in both V1 and V2, and allow feedback between the two layers. In V1 contrast-dependent responses are obtained, which are pooled according to orientation in V2. Feedback from V2 to V1 is shape sensitive. The method fits with recent findings in visual neuroscience, and offers good performance.

Event cameras sense the change in intensity at a pixel, rather than the intensity itself, and offer exciting new possibilities for high speed imaging. They are capable of capturing information about sparse events at mega-Hertz rates. They thus present new challenges to computer vision. Gottfried Munda, Christian Reinbacher and Thomas Pock address the problem of how to reconstruct the intensity values in the paper “Real-Time Intensity-Image Reconstruction for Event Cameras Using Manifold Regularisation”. Their approach exploits the concept of an event manifold to regularize the recovery of intensity images via an energy minimization approach. The method offers impressive image recovery.

A second by paper on event cameras describes how they can be used for real-time multiview stereo. In “EMVS: Event-based Multi-View Stereo—3D Reconstruction with an Event Camera in Real-Time”, Henri Rebecq, Guillermo Gallego, Elias Mueggler and Davide Scaramuzza show how a single event camera can be used for depth estimation without reconstructing the intensity image. To generate events associated with changes in depth, the camera is scanned across the scene. This generates an event stream that can be processed to reveal depth, without the need to solve the correspondence problem.

Finally, Katsushi Ikeuchi, Zengqiang Yan, Zhaoyuan Ma, Yoshihiro Sato, Minako Nakamura and Shunsuke Kudoh present “Describing upper-body motions based on labanotation for learning-from-observation robots,” extending the work presented in Ikeuchi’s keynote presentation. This fascinating line of work traces its roots right back to the ground-breaking robot reasoning and action research of the late 1960s. Here, the specific goal is for a robot to learn motion sequences by observing a human performance. In the keynote talk, this was motivated by the specific application of preserving intangible cultural heritage such as traditional dance. The novel contribution is to use an intermediate representation, labanotation, which can be used to map human motion onto a physically-realizable equivalent in the robot.

We thank the reviewers who took time to read several versions of the submitted manuscripts, and the editorial staff at Springer, who showed great patience as we progressed the sometimes laboriously slow process of acquiring reviews.

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.