



Special Issue on Machine Vision

Tae-Kyun Kim¹ · Stefanos Zafeiriou² · Ben Glocker² · Stefan Leutenegger²

Published online: 14 August 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

This special issue of the International Journal of Computer Vision (IJCV) presents the best papers of the 28th British Machine Vision Conference (BMVC) held in Imperial College London, UK, on 4th–7th September 2017. BMVC, since its first edition in 1990, has become a major conference in computer vision and related fields. BMVC 2017 was particularly successful, raising the bar in all aspects: the number of paper submissions and attendees, and the amount of industrial sponsorship were the highest in the history of BMVC. It received 635 full paper submissions, among them 188 were accepted (29.6% acceptance rate). 36 orals (5.6% acceptance rate), 20 spotlights, 132 poster papers were accepted. A full review process was carried out by 550 reviewers and 41 area chairs, offering a very strong program of accepted papers.

The special issue is by invitation only among the best accepted papers. Three award winning papers and three honourable mention papers of the conference were invited. The invitation was extended to selected oral and spotlight papers, based on the reviews and area chair award suggestions. Longer versions of 16 conference papers were submitted, and rigorously reviewed by at least three reviewers, who were assigned independently of the paper review process of the conference. In total, 12 papers were accepted for inclusion in this special issue.

The papers presented in this issue offer a snapshot of some of the best work in the field, on the topics of (1) learning

quantised representations by deep neural networks, (2) 3D shape estimation, (3) novel image representations, (4) image descriptors and matching, and (5) data generation.

Deep hashing and learning low-bit representations by DNNs are crucial to large-scale image retrieval and relevant problems, where time efficiency and low storage are required. Over the conventional projection methods, deep learning has been a promising alternative under supervised learning settings, but has a gap in unsupervised settings. Yuming Shen, Li Liu, and Ling Shao propose a novel method for unsupervised deep hashing by unveiling and exploiting data intrinsic structure using the conditional auto-encoding variational Bayesian networks. When representing data by low-bits, accuracy drop is generally non-trivial. While prior-arts are based on gradient optimisation, Yinpeng Dong, Renkun Ni, Jianguo Li, Yurong Chen, Hang Su, and Jun-Zhu present stochastic quantisation. The proposed method quantises a portion of elements/filters with a stochastic probability density, while keeping the other portion unchanged with full precision. The proposed methods are demonstrated over state-of-the-arts in various experiments.

Inferring 3D shapes from images is central to vision research. Quite diverse approaches and settings are covered in this special issue. Existing polarisation-based methods yield ambiguous shape estimation. The imaging formulation also differs for specular and diffuse reflection. Fotios Logothetis, Roberto Mecca, Fiorella Sgallari, and Roberto Cipolla derive a formulation that depends only on polarimetric images, allowing direct geometrical characterisation of the level-set of objects. Reflectance estimation in uncontrolled environments remains challenging, while shape estimation is more pervasive. Trung Thanh Ngo, Hajime Nagahara, Ko Nishino, Rin-ichiro Taniguchi, and Yasushi Yagi propose a solution that simultaneously recovers both reflectance and shape from natural illumination. They exploit a lightfield camera for objects and a 360-degree camera for ambient lighting. Olivia Wiles and Andrew Zisserman propose to predict 3D surfaces of sculptures by predicting depth and silhouette maps from multi-view images and their viewpoints. The

Communicated by Tae-Kyun Kim, Stefanos Zafeiriou, Ben Glocker and Stefan Leutenegger.

✉ Tae-Kyun Kim
tk.kim@imperial.ac.uk
Stefanos Zafeiriou
s.zafeiriou@imperial.ac.uk
Ben Glocker
b.glocker@imperial.ac.uk
Stefan Leutenegger
s.leutenegger@imperial.ac.uk

¹ Department of EEE, Imperial College London, London, UK

² Department of Computing, Imperial College London, London, UK

proposed network learns the visual hull of objects. It is trained to predict the silhouette of objects. The added depth term in the loss function helps capture concavities of 3D shapes.

Novel image representations are proposed by two papers. Deformable part-based representations for object detection and recognition were revisited by Taylor Mordan, Nicolas Thome, Gilles Henaff, and Matthieu Cord. They propose a fully convolutional network that learns to align parts in end-to-end. The method does not require part annotations. Interactions between parts are explicitly learned, and alignment is done with an in-network optimization. This paper received the best science paper award of BMVC17. Daniel Hernandez-Juarez, Lukas Schneider, Pau Cebrian, Antonio Espinosa, David Vazquez, Antonio Manuel Lopez, Uwe Franke, Marc Pollefeys, and Juan Carlos Moure present a novel scene representation based on Stixels. The new representation introduces a depth model to account for non-flat roads and slanted objects. Both geometric and semantic cues are used to infer the scene representation, and their optimization method achieves real-time performance. This work received the best industrial paper award of BMVC17.

Image descriptors and matching by local descriptors or deep neural networks remains an active research topic. Arun Mukundan, Giorgos Toliás, Andrei Bursuc, Herve Jegou, and Ondrej Chum propose a multiple-kernel local-patch descriptor based on efficient match kernels from pixel gradients. It combines Polar and Cartesian parametrization for errors in the patch dominant orientation and the feature point location respectively. Combined with whitening of the descriptor space, the performance competes with deep learning methods. The work of Bailey Kong, James Supancic, Deva Ramanan, and Charless Fowlkes is application-focused. The work tackles an interesting application i.e. shoe print recognition. The data variability e.g. traces of dust or oil on hard surfaces throws challenges for cross-domain image matching. They found the mid-level feature descriptors from pre-trained CNNs surprisingly effective for this specific domains. Multi-channel normalized cross-correlation is also proposed, its impact on the new descriptor is analysed.

Image/video synthesis, powered by deep learning and adversarial learning, is one of hottest research directions at present. Numerous works newly appear for novel applications and methodological breakthroughs on data augmentation. Amir Jamaludin, Joon Son Chung, and Andrew Zisserman present an encoder-decoder Convolutional Neural Network (CNN) that learns using a joint embedding of face and audio. The proposed method generates a video of a target face lip synched with an audio speech segment in real time, taking still face images as inputs. The encoder-decoder architecture has been widely used. In such an architecture, while encoders have been intensively studied, relatively

few studies address the decoder aspect. The work of Zbigniew Bogdan Wojna, Vittorio Ferrari, Sergio Guadarrama, Nathan Silberman, Liang-Chieh Chen, Alireza Fathi, and Jasper Uijlings, offers an extensive comparative study of various decoders on pixel-wise tasks: from classification, regression to synthesis by GANs. To this end, they propose new residual-like connections for decoders and a novel bilinear additive upsampling decoder. Mariko Isogawa, Dan Mikami, Kosuke Takahashi, Daisuke Iwai, Kosuke Sato, and Hideaki Kimata tackles the evaluation perspective of image inpainting problems. Image inpainting, which removes and restores unwanted regions in images, can be cast as a data generation problem. How to evaluate such methods is less clear, often resorting to subjective criteria. They propose a learning-based image quality assessment framework, where simulated results of inpainted images with controlled qualities are exploited as training data.

Overall, the papers in the special issue illustrate the breadth of ongoing challenges in the field, and each work presents in-depth analyses and novel ideas/methods on a topic in the highest quality. We thank all the authors, and the reviewers who carefully read papers and submitted their thorough reviews at multiple rounds. We would also like to thank the editorial staffs at Springer who helped the process of preparing the special issue. Below is the list of accepted papers.

Accepted Papers

1. **Unsupervised Binary Representation Learning with Deep Variational Networks**
Yuming Shen, Li Liu, and Ling Shao
2. **Stochastic Quantization for Learning Accurate Low-bit Deep Neural Networks**
Yinpeng Dong, Renkun Ni, Jianguo Li, Yurong Chen, Hang Su, and Jun Zhu
3. **Slanted Stixels: A way to represent steep streets**
Daniel Hernandez-Juarez, Lukas Schneider, Pau Cebrian, Antonio Espinosa, David Vazquez, Antonio Manuel Lopez, Uwe Franke, Marc Pollefeys, Juan Carlos Moure
4. **End-to-End Learning of Latent Deformable Part-based Representations for Object Detection**
Taylor Mordan, Nicolas Thome, Gilles Henaff, and Matthieu Cord
5. **A Differential Approach to Shape from Polarisation: a Level-Set Characterisation**
Fotios Logothetis, Roberto Mecca, Fiorella Sgallari, and Roberto Cipolla
6. **The Devil is in the Decoder: Classification, Regression and GANs**
Zbigniew Bogdan Wojna, Vittorio Ferrari, Sergio Guar-

- rama, Nathan Silberman, Liang-Chieh Chen, Alireza Fathi, and Jasper Uijlings
7. **Reflectance and Shape Estimation with a Light Field Camera under Natural Illumination**
Trung Thanh Ngo, Hajime Nagahara, Ko Nishino, Rin-ichiro Taniguchi, and Yasushi Yagi
 8. **Understanding and Improving Kernel Local Descriptors**
Arun Mukundan, Giorgos Toliás, Andrei Bursuc, Herve Jegou, and Ondrej Chum
 9. **Cross-Domain Image Matching with Deep Feature Maps**
Bailey Kong, James Supancic, Deva Ramanan, and Charless Fowlkes
 10. **Which is the better inpainted image? Training data generation without any manual operations**
Mariko Isogawa, Dan Mikami, Kosuke Takahashi, Daisuke Iwai, Kosuke Sato, and Hideaki Kimata
 11. **You said that?: synthesising talking faces from audio**
Amir Jamaludin, Joon Son Chung, and Andrew Zisserman
 12. **Learning to Predict 3D Surfaces of Sculptures from Single and Multiple Views**
Olivia Wiles, and Andrew Zisserman

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