Machine Vision and Applications (2018) 29:929–931
https://doi.org/10.1007/s00138-018-0948-5

EDITORIAL

Machine learning for big visual analysis

Jun Yu[1] · Xue Mei[2] · Fatih Porikli[3] · Jason Corso[4]

Published online: 23 June 2018

1 Introduction

With the proliferation of smartphones and handheld devices, the amount of visual data in our world has been exploding. Visual data have become so large and complex that the traditional data processing applications are inadequate to deal with it. As a result, the emergence of big visual data has brought a paradigm shift to many fields of visual data analysis. Visual data are not only big in volume and size, but also can be unstructured, incomplete, noisy, redundant, and heterogeneous. A variety of methods have been developed for addressing numerous challenging problems. However, many problems remain challenging especially on the scalability of computationally complex algorithms, the shortage of accurately annotated raw data, the issue of integrating heterogeneous data from different sources, the difficulty in discovering valuable knowledge from noisy and redundant data.

This special issue aims to demonstrate the contribution of machine learning techniques to the research and development of big visual data analysis. Many machine learning techniques have already been applied to address the relevant problems. For example, convolutional neural networks have demonstrated superior performance on large-scale image classification. Semi- and weakly supervised learning methods have significantly improved the performance when only small amount of annotated data are available. Correlation analysis, transfer learning, and multitask learning have shown the potential in integrating severely heterogeneous data. Sparse representation and clustering approaches have been exploited in denoising and selecting of exemplary samples from the raw data.

✉ Jun Yu
zju.yujun@gmail.com

1 School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou 310018, China

2 Toyota Research Institute, Ann Arbor, USA

3 Australian National University, Canberra, Australia

4 University of Michigan, Ann Arbor, USA

Many algorithms have already been successfully applied in this research area, e.g., sparse coding has been successfully used for visual object recognition that models human visual system; multitask learning can efficiently achieve neural generative question answering; discriminant analytical least squares metric learning has shown its excellent performance in person re-identification; and sparse representation has been efficiently used in face recognition. In total, we received 23 submissions from all around the world. The submissions cover a wide variety of areas including large-scale face recognition, neural generative question answering, and visual object recognition. After two rounds of vigorous review by at least two expert reviewers for each paper, we finally selected 9 high-quality articles to be included in this highly popular special issue.

2 Overview of articles

We give an overview of all the included papers summarizing the contributions and novel aspects of each paper. All the articles are divided into 3 subsections roughly according to their areas, namely deep learning for visual analysis, sparse learning for visual analysis and other related topics.

2.1 Deep learning for visual analysis

The article entitled "Panchromatic and multi-spectral image fusion for new satellites based on multi-channel deep model" proposed a novel method based on the multi-channel deep model to fuse images for new satellites. The deep model is implemented by convolutional neural networks and trained on each band to reduce the impact of spectral range mismatch. The proposed method also preserves the detailed information in multi-spectral images, which is ignored by the traditional methods. It also effectively alleviates the inconvenience caused in the data augmentation processing for remote sensing images. Visual and quantitative assessments of fusion results show that the proposed method clearly

improves the fusion quality compared to the state-of-the-art methods.

The article entitled "Two-stream Person Re-identification with Multi-task Deep Neural Networks" proposed a two-stream strategy to use parts and bodies simultaneously. It utilizes a multitask learning framework with deep neural networks (DNNs). Part detection and body recognition are performed as two tasks, and the features are extracted by two DNNs. The features are connected to multitask learning to compute the mapping model from features to identifications. With this model, re-id can be achieved. Experimental results on a challenging task show the effectiveness of the proposed method.

The article entitled "Hierarchical convolutional features for end-to-end representation based visual tracking" proposed a novel simple tracker with deep learning to complete the tracking task. A simple fully convolutional Siamese network is applied to capture the similarity between different frames. Nevertheless, the detailed information from lower layers, which is also important for locating the target object, is not considered into the tracking task. In this paper, the detailed information from two lower layers is considered into the response map to improve the performance and not to increase much time spent. This leads more significant improvement for feature representation and localization of the target object. The experimental results demonstrate that the proposed algorithm is efficient and robust compared with the baseline and the state-of-the-art trackers.

2.2 Sparse learning for visual analysis

The article entitled "Biological Modeling of Human Visual System for Object Recognition using GLoP Filters and Sparse Coding on Multi-manifolds" proposed an extended HMAX model, denoted as E-HMAX, by the following biologically plausible ways. First, contrast normalization is conducted on the input image to simulate the processing of human retina and LGN. Second, Log Polar Gabor (GLoP) filters are used to simulate the properties of V1 simple cells instead of Gabor filters. Then, sparse coding on multi-manifolds (SCMM) is modeled to compute the V4 simple cell response instead of Euclidean distance. Meanwhile, a template learning method based on dictionary learning on multi-manifolds (DLMM) is proposed to select informative templates during template learning stage. Experimental results demonstrate that the proposed model has greatly outperformed the standard HMAX model. It is also comparable to some state-of-the-art approaches such as EBIM and OGHM-HMAX.

The article entitled "Rank-Sparsity Balanced Representation for Subspace Clustering" proposed a new model that can balance the rank and sparsity well. This model adopts the log-determinant function to control the rank of solution. Meanwhile, the diagonals are penalized, rather than the strict zero restriction on diagonals. This strategy makes the rank-sparsity balance more tunable. Numerical experiments show that the new method, named as RSBR, can significantly increase the accuracy of subspace clustering on the real-world datasets.

The article entitled "An Extended Sparse Representation based Classification Method for Face Recognition" proposed a new sparse representation-based classification method which can strengthen the discriminative property of different classes and obtain a better representation coefficient vector. Authors introduced a weighted matrix, which can make small deviations correspond to higher weights and large deviations correspond to lower weights. Meanwhile, they improve the constraint term of representation coefficients, which can enhance the distinctiveness of different classes and make a better positive contribution to classification. In addition, motivated by the work of ProCRC algorithm, they take into account the deviation between the linear combination of all training samples and of each class. Thereby, the discriminative representation of the test sample is further guaranteed. Experimental results on the ORL, FERET, Extended_YaleB, and AR databases show that the proposed method has better classification performance than other methods.

2.3 Other related topics

The article entitled "Multi-Task Learning for Neural Generative Question Answering" investigated multitask learning (MTL) in neural network based method under a QA scenario. They define the main task as a generative QA via Seq2Seq learning and define the auxiliary task as a discriminative QA via binary QA classification. Both main task and auxiliary task are learned jointly with shared representations, allowing to obtain improved generalization and transferring classification labels as extra evidences to guide the word sequence generation of the answers. Experimental results on both automatic evaluations and human annotations demonstrate the superiorities of the proposed method over baselines.

The article entitled "Person Re-identification by Discriminant Analytical Least Squares Metric Learning" proposed a new metric learning method based on least squares for person re-identification. Specifically, the similar training images pairs are used to learn a linear transformation matrix by being projected to finite discrete discriminant points using regression model, and then, the metric matrix can be deduced by solving least squares problem with a closed form solution. In addition, authors develop the incremental learning scheme of DALS, which is particularly valuable in model retraining when given additional samples. Furthermore, DALS could be effectively kernelized to further improve the matching performance. Extensive experiments on the VIPeR, GRID,

PRID450S, and CUHK0 datasets demonstrate the effectiveness and efficiency of the approaches.

The article entitled "Distributed Kalman filter based on Metropolis-Hastings sampling strategy" proposed a novel distributed Kalman filter in multi-sensor observations based on Metropolis–Hastings (M–H) sampling strategy. Firstly, combined with the latest observation information and the accuracy information of sensor which is also used to describe the prior modeling knowledge of observation system, they design the bootstrapped observation sampling for linear observation system. Secondly, aiming to the consistency deviation phenomenon appearing in the bootstrapped observations of single sensor, through constructing the likelihood degree of multi-sensor bootstrapped observations and the accept probability of credible observations, meanwhile, combined with the M–H sampling strategy, They give the validation method of credible observations. Finally, the realization steps of new algorithm are constructed according to the weighted fusion criterion. The advantage of new algorithm is to improve greatly the filtering precision with additional less hardware costs. The theoretical analysis and experimental results show the feasibility and efficiency of the proposed algorithm.

Acknowledgements The work was supported in part by the NSFC-61622205 and in part the NSFC-61472110.

Jun Yu received his BEng and PhD from Zhejiang University, Zhejiang, China. He is currently a Professor with the School of Computer Science, Hangzhou Dianzi University. From 2009 to 2011, he worked in Singapore Nanyang Technological University. From 2012 to 2013, he was a visiting researcher in Microsoft Research Asia (MSRA). Over the past years, his research interests include multimedia analysis, machine learning and image processing. He has authored and co-authored more than 80 scientific articles including IEEE Transactions on Image Processing (TIP), IEEE Transactions on Multimedia (TMM), IEEE Transactions on Systems, Man and Cybernetics, Part B (TSMCB), Pattern Recognition (PR). He has won IEEE TMM best paper award in 2015 and IEEE Signal Processing Society best paper award in 2017. He has (co-)chaired for several special sessions, invited sessions and workshops. He served as a program committee member or reviewer top conferences and prestigious journals, such as CVPR, SMC, signal processing, IEEE TKDE, TSMCB, TCSVT, information sciences and neurocomputing. He is a Professional Member of the ACM, IEEE and the CCF.

Xue Mei is Senior Scientist at Future Mobility Research Department, Toyota Research Institute, North America. From 2008 to 2012, he was at the Automation Path-finding group in Assembly and Test Technology Development and Visual Computing Group at Intel Corporation. Dr. Xue Mei received the BS degree in Electrical Engineering from University of Science and Technology of China in 2002 and the PhD degree from the University of Maryland, College Park, in Electrical Engineering in 2009. His research interests include robotics, image processing, computer vision and machine learning. He served as a program committee member or reviewer top conferences and prestigious journals, such as CVPR, ICCV, IEEE TPAMI, IEEE TIP, IEEE TKDE, TSMCB, TCSVT.

Prof. Fatih Porikli is an IEEE Fellow and a Professor in the Research School of Engineering, Australian National University (ANU), Canberra. He is also acting as the Computer Vision Group Leader at NICTA, Australia. He has received his PhD from New York University (NYU), New York, in 2002. Previously he served a Distinguished Research Scientist at Mitsubishi Electric Research Laboratories (MERL), Cambridge. Before joining MERL in 2000, he has contributed broadly to object and motion detection, tracking, image-based representations and video analytics. His current research interests include computer vision, pattern recognition, manifold learning, robust and sparse optimization, multimedia processing, data analysis, online learning and classification, grid and parallel computing and data mining with many commercial applications including video surveillance, medical systems, car navigation, intelligent transportation, logistics, satellite systems, automation, visualization and consumer electronics. Prof Porikli authored more than 150 publications and invented 66 patents. He has h-index 40 and i10-index 88, citation count 7700+.

Dr. Jason Corso is currently an Associate Professor of Electrical Engineering and Computer Science at the University of Michigan. He received his PhD in Computer Science at The Johns Hopkins University in 2005. He is a recipient of the NSF CAREER Award (2009), ARO Young Investigator Award (2010), Google Faculty Research Award (2015) and on the DARPA CSSG. His main research thrust is high-level computer vision and its relationship to human language, robotics and data science. He primarily focuses on problems in video understanding such as video segmentation, activity recognition and video to text. He primarily studies the coupled problems of segmentation and recognition from a Bayesian perspective emphasizing the role of statistical models in efficient visual inference. His long-term goal is a comprehensive and robust methodology of automatically mining, quantifying and generalizing information in large sets of projective and volumetric images and video.