



Editorial

Big data meets multimedia analytics



1. Introduction

With the rapid development of computing and sensing technologies, such as the emergence of social networking websites and wearable devices, many new research opportunities and challenges for multimedia content analysis have arisen.

Many big data modeling methods, computing algorithms, and signal processing technologies have recently been successfully developed and applied to multimedia content analysis: for example, multi-view learning algorithms have been proposed for exploring the variety of multimedia content; sparse and manifold learning have been developed for high dimensional multimedia data representation; deep learning has produced promising results in large scale multimedia retrieval; and compressive sensing and new sampling schemes have been investigated for big data analytics.

Motivated by the inclination to collect a set of recent advances and results in these related topics, provide a platform for researchers to exchange their innovative ideas on big modeling and computing solutions for multimedia content analytics, and introduce interesting utilizations of modeling and computing algorithms for particular social/personal media applications, this special issue will target emergent big modeling and computing methods for multimedia signal processing and understanding.

To summarize, this Signal Processing (Elsevier) special issue offers a venue to present innovations for using artificial intelligence and computing techniques for multimedia analytics, including: (1) theoretical advances as well as algorithm developments in big data technology for specific social/personal media analytics problems; (2) reports of practical applications and system innovations in social/personal media analytics; and (3) novel datasets as test beds for new developments, preferably with implemented standard benchmarks.

This special issue attracted 67 paper submissions from all around the world. Eventually, 26 high-quality articles were accepted after two rounds of vigorous reviews by at least two expert reviewers for each paper. We strongly

believe that it is now an ideal time to publish this special issue with the aforesaid selected papers. The contents will provide readers of Signal Processing with cutting-edge and topical information for their related research.

2. Overview of articles

In this section, we give an overview of all the included papers summarizing the contributions and novel aspects of each paper. We roughly divide the 26 papers into three subsections according to their areas, namely, retrieval and classification, detection and prediction and other related topics.

2.1. Retrieval and classification

The paper entitled “Investigating the Impact of Frame Rate Towards Robust Human Action Recognition” investigates the impact of frame rate on human action recognition with several state-of-the-art approaches and three benchmark datasets. In addition, It investigates key-frame selection techniques for choosing a set of suitable frames from an action sequence for action recognition. Promising results indicate that well designed key-frame selection methods can produce a set of representative frames and eventually reduce the impact of frame rate on the performance of human action recognition.

The paper entitled “Multi-Task Human Action Recognition via Exploring Super-Category” proposes a novel human action recognition method in multi-task learning framework with super-category. By applying the explored super-category information as a prior, feature sharing within super-category and feature competition between super-category are simultaneously encouraged in multi-task learning framework. Experimental results on large and realistic datasets HMDB51 and UCF50 show that the proposed method achieves higher accuracy with fewer dimensions of features over several state-of-the-art approaches.

The paper entitled “On Very Large Scale Test Collection for Landmark Image Search Benchmarking” presents a large-scale test collection to support robust performance evaluation of landmark image search and corresponding construction methodology. It develops a very large scale test collection consisting of three key components: (1) 355,141 images of 128 landmarks in five cities across three continents crawled from Flickr; (2) different kinds of textual features for each image, including surrounding text, contextual data, and metadata; and (3) six types of low-level visual features. In order to support robust and effective performance assessment, a series of baseline experimental studies have been conducted on the search performance over both textual and visual queries. The results demonstrate importance and effectiveness of the test collection.

The paper entitled “Cross-Modal Self-Taught Hashing for Large-Scale Image Retrieval” proposes Cross-Modal Self-Taught Hashing (CMSTH) for large-scale cross-modal and unimodal image retrieval. Its learning process contains three steps: First it proposes Hierarchical Multi-Modal Topic Learning (HMMTL) to detect multi-modal topics with semantic information. Then it uses Robust Matrix Factorization (RM-F) to transfer the multi-modal topics to hash codes that are more suited to quantization, and these codes form a unified hash space. Finally it learns hash functions to project all modalities into the unified hash space. Experimental results on two web image datasets demonstrate the effectiveness of CMSTH compared to representative cross-modal and unimodal hashing methods.

The paper entitled “Chart Classification By Combining Deep Convolutional Networks and Deep Belief Networks” proposes a novel framework to classify charts by combining convolutional networks and deep belief networks. It firstly extracts deep hidden features of charts, which are taken from the fully connected layer of deep convolutional networks. It then utilizes deep belief networks to predict the labels of the charts based on their deep hidden features. Compared with previous methods using primitive feature extraction, the deep features give the framework better scalability and stability. It collects a 5-class chart data set with more than 5000 images and show that the proposed framework outperforms existing methods greatly.

The paper entitled “Modeling object recognition in visual K-means and Non-negative Sparse Coding” extends the model in some biologically plausible ways and construct a five-layer computational model, denoted as Sparse-HMAX model. It first applies Gabor filters to describe the response properties of V1 neurons as in original HMAX model and characterize C1 image patches with HOG descriptors. Then it integrates multiple firing k-means into the HMAX model to emulate the V2 neural responses and non-negative sparse coding to model V4 neurons. Experimental results on three public image databases show that the proposed method exhibits great improvements over the original HMAX model both in accuracy and speed for object recognition.

The paper entitled “CDMMA: Coupled Discriminant Multi-Manifold Analysis for Matching Low-Resolution Face Images” develops a novel method called coupled discriminant multi-manifold analysis (CDMMA) to address

the problem of matching a (low-resolution) LR or poor quality face image to a gallery of high-resolution (HR) face images. It first explores the neighborhood information and local geometric structure of the multi-manifold space spanned by the samples. And then, it explicitly learns two mappings to project LR and HR faces to a unified discriminative feature space (UDFS) through a supervised manner followed with a conventional classification. Experimental results conducted on two standard face recognition databases demonstrate the superiority of the proposed CDMMA.

The paper entitled “A Novel Label Learning Algorithm for Face Recognition” proposes a novel FME algorithm for face recognition. It has two main contributions. First, it devices a score fusion scheme to predict the label of the original unknown sample. Second, it obtains mirror images of all original face images and views both mirror images and original face images as available samples. The experimental results demonstrate that algorithm proposed in this paper can perform very well in face recognition.

The paper entitled “Local Coordinate Based Graph-Regularized NMF for Image Representation” proposes a local coordinate based graph regularized NMF method (LCGNMF) to enforce the learned coefficients to be sparse by incorporating the local coordinate constraint meanwhile preserving the geometric structure of the data with graph regularization. It further develops a multiplicative update rule to solve LCGNMF and theoretically proved its convergence. Experiments of image clustering on several popular image datasets verify the effectiveness of LCGNMF compared to the representative methods in quantities.

2.2. Detection and prediction

The paper entitled “A Novel Disease Severity Prediction Scheme via Big Pair-wise Ranking and Learning Techniques using Image-based Personal Clinical Data” proposes a novel dementia disease severity prediction scheme using new big ranking and learning techniques. First, a single-pixel based method is presented to correct the partial volume effect in arterial spin labeling images. Second, novel big pair-wise ranking and learning techniques is proposed to realize the dementia disease severity prediction task using arterial spin labeling images after partial volume effects correction. Extensive experiments using a big database composed of images acquired from 350 real demented patients are carried out with several conventional methods being compared. Comprehensive statistical analysis is performed and it suggests that the new scheme is promising in dementia disease severity prediction.

The paper entitled “A Unified Model Sharing Framework for Moving Object Detection” proposes a unified model-sharing framework for moving object detection. Firstly, it establishes a many-to-one correspondence by model sharing between pixels, and a pixel is labeled into foreground or background by searching an optimal matched model in the neighborhood. Then a random sampling strategy is introduced for online update of the shared models. In this way, it can reduce the total number of models dramatically and match a proper model for each pixel accurately. Two popular approaches, statistical model

and sample consensus model, are used to verify the effectiveness. Experiments on ChangeDetection benchmark 2014 demonstrate the superiority of the model sharing solution.

The paper entitled “Robust Text Detection via Multi-degree of Sharpening and Blurring” proposes an approach to scene text detection via multi-degree of sharpening and blurring. Input image is sharpened and blurred with unsharp masking (USM) and bilateral filter. Then components are extracted with Maximally Stable Extremal Regions (MSER) from origin and processed images. Color, spatial layout and distance features of component are calculated, and features are weighted to construct the text candidates with distance function. At last, text candidates are estimated with a character classifier and the non-text candidates are eliminated. Experiments show that the proposed approach is robust to complex backgrounds and low image quality.

The paper entitled “A Spectral-Spatial based Local Summation Anomaly Detection Method for Hyperspectral Images” proposes a novel local summation anomaly detection method (LSAD) that combines the multiple local distributions from neighboring local windows surrounding the pixel under test (PUT) with spectral-spatial feature integration. The experimental results demonstrate that the proposed anomaly detection strategy outperforms the other traditional anomaly detection methods.

The paper entitled “A Picture Tells a Thousand Words-About You! User Interest Profiling from User Generated Visual Content” posits that images posted by users on online social networks are a reflection of topics they are interested in and proposes an approach to infer user attributes from images posted by them. It analyzes the content of individual images and then aggregates the image-level knowledge to infer user-level interest distribution. It employs image-level similarity to propagate the label information between images, as well as utilize the image category information derived from the user created organization structure to further propagate the category-level knowledge for all images. A large-scale social network dataset of 1.5+ million images created from Pinterest is used for evaluation and the experimental results demonstrate the effectiveness of the proposed approach.

The paper entitled “A Scene Change Detection Framework for Multi-Temporal Very High Resolution Remote Sensing Images” explores a scene change detection framework for VHR images, with a bag-of-visual-words (BOVW) model and classification-based methods. Image scenes are represented by a word frequency with three kinds of multi-temporal learned dictionary, i.e., the separate dictionary, the stacked dictionary, and the union dictionary. Post-classification and compound classification were evaluated for their performances in the “from-to” change results. Two multi-temporal scene datasets were used to quantitatively evaluate the proposed scene change detection approach. The results indicate that the proposed scene change detection framework can obtain a satisfactory accuracy and can effectively analyze land-use changes, from a semantic point of view.

The paper entitled “A Robust Vision Inspection System for Detecting Surface Defects of Film Capacitors” presents a

robust vision inspection system for detecting the surface defects of film capacitors. In particular, it uses a novel Non-subsampled Contourlet Transform (NSCT) based algorithm to detect the surface defects. Then, the detection results are sent to the mechanical separation system via a serial port. The defective capacitors are peeled off the production line by motor. The proposed system can improve the detection efficiency. It thus can improve the product quality and reduce production costs. Experimental results have demonstrated that the system achieves superior performance over other state-of-the-art solutions. Moreover, with the system, large-scale vision data of capacitor surfaces can be collected and used to supervise capacitor manufacturing process.

2.3. Other related topics

The paper entitled “Hypergraph Regularized Autoencoder for Image-Based 3D Human Pose Recovery” proposes a novel feature extractor with deep learning. It is based on denoising autoencoder and improves traditional methods by adopting locality preserved restriction with hypergraph Laplacian. Experimental results on two datasets show that the recovery error has been reduced by 10–20%, which demonstrates the effectiveness of the proposed method.

The paper entitled “Real-time Tracking-by-Learning with High-order Regularization Fusion for Big Video Abstraction” proposes a novel tracking strategy based on semi-supervised learning with high order regularization fusion. It employs two different types of regularizers to achieve more accurate label assignment based on kernelized confidence prediction and graph-based bi-directional trace from motion. Via dynamic budget maintenance for model updating, the proposed tracking method demonstrated to outperform most state-of-art trackers on challenging benchmark videos with a fixed parameter configuration.

The paper entitled “Supervised Sampling for Networked Data” models a network as a Markov chain and derives supervised random walks to learn a stationary distribution of the sampled network. The learned stationary distribution can help identify the best node to be sampled next. With more information acquired about sampled nodes, supervised sampling can be strengthened in turn. Experiments on synthetic as well as real-world networks show that the supervised sampling algorithm outperforms other existing methods in obtaining more target nodes in the sampled networks.

The paper entitled “Real-Time People Counting for Indoor Scenes” proposes a counting approach for indoor scenes, which can count not only moving crowds but also stationary crowds efficiently. Firstly, a foreground extraction assisted by detection is introduced for crowd segmentation and noise removal with a feedback update scheme. Then it builds a multi-view head-shoulder model for people matching in the foreground and estimate the number of people with an improved K-mean clustering approach. Finally, it presents a temporal filter with frame-difference to further refine the counting results. A new indoor counting dataset including about 570,000 frames was collected from four different scenarios. Experiments and comparisons show the superiority of the proposed approach.

The paper entitled “On Characterizing Scale Effect of Chinese Mutual Funds via Text Mining” investigates the correlation between mutual funds’ scale and return in China by text mining on the sheer volume of online financial reports. It starts with a data-fitting test to demonstrate that the tail of fund scale fits best in a distribution between Power-Law and Log-Normal. It then introduces K-means clustering for fund categorization, which enables reliable examination of correlations between fund scale and return. Empirical study unveils some interesting findings on the scale effect of funds under different market conditions. These findings highlight the uniqueness of emerging markets while providing interesting guidelines for exploiting big data analytics for financial studies.

The paper entitled “Biologically Inspired Image Quality Assessment” presents a novel IQA approach named biologically inspired feature similarity (BIFS), which is demonstrated to be highly consistent with the human perception. Biologically inspired features (BIFs) of the test image and the relevant reference image are first extracted. Afterwards, local similarities between the reference BIFs and the distorted ones are calculated and then combined to obtain a final quality index. Thorough experiments on a number of IQA databases demonstrate that the proposed method is highly effective and robust, and outperform state-of-the-art FR-IQA methods across various datasets.

The paper entitled “Dynamic Texture Modeling and Synthesis using Multi-Kernel Gaussian Process Dynamic Model” proposes a Bayesian-based nonlinear dynamic texture modeling method for dynamic texture synthesis. It firstly utilizes the Gaussian process latent variable model for dimensional reduction. Furthermore, it designs a multi-kernel dynamic system for the latent dynamic behavior modeling. The model automatically constructs a suitable nonlinear kernel for dynamic modeling and therefore is capable of fitting various types of dynamics. Experimental results on the DynTex database show that the proposed method can achieve synthesis results with higher visual quality.

The paper entitled “Superpixel-guided Nonlocal Means for Image Denoising and Super-resolution” proposes a novel superpixel-guided nonlocal means (SNLM) algorithm. The thorough quantitative and qualitative results demonstrate that SNLM is more effective for image denoising and super-resolution than the conventional NLM-based method.

The paper entitled “Global Image Completion with Joint Sparse Patch Selection and Optimal Seam Synthesis” presents a global exemplar-based image completion method for filling large missing or unwanted regions in an image. Image completion problem is formulated as a global discrete optimization problem with a well-defined energy function. The energy function can evaluate image consistency globally and is minimized with an expectation-maximization (EM) like algorithm. It analyzes the proposed global energy function and optimization method in theory. Simulation comparisons with other state-of-the-art methods show the superiority of the proposed method in ensuring global coherent and avoiding image blurring.

The paper entitled “A Service Components Pipeline Model Based on Multi-Source Data Extraction” proposes a service components and its pipeline service mechanism on

the basis of multi-source data extraction. It builds models of open multi-source data extraction and service data package as well as realizes the transparent access to multi-source data service. Finally, it takes ethnic minorities featured villages GIS service workflow to verify the effectiveness of the proposed experiments and methods.

Acknowledgment

We would like to thank authors for their continuous efforts and enduring patience as this issue came together. We would also like to thank the Signal Processing Editor-in-Chief Professor Björn Ottersten, for his constant and significant support. Moreover, we also acknowledge the support from the Elsevier colleagues, and finally, we would like to express our deepest appreciation to all anonymous referees who responsibly and effectively reviewed the special issue submissions. Without their timely and important assistance, it is impossible to handle such a large number of submissions and finalize this high-quality special issue as scheduled.

Tat-Seng Chua,
*School of Computing, National University of Singapore,
Singapore*

Xiangjian He,
*Global Centre for Big Data Technologies (GCBDT), University
of Technology, Sydney*

Weifeng Liu¹,
*College of Information and Control Engineering, China Uni-
versity of Petroleum, Qingdao 266580, China*

Massimo Piccardi,
*Global Centre for Big Data Technologies (GCBDT), University
of Technology, Sydney*

Yonggang Wen,
*School of Computer Engineering, Nanyang Technological
University, Singapore*

Dacheng Tao
*Centre for Quantum Computation and Intelligent Systems,
University of Technology, Sydney*

Available online 12 December 2015

¹ Corresponding author.