Guest Editorial

Multimedia Modeling

The effective and efficient modeling of multimedia data has attracted extensive research interests over last decades. Nowadays, multimedia modeling forms the basis of a wide variety of applications and services, such as search, recommendation, advertising, and personalization. On the other hand, we have witnessed the explosive growing of multimedia data. Such scale brings significant challenges and profound impacts to multimedia modeling in different aspects. For example, it is very challenging for many classification, clustering, and labeling algorithms to effectively and efficiently handle large-scale multimedia data, especially when the scale comes up from tens of thousands to tens of millions or even billions. Feature extraction process also needs to speed up. Fortunately, along with the growth of multimedia data, more and more resources also become available, such as the associated metadata, context and social information. In addition, collaborative tagging, a representative behavior of web 2.0, enables the availability of tags for a large amount of multimedia signals on the Internet. These facts have provided opportunities to tackle the difficulties in multimedia modeling.

This special issue is organized with the purpose of introducing novel research work on multimedia modeling. Submissions have come from an open call for paper. With the assistance of professional referees, twenty papers are selected after at least two rounds of rigorous reviews. These papers cover widely subtopics of multimedia modeling. We divide them into five parts according to the themes of the papers.

The first part contains three papers that are related to multimedia representation. The first paper, "*Feature Selection with Spatial Path Coding for Multimedia Analysis*", introduces a supervised structured sparse feature selection approach. In the approach, different dimensions of features are modeled with a graph and the spatial correlations among features are considered as directed edges. Thus, supervised feature selection is formulated as a path selection problem with minimized cost. In the second paper, "*A Data-Driven Study on Image Feature Extraction and Fusion*", Wang et al. conduct a study on several representative image feature extraction algorithms. They also design a feature fusion strategy based on their observations to boost image classification accuracy. In the third paper, "*GA-SIFT: A New Scale Invariant Feature Transform for Multi-Spectral Image Using Geometry Algebra*", Li et al. introduce a scale invariant feature transform for multi-spectral images. As the conventional SIFT features cannot be directly applied to multi-spectral images, this work uses the theory of geometry algebra to detect and describe local feature points.

The second and also the largest part contains nine papers about multimedia content analysis and prediction. The first paper, "*Low Level and High Level Prior Learning for Visual Saliency Estimation*", introduces a hybrid feature driven visual saliency analysis

method that takes into account both low-level and high-level features. In the second paper, "*Salient Region Detection for Complex Background Images using Integrated Features*", Zhang et al. summarize five principles for designing a salient region detector, and they then propose a novel method that detects highlighted salient regions by utilizing saliency values and spatial weights. In the third paper "*A Multi-dimensional Image Preference Prediction Model for User Generated Images in Social Networks*", Yang et al. build two models for presentation measurement and distortion measurement respectively by exploring image, text, and social link information. The outcomes of these two models are pooled to obtain a final preference score. The fourth paper, "*Directional Projection Based Image Fusion Quality Metric*", introduces a quality evaluation metric for image fusion. Input images are projected onto the fused image's signal characterization to obtain project vector, and the difference between the source images and the fused image are measured by the projection vectors. In the fifth paper, "*A Collaborative Approach for Face Verification and Attributes Refinement*", Zhang et al. explore the relationship of face verification and human attribute learning and propose a collaborative approach allowing them interact with each other to iteratively refine the attribute and face verification results. In the sixth paper, "*Action Recognition Based on Overcomplete Independent Components Analysis*", the authors introduce an action recognition framework based on overcomplete ICA. In contrast to many conventional action recognition methods that rely on a SVM classifier, the proposed method exploits the response properties of overcomplete ICA to perform classification. The seventh paper, "*Automatic Image Annotation by Semi-Supervised Manifold Kernel Density Estimation*", introduces a novel semi-supervised learning method for image classification. It extends a conventional manifold kernel density estimation algorithm to a semi-supervised learning approach by exploring unlabeled data. In the eighth paper, "*Learning the Object Location, View and Scale for Image Categorization with Adapted Classifier*", Yan et al. propose a method to model object location, view and scale and the sharing information between different categories for image categorization task. The ninth paper, "*Semantic Preserving Distance Metric Learning and Applications*", proposes a novel distance metric learning method that is able to encode the visual feature similarity and semantic similarity in a unified feature space.

In the third part of the special issue, we have three papers focusing on multimedia indexing and search. The paper "*A GPU-Accelerated Non-negative Sparse Latent Semantic Analysis Algorithm for Social Tagging Data*" introduces a GPU-accelerated fast optimization algorithm for non-negative sparse latent semantic analysis In order to utilize social tagging data in tag recommendation and image classification tasks. The second paper, "*3D Model Retrieval and Classification by Semi-Supervised Learning with Content-based Similarity*", proposes a method to estimate the distance between two 3D models. This distance measure is then employed for 3D object retrieval and classification. In the third paper, "*Nonnegative Sparse Locality Preserving Hashing*", Liu et al. introduce a novel nonnegative sparse locality preserving hashing method. In the method, nonnegative and sparse constraints are

imposed for a more accurate solution, and semantic information can also be well preserved.

The fourth part contains five papers that are related to modeling-based multimedia processing. The first paper "*Re-Texturing by Intrinsic Video*" introduces a novel video re-texturing approach that can produce high quality video re-texturing results with a variety of sample textures. In the second paper, "*Single-image Motion Deblurring Using an Adaptive Image Prior*", He et al. introduce a single image deblurring algorithm that adapts the image prior to the underlying detailed structures. In the paper "*Gradient-domain-based Enhancement of Multi-view Depth Video*", Liu et al. propose a gradient-domain based enhancement method for multi-view depth videos. Different from conventional enhancement methods that only considered a single dimension or two dimensions, the proposed method optimizes the depth images in the spatial, temporal and inter-view dimensions with a joint gradient field constraints. The fourth paper, "*An Approach to Generate and Embed Sign Language Video Tracks into Multimedia Contents*", proposes a solution for automatic generation and insertion of sign language video tracks into captioned digital multimedia content in order to help deaf audience. The fifth paper, "*Video Abstraction Based on fMRI-Driven Visual Attention Model*", is about video abstraction. The authors introduce a novel video abstraction paradigm which leverages functional magnetic resonance imaging to monitor and quantify the brain's responses to video stimuli. The responses are used to guide the extraction of visually informative segments from videos.

To conclude, the papers in this special issue cover different techniques for multimedia modeling. We believe this special issue will benefit researchers and practitioners working in this area.

Acknowledgement
We would like to thank Prof. Witold Pedrycz for providing us the chance to organize this special issue. We thank the reviewers for their great efforts. Their professional evaluations and constructive comments are vital for securing the high quality of the special issue. Finally, we express our thanks to all the authors who have contributed to the special issue.

Meng Wang, Hefei University of Technology
Dacheng Tao, University of Technology, Sydney
Benoit Huet, EURECOM, France