Guest Editorial Multimodal Modeling and Analysis Informed by Brain Imaging—Part I

Abstract—**Human brains are the ultimate recipients and assessors of multimedia contents and semantics. Recent developments of neuroimaging techniques have enabled us to probe human brain activities during free viewing of multimedia contents. This special issue mainly focuses on the synergistic combinations of cognitive neuroscience, brain imaging, and multimedia analysis. It aims to capture the latest advances in the research community working on brain imaging informed multimedia analysis, as well as computational model of the brain processes driven by multimedia contents.**

Index Terms—**Affective computing, brain computer interface, brain imaging, brain-informed multimedia analysis.**

I. The Scope of this Special Issue

A challenging problem facing multimedia content analysis is the semantic gap between the high-level perception and cognition in the human brain and the low-level features embedded in digital contents. The human brain is the ultimate recipient and assessor of multimedia contents and semantics. Deep understanding of the brain responses to multimedia will fundamentally advance the computational strategies for multimodal representation, classification and retrieval.

We envision a future with a seamless integration between cognitive neuroscience, a discipline related to the principle and mechanisms of the brain, and computer science, a discipline designing automated digital algorithms. Examples of such integration include neural network algorithms, which could reduce the semantic gap by mimicking the neural processes in the brain. Conversely, applications of automated computer algorithms have advanced our understanding of the brain. In the recent years, we have witnessed the emergence of novel brain-guided or brain-informed techniques in multimedia analysis and modeling, including computational visual attention models, sparse representation techniques, and deep learning techniques. These techniques have been applied to object recognition, image categorization, image/video compression, image/video retrieval, and video summarization.

The remarkable development of neuroimaging techniques such as functional magnetic resonance imaging (fMRI), electroencephalography (EEG), and magnetoencephalography (MEG), has enabled us to probe the human brain in natural settings such as free viewing of multimedia contents. This development is leading a new trend that applies neuroscience and neuroimaging to assisting multimodal analysis and modeling. This new methodology has considerably narrowed the gaps between the low-level multimedia features and the high-level semantics. In parallel, neuroimaging combined with naturalistic stimuli such as films and music also provide neuroscientists an intriguing opportunity to examine the brain circuitry underlying natural experience.

II. Contribution To The Special Issue

This Special Issue incorporates 17 papers. These papers basically can be classified into three themes. The first theme is to probe human brain activities during viewing of multimedia contents using fMRI or EEG and then apply neurophysiological signals or features to facilitate multimedia analysis such as affective computing, video categorization, image retrieval, and face recognition. The second theme is to explore how to use fMRI or EEG or the combination of fMRI and EEG to examine the functional and structure of the brain and apply the neuroimaging techniques to clinical applications. The third theme is about brain–computer interface (BCI), which interprets commands in the brain based on invasive or noninvasive neural recordings, and uses these command to control computers and machines. This issue contains Part I with the first nine papers. Below, we briefly summarize the highlights of each paper.

Affective computing is a rapidly growing field that bridges affective neuroscience and computer science. Much focus is on how to infer emotional states based on neurophysiological signals. EEG is a particularly practical tool for affective computing offering great potential for a variety of real life applications. The question is how to extract the emotion-relevant information from the multi-channel, multi-frequency EEG signals. In their paper, "*Investigating Critical Frequency Bands and Channels for EEG-based Emotion Recognition with Deep Neural Networks*," Zheng and Lu [1] employ deep belief networks (DBN) to identify the critical frequency and channels for valence recognition. They find the profiles trained by DBN, based on a subset of channels and frequency bands, could reliably and accurately classify emotional valence. Their findings further suggest that neural signatures associated with different emotions do exist and they share commonality across sessions and individuals.

How music arouses the listener's specific emotion response in the listener is not well-understood yet [2]. Inspired by the fact that brain imaging techniques such as EEG and fMRI provide an effective tool to probe the human brain activities during music listening, Liu *et al.*, [3] in their paper entitled "*What Strikes the Strings of Your Heart?—Multi-Label Dimensionality Reduction for Music Emotion Analysis via Brain Imaging*," propose a novel computational framework via EEG to explore the relationship between the music and evoked emotions. The user provided emotion labels are refined by using the EEG consistency and then a multi-label dimensionality reduction technology is applied to discover the genuine correlation between music and emotion.

While much progress has been made in automated speech and face recognition, accurate emotion recognition remains a major

Digital Object Identifier 10.1109/TAMD.2015.2495698

challenge. While earlier works mostly rely on unimodal information such as neurophysiological signals and audio-visual features separately, recent advances are made toward multimodal integration for emotion tagging. In the article "*Emotion Recognition with the Help of Privileged Information*," Wang *et al.* [4] propose a novel integration method that utilizes information that is only available during training but not testing. Using three multimodal databases that contain EEG response to emotional video clips, they examine emotion recognition when training was done with EEG responses alone, with the audio-visual features of video clips alone, or with both together. They find when both information is available during training, hence the aid of privileged information, the accuracy of emotion classification is greatly improved. Their findings have practical significance for future work on emotion tagging.

A major challenge in multimedia analysis and application is the retrieval of semantics from low-level audio-visual features. Modern computational algorithms can exact this low-level information with ease; they, however, have great difficulties in forming semantic concepts based on these low-level features. How to improve the performance of semantic categorization, hence, is a major research goal of multimedia content analysis. Recent progress has been made by learning from the neural mechanisms of semantic processing excelled by the human brain. In their article, "*Decoding Semantics Categorization during Natural Viewing of Video Streams*," Hu *et al.* [5] tackle this problem of semantic categorization by integrating neuroimaging data during complex, naturalistic video streams. They propose a novel framework that models semantic processing based on functional interactions between large-scale brain networks. Their results demonstrate this framework can accurately retrieve semantics during naturalistic video viewing and further reveal a significant contribution from the working memory network to semantic categorization.

As shown in [6], human brain is capable of recognizing faces across various viewpoints, illumination conditions, even in the case of partial occlusion. In the paper entitled "*An Iterative Approach for EEG based Rapid Face Search: A Refined Retrieval by Brain Computer Interfaces*," Wang *et al.* [7] propose a face image retrieval system by integrating face recognition algorithm in computer vision and powerful cognitive function of human brain measured by EEG signals, which achieves promising performance.

Whether there is an aging effect in the early ERP to facial emotion or not is debated amid controversial results [8], [9]. In the paper entitled "*Age Effect in Human Brain Responses to Emotion Arousing Images: The EEG 3-D-Vector Field Tomography Modeling Approach*," Papadaniil *et al.* [10] develop a novel method called 3-D-vector field tomography to investigate the age-related differences in the N170 component of EEG responses to emotionally arousing pictures. The results reported in the paper show statistically significant age-related difference in the amplitude of the N170 component for two emotions, i.e., anger and fear.

In the paper entitled "*Perceptual Experience Analysis for Tone-Mapped HDR Videos based on EEG and Peripheral Physiological Signals*," Moon and Lee [11] examine the perceptual experience of tone-mapped high dynamic range (HDR) videos both explicitly by conducting a subjective questionnaire assessment and implicitly by using EEG and peripheral physiological signals. Their results show that tone-mapped HDR videos are more interesting and more natural, and they offer better quality than low dynamic range (LDR) videos. They also report significant difference in the physiological signals between tone-mapped HDR and LDR videos in the classification under both subject-dependent and subject-independent scenarios. This work potentially contributes to the field of high dynamic range imaging.

FMRI has been shown to be a powerful tool for understanding human brain functional activities. Wang *et al.* [12] in their paper entitled "*Predicting Purchase Decisions based on Spatio–Temporal Functional MRI Features Using Machine Learning*," present a computational model to classify purchase decisions based on the combination of spatio-temporal fMRI features extracted from a few significantly activated brain regions during decision-making. This model can achieve 71% prediction accuracy.

In the paper entitled "*A Robust Gradient based Algorithm to Correct Bias Fields of Brain MR Images*," Ling *et al.* [13] develop a new method to estimate bias fields from brain MRI images by using a gradient based method and a low-order polynomial fitting method. Experiments using both simulated and real MRI images demonstrate the improved accuracy and robustness to noises. This work has potential applications in human brain mapping.

This issue includes the first part of the Special Issue with nine papers. The next issue (December 2015) will include the second part of the Special Issue with the publication of eight additional papers.

Junwei Han*, Guest Editor*
Northwestern Polytechnical University
School of Automation
Xi'an, 710072 China
E-mail: junweihan2010@gmail.com

Tianming Liu*, Guest Editor*
The University of Georgia
Department of Computer Science
Cortical Architecture Imaging and Discovery Lab
Athens, GA 30602 USA
E-mail: tliu@cs.uga.edu

Christine C. Guo*, Guest Editor*
QIMR Berghofer
Medical Research Institute
Herston, QLD 4006 Australia
E-mail: christine.cong@gmail.com

Juyang (John) Weng*, Guest Editor*
Michigan State University
Department of Computer Science and Engineering
Cognitive Science Program and Neuroscience Program
East Lansing, MI 48824 USA
E-mail: weng@cse.msu.edu

References

[1] W. Zheng and B. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," *IEEE Trans. Autonom. Mental Develop.*, vol. 7, no. 3, pp. 162–175, Sep. 2015.

[2] Y.-H. Yang and H. H. Chen, "Machine recognition of music emotion: A review," *ACM Trans. Intell. Syst. Technol.*, vol. 3, no. 3, pp. 40:1–40:30, May 2012.

[3] Y. Liu, C. Wang, X. Wang, P. Zhou, G. Yu, and K. Chan, "What strikes the strings of your heart?—multi-label dimensionality reduction for music emotion analysis via brain imaging," *IEEE Trans. Autonom. Mental Develop.*, vol. 7, no. 3, pp. 176–188, Sep. 2015.

[4] S. Wang, Y. Zhu, L. Yue, and Q. Ji, "Emotion recognition with the help of privileged information," *IEEE Trans. Autonom. Mental Develop.*, vol. 7, no. 3, pp. 189–200, Sep. 2015.

[5] X. Hu, L. Guo, J. Han, and T. Liu, "Decoding semantics categorization during natural viewing of video streams," *IEEE Trans. Autonom. Mental Develop.*, vol. 7, no. 3, pp. 201–210, Sep. 2015.

[6] A. Gosling and M. Eimer, "An event-related brain potential study of explicit face recognition," *Neuropsychologia,*, vol. 49, pp. 2736–2745, 2011.

[7] Y. Wang, L. Jiang, Y. Wang, B. Cai, Y. Wang, W. Chen, S. Zhang, and X. Zheng, "An iterative approach for EEG based rapid face search: a refined retrieval by brain computer interfaces," *IEEE Trans. Autonom. Mental Develop.*, vol. 7, no. 3, pp. 211–222, Sep. 2015.

[8] S. Daniel and S. Bentin, "Age-related changes in processing faces from detection to identification: Erp evidence," *Neurobiol. Aging*, vol. 33, no. 1, pp. e1–28, 2012.

[9] M. Bieniek, L. Frei, and G. Rousselet, "Early ERPs to faces: aging, luminance, and individual differences," *Frontiers Psychol.*, vol. 4, p. 268, 2013.

[10] C. Papadaniil, V. Kosmidou, A. Tsolaki, L. Hadjileontiadis, M. Tsolaki, and I. Kompatsiaris, "Age effect in human brain responses to emotion arousing images: The EEG 3D-vector field tomography modeling approach," *IEEE Trans. Autonom. Mental Develop.*, vol. 7, no. 3, pp. 223–235, Sep. 2015.

[11] S. Moon and J. Lee, "Perceptual experience analysis for tone-mapped HDR videos based on EEG and peripheral physiological signals," *IEEE Trans. on Autonomous Mental Development*, vol. 7, no. 3, pp. 236–247, Sep. 2015.

[12] Y. Wang, V. Chattaraman, H. Kim, and G. Deshpande, "Predicting purchase decisions based on spatiotemporal functional MRI features using machine learning," *IEEE Trans. Autonom. Mental Develop.*, vol. 7, no. 3, pp. 248–255, Sep. 2015.

[13] Q. Ling, Z. Li, Q. Huang, and X. Li, "A robust gradient based algorithm to correct bias fields of brain MR images," *IEEE Trans. Autonom. Mental Develop.*, vol. 7, no. 3, pp. 256–264, Sep. 2015.

Junwei Han received the Ph.D. degree in pattern recognition and intelligent systems from the School of Automation, Northwestern Polytechnical University, Xi'an, China, in 2003.

He is currently a Professor with Northwestern Polytechnical University. His research interests include multimedia processing and brain imaging analysis.

Dr. Han is an Associate Editor of the IEEE Transactions on Human-Machine Systems, Neurocomputing, and Multidimensional Systems and Signal Processing.

Tianming Liu is a Professor of Computer Science at The University of Georgia. His research interest focuses on brain imaging and mapping. He has published over 160 peer-reviewed papers in this area.

Dr. Liu is the recipient of the NIH Career Award and the NSF CAREER Award, both in the area of brain mapping.

Christine C. Guo received the B.Sc. degree in biological sciences from Peking University, Beijing, China, and the Ph.D. degree in neuroscience from the School of Medicine, Stanford University, Stanford, CA, USA, followed by postdoctoral training at the Memory and Aging Center (UCSF).

She has broad research experience, from molecular biology and genetics, to electrophysiology and systems neuroscience. Her work focuses on understanding selective vulnerability at the network level in health and in neurodegenerative diseases, using modern neuroimaging techniques. She is also developing neuroimaging methods to understand the body-brain interaction and its breakdown in neurological and psychiatric disorders.

Juyang (John) Weng (S'85–M'88–SM'05–F'09) received the B.S. degree from Fudan University, in 1982, and the M.S. and Ph.D. degrees from the University of Illinois at Urbana-Champaign, Urbana, IL, USA, in 1985 and 1989, respectively, all in computer science.

He is a Professor at the Department of Computer Science and Engineering, the Cognitive Science Program, and the Neuroscience Program of Michigan State University, East Lansing, MI, USA. From August 2006 to May 2007, he was also a visiting professor at the Department of Brain and Cognitive Science of MIT. His research interests include computational biology, computational neuroscience, computational developmental psychology, bioinspired systems, vision, audition, touch, behaviors, and robotics. He is the author of over 300 papers, including the book "Natural and Artificial Intelligence: Introduction to Computational Brain-Mind."

Dr. Weng is an Editor-in-Chief of the *International Journal of Humanoid Robotics*, an Associate Editor of the IEEE TRANSACTIONS ON AUTONOMOUS MENTAL DEVELOPMENT, and the Editor-in-Chief of *Brain-Mind Magazine*. He was the Chairman of the Governing Board of the International Conferences on Development and Learning (ICDLs) (2005–2007, http://cogsci.ucsd.edu/~triesch/icdl/), the Chairman of the Autonomous Mental Development Technical Committee of the IEEE Computational Intelligence Society (2004–2005), an Associate Editor of IEEE TRANSACTIONS ON PATTERN RECOGNITION AND MACHINE INTELLIGENCE (2001–2004), and an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING (1994–1997).