Guest Editorial: Challenges and Perspectives for Affective Analysis in Multimedia

Mohammad Soleymani, *Member, IEEE*, Yi-Hsuan Yang, *Member, IEEE*, Go Irie, *Member, IEEE*, and Alan Hanjalic, *Senior Member, IEEE*

THERE has been an increasing interest in the research on multimedia indexing and retrieval based on subjective concepts, such as emotion, preference and aesthetics. The research efforts in this direction are considered human-centered and beyond the conventional keyword- or (semantic) object-based multimedia analysis paradigm [1]. In addition, the problems addressed by these efforts are considered challenging since they require multidisciplinary understanding of human behavior and perception as well as multimodal approaches integrating different modalities, e.g., audio, visual and text, to reach an acceptable level of performance. In addition to the multimedia content and to better take into account the human factors, users' spontaneous responses, either explicit (e.g., rating) or implicit (laughing, gaze direction, physiological responses) are increasingly also integrated as input in such approaches.

This special section focuses on affective analysis of multimedia, that is, the analysis focusing on estimating the expected emotional state of the user when interacting with multimedia content. The result of such analysis can enhance multimedia retrieval and recommendation systems by adding the emotion-related factor into the assessment of the appropriateness of the content for the user [2]. The mission of the special section is to provide a view at the state-of-the-art in this domain and point to the potentially interesting future research directions. It therefore features articles concerning multimedia affective understanding from different perspectives, including predicting likability and emotions from users' responses [3], [4], content analysis for music affective understanding [5], [6], [7], visual content analysis for aesthetics and emotional understanding [8], [9], [10], and an application for sentiment analysis [11].

In response to an open call for our special section, we initially received 35 submissions [4], [10] from which we accepted nine original articles. In the next section, we summarize the articles and their key findings. In the final section, we summarize the challenges for future research on affective analysis in multimedia.

Featured Work

Users spontaneous reactions can be used for understanding the content they watch or consume. In [3], Abadi et al. introduced a new database for emotion recognition in response to videos. The database includes near-infra-red (NIR) facial videos, horizontal Electrooculogram (hEOG), Electrocardiogram (ECG), trapezius-Electromyogram (tEMG) and most importantly Magnetoencephalogram (MEG) responses from participants who were watching short video clips. The self-assessed emotions and continuous annotation of the stimuli are provided with the goal of serving as ground truth. Correlation analysis and emotion recognition results were also given to serve as a baseline for the future database users. Emotional expressions in response to videos can be translated into how much the video was likable. McDuff et al. [4] first collected 12,000 sequences of facial expressions in response to 170 video advertisements. Then, using machine learning methods and face tracking techniques, they demonstrated the feasibility of detecting advertisement liking and potential purchase intent. This work, co-developed by an industrial partner, i.e., Affectiva Inc., presents a significant step in commercializing multimedia affective computing technologies.

Emotional content in music can be used for music indexing and recommendation. The special section contains three articles that characterize musical emotion as discrete classes. Ren et al. [5] proposed the use of a two-dimensional representation of frequency and modulation features, and demonstrated the significance of these novel features over three music mood datasets. They also described a leading solution for the music information retrieval evaluation exchange (MIREX) contest on audio mood classification. Liu et al. [6] considered music mood classification as a multi-label classification problem, and proposed a multi-label dimension reduction method to discover the intrinsic factors in music that convey emotions. Extensive numerical results showed that the dimensionality-reduced features lead to better classification accuracy, comparing to the original features. Wang et al. [7] explored a non-parametric Bayesian approach for emotional characterization of music. To construct a discriminative latent space while capturing correlations between emotions, they used a hierarchical Dirichlet process (HDP) mixture and modified it by introducing linear discriminant idea into the sampling distributions of HDP latent topics.

- *M. Soleymani is with the Centre Interfacultaire En Sciences Affectives (CISA), University of Geneva, Switzerland. E-mail: mohammad.soleymani@unige.ch.*
- *Y.-H. Yang is with the Academia Sinica, Taiwan. E-mail: yang@citi.sinica.edu.tw.*
- *G. Irie is with NTT the Corporation, Japan. E-mail: irie.go@lab.ntt.co.jp.*
- *A. Hanjalic is with the Department of Intelligent Systems, Delft University of Technology, the Netherlands. E-mail: A.Hanjalic@tudelft.nl.*

For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TAFFC.2015.2445233

Experiments demonstrated the effectiveness of the proposed approach for both music mood classification and emotion-based music retrieval.

Visual content expresses and evokes emotions in both utilitarian and aesthetics forms. Hence, image indexing will benefit from understanding their affective content. Aesthetics are closely related to natural responses of humans to multimedia content. To measure the consensus of aesthetic quality, the variance of the aesthetic quality scores is usually considered. Park and Zhang [8], however, argued that further statistical moments such as skewness and kurtosis could be more adequate measures, because such score distributions tend to be non-Gaussian. Based on the higher order statistical moments, they designed a dynamical model named drift-diffusion model for aesthetic perception (DDM4AP), and showed that this model can better explain the properties of the behaviors. Social networking services are now a common platform to share their emotional experiences with their friends by exchanging photos, short clips, and their comments to those. Wang et al. [9] focused on modeling of emotion influences over social networks to improve affective image content analysis performance. They proposed a new graphical model that takes into account different types of relationships such as user-content, user-user (friendship), and temporal dynamics. Using this model, they showed that such relationships indeed contribute to improving affective content analysis performance. Chen et al. [10] explored how visual content is used to convey affects by publishers of images in online social media, and how such content evokes affective responses accordingly on the viewer side. From the visual content and the text comments associated with the images, the authors proposed to learn a set of publisher affect concepts (PACs) and viewer affect concepts (VACs). By modeling the correlation between PACs and VACs, they developed and evaluated an innovative system that can automatically generate multi-sentence comments related to the affect of images, with high perceptual plausibility.

Sentiment analysis deals with the intrinsic affect in content and has a large body of applications, e.g., opinion mining and content characterization. Nguyen et al. [11] presented a comprehensive study of online blogs related to the Autism Spectrum Disorder, using three types of textual features: sentiment information, topics of interest, and language style. These feature were found to be effective in detecting autism related blogs in both personal and social settings.

Challenges and Perspectives

Articles in this issue are covering multimedia affective understanding using different modalities including textual, visual and auditory modalities. What they all have in common is the affective understanding of the content and its users. The featured work demonstrated how affective analysis in multimedia has advanced to the point of commercialization. However, we are still in need of a common understanding with regard to the psychology of affect to advance this topic. Ideally, we need new models developed by an inter-disciplinary effort, including researchers from computing, communication and psychology, that can be used in computational modeling of affect in multimedia.

A great challenge in affective computing is understanding the true emotions that people feel, i.e., the ground truth. In affective analysis in multimedia, affect might refer to different aspects of emotional characteristics [1], [12], [13], including intended, expected or felt emotions. Emotions that the content creator is intending to elicit in the audience are expressed or intended emotions; regardless of the fact that the users feel those emotions or not. Expected emotions arise in response to a content in most of its audience. Emotions that audience feel in response to a content are felt-emotions. Affect might be also referring to an intrinsic characteristic of the content, e.g., the subject of sentiment analysis. Each and every one of these aspects requires its own strategy for data collection. There is also some confusion over what we mean by affective terms in multimedia community, e.g., mood might refer to the intrinsic affect of a given content as opposed to its classic definition of a long and diffused affect with no certain stimulus [14]. A clearer definition for affective terminology will benefit the future work by dividing the affective analysis depending on its aim.

We ideally want to have access to a large number of emotional responses or labels, e.g., [4]. However, we cannot expect gathering this amount of data in a laboratory, e.g., [3]. The rise of crowdsourcing as an alternative for data collection to the laboratory settings is enabling efficient and effective data collection [1]. Also the proliferation of wearable devices will result in a massive generation of physiological and behavioral responses that can be used for this type of research. We should therefore position ourselves to benefit from these new opportunities and at the same time encourage developing and sharing databases such as the ones developed in MediaEval multimedia benchmarking campaign (http://www.multimediaeval.org) [15] or other publicly available databases such as [3].

Content-based affective understanding of multimedia has been mainly focused on directly translating low level features to emotions. A better understanding of what affective analysis entails will strengthen content-based methods by further identifying and learning the mid-level affective attributes.

Finally, academic and industrial researchers yet need to find best practices in incorporating affect in multimedia applications and take it beyond its current applications, e.g., marketing and content retrieval. We hope this cross-disciplinary special section motivates further research on affective analysis in multimedia in both multimedia computing and affective computing communities.

Acknowledgments

The guest editors of this special section would like to thank all the anonymous reviewers for their valuable comments. We would also like to thank Samantha Jacobs for her conscientious coordination during the review process.

References

[1] M. Soleymani, M. Larson, T. Pun, and A. Hanjalic, "Corpus development for affective video indexing," *IEEE Trans. Multimedia*, vol. 16, no. 4, pp. 1075–1089, Jun. 2014.

[2] R. W. Picard, *Affective Computing*. Cambridge, Ma, Usa: MIT Press, 1997.

[3] M. K. Abadi, R. Subramanian, S. M. Kia, P. Avesani, I. Patras, and N. Sebe, "Decaf: MEG-based multimodal database for decoding affective physiological responses," *IEEE Trans. Affective Comput.*, vol. 6, no. 3, pp. 209–222, Jul.-Sep. 2015.
[4] D. McDuff, R. Kaliouby, J. Cohn, and R. Picard, "Predicting ad liking and purchase intent: Large-scale analysis of facial responses to ads," *IEEE Trans. Affective Comput.*, vol. 6, no. 3, pp. 223–235, Jul.-Sep. 2015.
[5] J.-M. Ren, M.-J. Wu, and J.-S. R. Jang, "Automatic music mood classification based on timbre and modulation features," *IEEE Trans. Affective Comput.*, vol. 6, no. 3, pp. 236–246, Jul.-Sep. 2015.
[6] Y. Liu, Y. Liu, Y. Zhao, and K. A. Hua, "What strikes the strings of your heart?—Feature mining for music emotion analysis," *IEEE Trans. Affective Comput.*, vol. 6, no. 3, pp. 247–260, Jul.-Sep. 2015.
[7] J.-C. Wang, Y.-R. Chen, W.-C. Hsieh, Y.-S. Lee, and Y.-H. Chin, "Hierarchical Dirichlet process mixture model for music emotion recognition," *IEEE Trans. Affective Comput.*, vol. 6, no. 3, pp. 261–271, Jul.-Sep. 2015.
[8] T.-S. Park and B.-T. Zhang, "Consensus analysis and modeling of visual aesthetic perception," *IEEE Trans. Affective Comput.*, vol. 6, no. 3, pp. 272–285, Jul.-Sep. 2015.
[9] X. Wang, J. Jia, J. Tang, B. Wu, L. Cai, and L. Xie, "Modeling emotion influence in image social networks," *IEEE Trans. Affective Comput.*, vol. 6, no. 3, pp. 286–297, Jul.-Sep. 2015.
[10] Y.-Y. Chen, T. Chen, T. Liu, H.-Y. M. Liao, and S.-F. Chang, "Assistive image comment robot—A novel mid-level concept-based representation," *IEEE Trans. Affective Comput.*, vol. 6, no. 3, pp. 298–311, Jul.-Sep. 2015.
[11] T. Nguyen, T. Duong, D. Phung, and S. Venkatesh, "Autism blogs: Expressed emotion, language styles and concerns in personal and community settings," *IEEE Trans. Affective Comput.*, vol. 6, no. 3, pp. 312–323, Jul.-Sep. 2015.
[12] Y.-H. Yang, and J.-Y. Liu, "Quantitative study of music listening behavior in a social and affective context," *IEEE Trans. Multimedia*, vol. 15, no. 6, pp. 1304–1315, Oct. 2013.
[13] N. Malandrakis, A. Potamianos, G. Evangelopoulos, and A. Zlatintsi, "A supervised approach to movie emotion tracking," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Proc.*, 2011, pp. 2376–2379.
[14] K. R. Scherer, "What are emotions? And how can they be measured?" *Soc. Sci. Inf.*, vol. 44, no. 4, pp. 695–729, 2005.
[15] M. Larson, M. Soleymani, M. Eskevich, P. Serdyukov, R. Ordelman, and G. Jones, "The community and the crowd: Multimedia benchmark dataset development," *IEEE MultiMedia*, vol. 19, no. 3, pp. 15–23, Jul.–Sep. 2012.

Mohammad Soleymani (S'04-M'12) received the PhD degree in computer science from the University of Geneva, Switzerland, in 2011. From 2012 to 2014, he was a Marie Curie fellow at the Intelligent Behaviour Understanding Group (iBUG), Imperial College London, where he conducted research on sensor-based and implicit emotional tagging. He is currently a Swiss SNF Ambizione fellow at the University of Geneva, Switzerland. His research interests include affective computing, multimedia information retrieval, and multimodal interactions. He is one of the founding organizers of the MediaEval benchmarking campaign. He has served as an associate editor and guest editor for the *IEEE Transactions on Affective Computing* and *Journal of Multimodal User Interfaces*. He has also served as an area chair, program committee member, and a reviewer for multiple conferences and workshops including ACM MM, ACM ICMI, ISMIR and IEEE ICME. He is a member of the IEEE.

Yi-Hsuan Yang (M'11) received the PhD degree in communication engineering from National Taiwan University in 2010. Since 2011, he has been affiliated with Academia Sinica as an assistant research fellow. He is also an adjunct assistant professor with the National Cheng Kung University. His research interests include music information retrieval, machine learning, and affective computing. He received the 2011 IEEE Signal Processing Society (SPS) Young Author Best Paper Award, the 2012 ACM Multimedia Grand Challenge First Prize, and the 2014 Ta-You Wu Memorial Research Award of the Ministry of Science and Technology, Taiwan. He is an author of the book *Music Emotion Recognition* (CRC Press 2011) and a tutorial speaker on music affect recognition in the International Society for Music Information Retrieval Conference (ISMIR 2012). In 2014, he serve as a Technical Program co-chair of ISMIR and a guest editor of the *IEEE Transactions on Affective Computing*. He is a member of the IEEE.

Go Irie received the PhD degree in information science and technology from The University of Tokyo in 2012. Currently, he is working as a research engineer at NTT Corporation, Japan. From 2012 to 2013, he was a visiting research scholar at Columbia University. His research interests include multimedia information retrieval and multimodal analysis. He is a member of the IEEE.

Alan Hanjalic (M'99-SM'08) is a professor of computer science and head of the Multimedia Computing Group, Delft University of Technology, Netherlands. His research focus is on multimedia information retrieval and recommender systems. He is a chair of the Steering Committee of the *IEEE Transactions on Multimedia*, associate editor-in-chief of the *IEEE MultiMedia Magazine* and member of editorial boards of the *ACM Transactions in Multimedia*, *IEEE Transactions on Affective Computing* and the *International Journal of Multimedia Information Retrieval*. He has been involved in organization of top conference venues in the multimedia domain, including ACM Multimedia (General Chair 2009 and 2016, Program Chair 2007 and 2014). He is a senior member of the IEEE.