Gang Hua
Stevens Institute of Technology

Emotional and Social Signals: A Neglected Frontier in Multimedia Computing?

Hatice Gunes
Queen Mary University of London

Hayley Hung
Delft University of Technology

Although most multimedia data is made by people and for people, the role of emotional and social signals in multimedia has not been a core concern of the multimedia research community. At the 22nd ACM International Conference on Multimedia, a panel titled "Looking for Emotional and Social Signals in Multimedia: Where Art Thou?" aimed to investigate this further and revealed major gaps in the formulation, understanding, and application of emotional and social signal processing in the multimedia domain. Here, we attempt to break down and understand the challenges in bringing this new domain to multimedia, summarizing current feelings in the research community based on discussions during the panel.

What Are Emotional and Social Signals?

In the spirit of encouraging more people in the multimedia community to recognize the potential of working in this domain, we first define emotional, affective, and social signals. (Also see the "Definitions of the Different Fields" sidebar for more information.)

Emotional and Affective Signals

In keeping with the new area of "emotional and social signals in multimedia" at ACM Multimedia '14, we use the term *emotional signals*. In reality, what we consider as part of such emotional signals is much broader, ranging from *affect* to *feelings* and *moods*. Robert Masters makes the following distinctions between affect, feeling, and emotion:

> Affect is an innately structured, non-cognitive evaluative sensation that may or may not register in consciousness; feeling is affect made conscious, possessing an evaluative capacity that is not only physiologically based, but that is often also psychologically (and sometimes relationally) oriented; and emotion is psychosocially constructed, dramatized feeling.[1]

Further distinctions between attitudes, moods, affect dispositions, interpersonal stances, and various emotions appear elsewhere.[2]

Affect can be recognized from visible/external signals—such as text, gestures (facial expressions, body gestures, head movements, gait, and so on), and speech (what we say and how we say it)—or invisible/internal signals—such as physiological signals (heart rate, skin conductivity, salivation, and so on) and brain and scalp signals.

Social Signals

Traditional social signals (as shown in Figure 1) are signals that can be transmitted by a sender and perceived by one or more receivers. In the context of human-human social interactions, such signals are used by the receiver in forming judgements about the sender, which then affects the receiver's behavior. Social signals include verbal behavior, such as text and speech, or nonverbal behavior, such as face, head, and body behavior; physical appearance; and positioning of oneself with respect to others and the environment.

With the growth of online social media, the traditional understanding of social signals has expanded to include any socially relevant information that is broadcast online indicating marital status, "likes," +1s, tweets, retweets, shares, and status messages. Sentiment-based social and emotional signals are perhaps the most exploited in multimedia research today. However, this panel discussion hoped to open the debate about the use of social signals from direct face-to-face human-human and human-computer interaction, which has received less attention in multimedia research.

1070-986X/15/$31.00 © 2015 IEEE

Published by the IEEE Computer Society

Definitions of the Different Fields

Affective computing aims to equip (multimedia) computing devices, such as PCs, smartphones, and media players, with the means to retrieve, interpret, understand, and respond to emotions, affect, and moods—similar to the way humans rely on their senses to assess each other's communicative and affective states.

Social signal processing aims to develop methods to automatically detect nonverbal behavioral cues that can be used to further infer social evaluations such as attitudes, traits, or social hierarchy from sensor data. From such estimates, the aim is also develop methods to respond in socially recognizable ways via the synthesis of nonverbal behavior.

Multimedia computing aims to understand how to develop methods to effectively create, store, analyze, search, and distribute multimedia content. To achieve this effectively requires an understanding of the human needs of multimedia services in terms of interaction with the system and large-scale processing, delivery, and storage.

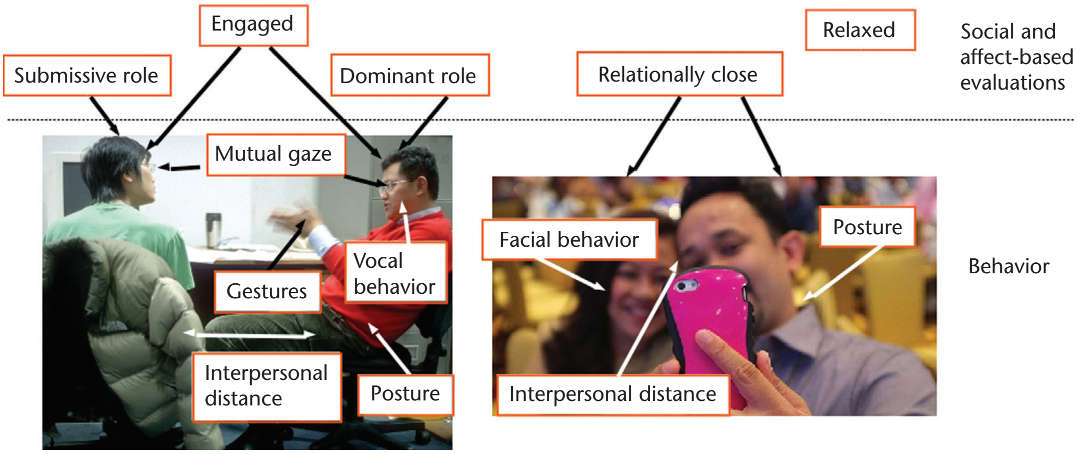

Figure 1. Illustrating examples of social and emotional signals and social evaluations.

Where Are Emotional and Social Signals?

Emotional and social signals exist in multimedia as signals evoked by media and in the context that surrounds multimedia.

Signals in Multimedia

Affective and social signals in media refer to people in multimedia (video or audio files) who are interacting and communicating affective and social signals using language, vocal intonation, facial expression, head movement, body movement, and posture. The widespread nature of such content has also brought about interest in emotion and mood classification in multimedia—such as sentiment and opinion mining in text or mood classification in music.

Signals Evoked by Media

Affective and social signals evoked by media can be analyzed by focusing on the behavior evoked in human observers when they hear sounds or see images and videos displayed. Analyzing media for affect-related tagging and annotation is a relatively new field of research appearing at multimedia conferences.

Signals in the Context Surrounding Multimedia

Affective and social signals also exist indirectly via the social context that surrounds the multimedia. For example, the social context when taking a photo represents information about who (if anyone) we were with when we took the photo, our relationship with those people, our attraction to them, our intentions for sharing the content, and so on.

Figure 2 summarizes the types of social context that can surround multimedia at the point of creation. For example, a photo of a beach taken during a conference with colleagues clearly has a different meaning from a photo

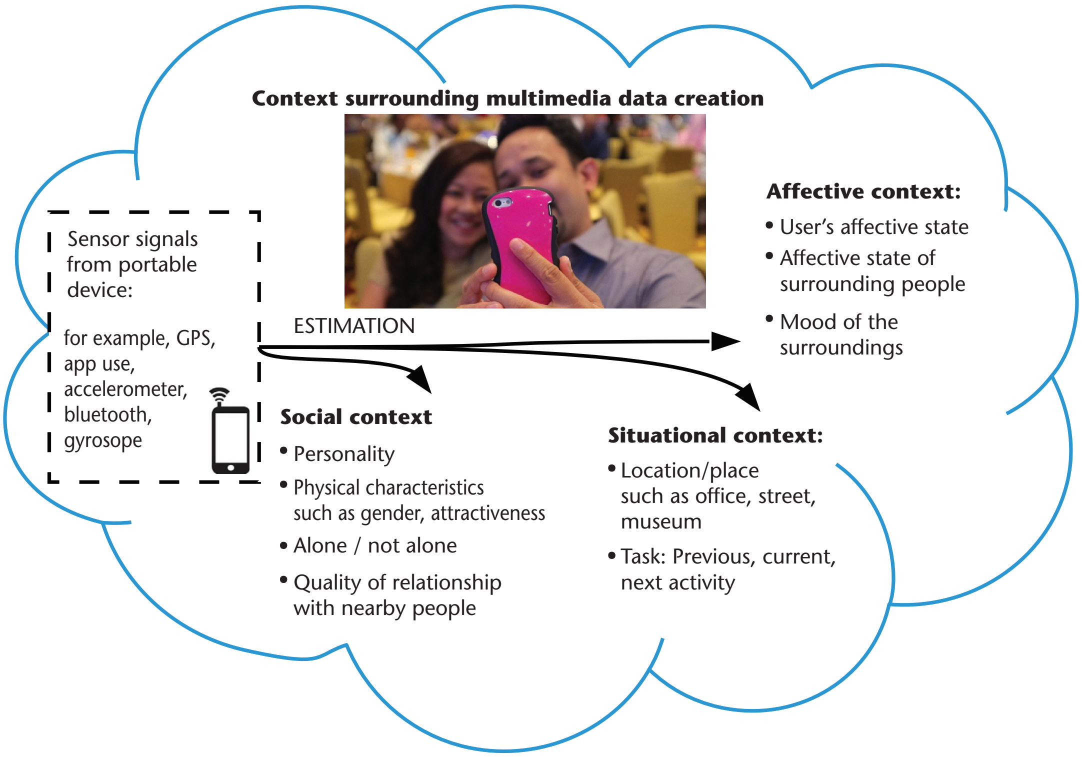

Figure 2. The emotional and social context that surrounds multimedia content at the point of creation.

taken of the same scene during a holiday with a romantic partner. In the latter, the photographer would experience a certain affect due to the romantic holiday, which would not be experienced with colleagues. Moreover, the mood of the environment, as characterized by the smells, temperature, sounds, and sights of the surrounding area, would also influence the person's affective state when taking the photo.

Even the mere act of sharing an image can be considered a social act; certain images are only shared with family while others are only shared with a certain group of friends. Sharing a picture of live train times might be used to express frustration about someone's traveling options. Social media has enabled an explosion in the metadata associated with multimedia content, such as tags, reactions, likes, and comments. Therefore, the number of likes that are associated with a shared photo can also be viewed as a representation of its context, because it represents a form of sentiment associated with the content.

Finally, the experiences related to multimedia—such as watching videos, video conferencing, or listening to music—are all influenced by social context. You might expect a higher fidelity audio signal from music you are listening to on your own compared to music played at a party.

The State of the Art

Figure 3 summarizes the core differences in current multimedia research and affective and social signal processing research with respect to four key areas in multimedia computing: creation and authoring, delivery and storage, analysis and understanding, and application. The separation of work in these fields is more for illustrative purposes, and exceptions exist where research spans the two domains.

Perhaps most noteworthy is that while much of multimedia research occurs within the analysis and the understanding and recommendation domains, affective computing and social signal processing have tended to stay within the analysis area. Moreover, research in affective and social signal processing tends to consider more narrow application domains. This is surprising, given the potential for exploiting affective and social signals for search and recommendation, delivery, and quality of experience—to name but a few points of connection.

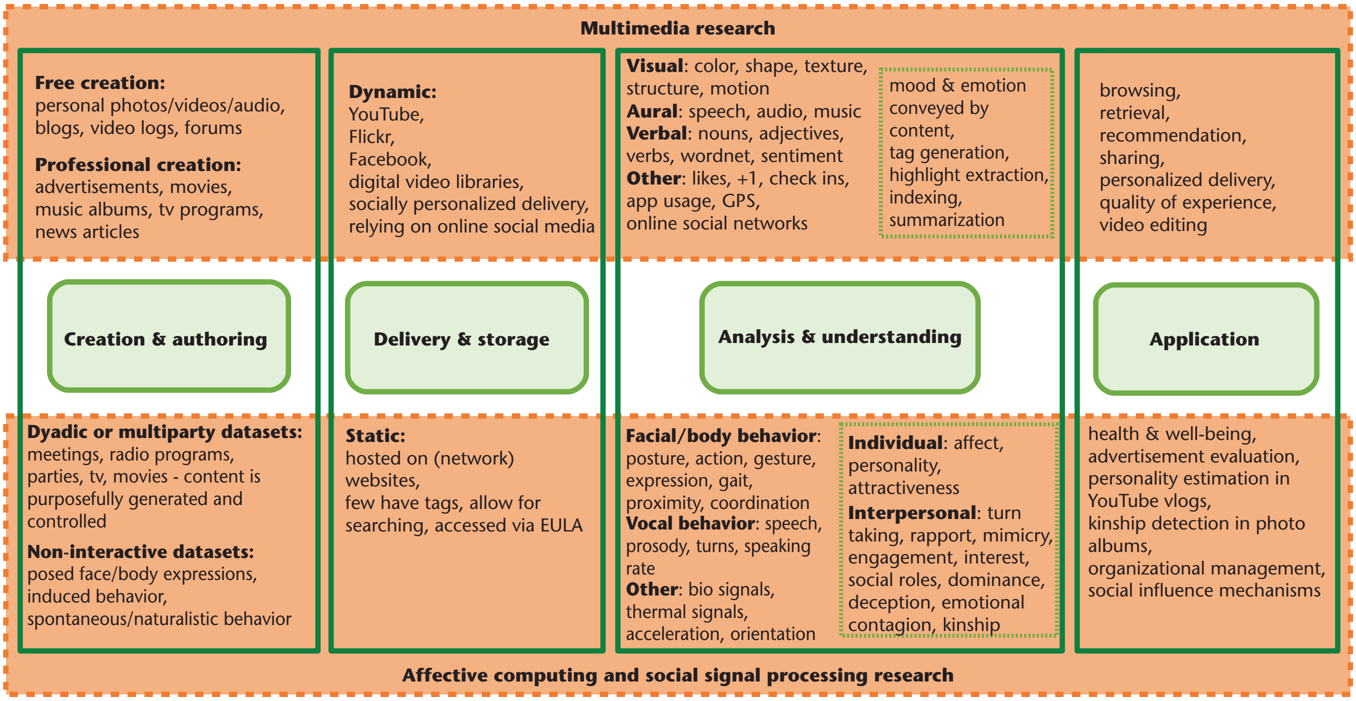

Figure 3. Core differences in current multimedia research and affective and social signal processing research, with respect to four key areas in multimedia computing: creation and authoring, delivery and storage, analysis and understanding, and application.

Annotating Images and Videos

The lowest entry point into the use of social and affective signals in multimedia systems has been on the level of social media data used to annotate image or video content. The mining of verbal content from social media has been easily accomplished by exploiting the word-level statistics rather than more complex language understanding. Thus far, and particularly in large areas of the multimedia community (such as content analysis, search, and recommendation), such verbal content has been exploited on a systematic level.

Analyzing Multimedia

Aside from textual and written information that users provide alongside multimedia content, multimodal human behavioral information that is present within the media provides a vast source of information, which we refer to as *affective and social content*. Multimedia content is heavily loaded with affective and social signals in speech, music, sound, text, and video contained in the media itself (TV news, advertisements, and movies). The majority of the work in automatic affect analysis can be categorized as generating and analyzing media (videos, images, audio clips) with affective content. The main focus has been on generating and analyzing exaggerated "single" human behavior content using a single modality—namely, visual or vocal—and classifying this content into the seven predefined basic emotion categories of neutral, happiness, sadness, surprise, fear, anger, and disgust.[3]

More recently, the focus has shifted toward

- analyzing multimodal affective human behavior content in terms of emotion dimensions (such as arousal, valence, or power), which enable the representation of any emotion encountered in daily life interactions;
- analyzing the nature of the behavior exhibited in the media (such as acted versus naturalistic behavior); and
- analysis in context—that is, in situations encountered in daily life, including the analysis of pain and depression, embarrassment, engagement, enjoyment, and boredom.[4]

There have also been pioneering attempts to acquire and analyze big (in terms of number, variety, and scope) multimodal data with affective human behavior content.[5] In all of these works, although the data itself can be considered multimedia, the research was mostly

Some recent works have started focusing on the automatic prediction of social evaluations from multimedia data, including the estimation of kinship in photo albums.

conducted without considering the needs of the multimedia computing community.

Analyzing the Evoked Affect

Analyzing the affect that is evoked from consuming multimedia is one new field of research appearing at multimedia conferences. Affect evoked in human observers by the media content, such as objects, colors, lights, or higher level phenomena (such as a sunset), have been studied extensively. Similar to analyzing media with evoked affective content, the main focus has been on the use of the seven predefined basic emotion categories.

In recent years, the use of the emotion dimensions of arousal and valence as a continuum with different scales has also been adopted. According to the dimensional approach, emotions and affective states relate to one another in a systematic manner. In this approach, the majority of affect variability is covered by three dimensions: valence, arousal, and potency (dominance).[6,7] The valence dimension refers to how positive or negative the emotion is and ranges from unpleasant feelings to pleasant feelings of happiness. The arousal dimension refers to how excited the emotion is and ranges from sleepiness or boredom to frantic excitement. The power dimension refers to the sense of control the person has over the emotion.

An interesting approach to quantify the emotions evoked by media has been that of *implicit human-centered tagging,* where the aim is to automatically tag multimedia data by analyzing users' nonverbal reactions (such as laughter and accelerated heartbeat)[8] to such data or body movements.[9] Related works have analyzed smiles to predict "liking" and "desire to view again" of online video ads;[10] analyzed audiovisual laughter and mirth to judge whether the joke in the video presented was funny;[11] looked at facial expressions and EEG signals simultaneously, and other multimodal behavioral cues of the observers, for affective tagging;[12] and analyzed the emotions of the observers in response to music videos.[13]

Analyzing Evoked Social Signals

Pioneering work by Alex Pentland and his colleagues has showed that it is possible to predict the outcome of a job interview, including the prediction of which candidate would get the job, by analyzing the multimedia content of the interviews. Overall, the nonverbal signals of the interviewees were fundamental in determining how the interviewers perceived them.[14] Such social signals, which tend to exist in the multimedia data itself, have mostly been analyzed for the development of multimedia systems for smart meeting rooms.[15] This has led to developing techniques to automatically interpret what happens in meetings to help browse meeting content (for example, meeting segmentation work, speaker localization, dominance estimation, interest estimation, personality estimation, and functional role estimation), mediate meetings,[16] or to automatically retrieve relevant documents during meetings.[17]

Some recent works have started focusing on the automatic prediction of social evaluations from multimedia data, including the estimation of kinship in photo albums;[18] personality from speech,[19] video blogs,[20] and dyadic interaction recordings;[21] beauty and attractiveness from facial images and facial behavior;[22] and likeability from speaker styles. Finally, there has been some promising research on transferring knowledge from large-scale online data to better predict perceptions of personality types during meetings.[23]

Exploiting Affective and Social Context

Context analysis attempts to answer the "W5+" questions of *who* (user), *with whom* (other people), *what* (task), *where* (location), *when* (the time of the observed behavior), and *how* and *why* (meaning and interpretation) that surround the multimedia content. Context significantly affects the way multimedia is captured, analyzed, shared, and interpreted. In the domain of multimedia delivery, recent studies have shown that the social context influences perceptions of video quality[24]—there are differences in perception across genders when watching sports games in groups. Social context, and in particular, who is conversing with whom, is vitally important for establishing boundaries for photo taking in crowded social events.[25]

In the context that surrounds Flickr images, liking something transmits a social message about us, such as our personality.[26,27] Social media is based around this notion of sharing. Text, images, videos, and music are shared, tagged, and grouped, and reactions via likes and comments are possible. Therefore, content can be indexed, searched, and clustered based on tags and preferences.

Context has also been explored for use in automatically analyzing attractiveness and personality traits. For example, facial behavioral features (such as smiling or blinking), in addition to the traditionally employed static features (such as facial proportions), improve the prediction of facial attractiveness.[22] Also, changes in situational context (with whom the person is interacting) and availability of different modalities (visual versus audio-visual) cause changes in the perception and automatic prediction of personality traits.[21]

Figure 4. Panelists discussing "Looking for Emotional and Social Signals in Multimedia: Where Art Thou?" at ACM Multimedia 2014 (right to left): Elisabeth Andre (University of Augsburg, Germany), Dick Bulterman (FXPAL, US), Alex Hauptmann (Carnegie Mellon University), Rainer Lienhart (University of Augsburg, Germany), and Nicu Sebe (University of Trento, Italy).

Panel Summary

The panel consisted of two parts: presentations and discussion. We started with an introduction of affective and social signals in multimedia, providing the motivation for organizing the panel together with a summary of progress in relevant fields. The panelists were asked to tell the audience about the relevance of emotional and social signals to their area of expertise (see Figure 4). The panel discussion then focused around the following questions:

- Where and what—where are emotional and social signals in multimedia?
- Context or content—what if the meaning of the content can be better obtained from the context surrounding the content? Do emotional and social dimensions help?
- Closing the gap—should future multimedia systems focus on incorporating emotional and social signals? Where are the gaps?

During the discussions, we could see and hear what the "believers" saw as achievements and challenges, and what was really bothering the "doubters."

The Believers

Analyzing and fusing data from multiple modalities and doing context-sensitive analysis have been two approaches that have worked well, although working on naturalistic multimodal data has been much more challenging than working on single-modal and posed data. Panelists reflected on the challenge of ascribing personal meaning to multimedia data, as this can be influenced by a highly time-varying context. Real-time and multimodal analysis, leading toward emotional and social signal processing, understanding, and interpretation in the wild, are the topics that have been neglected by researchers in relevant fields.

Emotional and social signals have been relevant also for affective cinema that understands (personalized) emotions of the viewer, and for attention and personality analysis in unrestricted social scenarios, such as parties versus meetings, with a pre-defined agenda (that is, "party behavior" might be more reflective of an individual's personality via his or her use of physical space.)

The multimedia community currently looks from a very narrow angle at images or multimedia, and the field is in need of tools that will allow richer descriptions.

The Doubters

One panelist mentioned that his or her research work on digital video libraries and viral videos focused mostly on analyzing and working with metadata and that the use of emotions in training classifiers gave no performance improvement. This could be because the intent when creating multimedia content is different from the consumers' reactions to that content, the emotions are context dependent and thus cannot be generic or generalized, or the emotions are subtle and beyond the current crude tools that the various research fields have to offer.

Another panelist voiced that his work so far would not be actually considered in the scope of the panel discussion, and he saw himself as representing the doubters for whom the emotional and social signals have been fuzzy concepts that are difficult to define.

The Challenges

A major challenge experienced by the believers was that reviewers in multimedia venues could not see how emotional and social signals fit into multimedia research, particularly in cases where nontraditional modalities such as biosignals were used. Once this challenge was voiced, even the doubters acknowledged that multimedia is present in emotional and social signal analysis, as long as it considers the user experience from multiple streams. However, the multimedia community currently looks from a very narrow angle at images or multimedia, and the field is in need of tools that will allow richer descriptions.

One might argue that the signals around a photo are more valuable than the photo itself, so we need to capture the richness around the photo, but the multimedia research field and community currently are not building systems to capture this content or exploit it for understanding how the users ascribe meaning to multimedia content they create or share.

Problems to Address

Despite reservations, the panel emphasized that there is a huge potential in researching emotional and social signals—particularly in the context of big multimodal data from personal devices—for better understanding of what is going on and identifying the real needs. This in turn will lead to efficiency in energy, space, time, and services; for example, in terms of infrastructure (municipality and city), health (what impacts long-term health?), and humanity and society (what influences the role of the individual in society and how do we become a better society?).

Because emotions are personal, and most people want to hide their emotions except for special occasions, how do we then address privacy preservation issues? How does the research field move away from individuals and apps in controlled environments to individuals and groups in the wild? How can emotional and social signals benefit from more dynamic content sharing (toward "together anywhere, together any time") where real events are being recorded, streamed, shared, exposed, and experienced by users and people around the globe? How can multimedia represent multiple meanings for multiple individuals or sets of individuals? Would introducing emotional and social dimensions help?

There were both "yes" and "no" answers to these questions. Focusing on emotional and social dimensions would help with understanding the content through the context, particularly because the impact of multimedia and what it contains can get measured only later, which indeed becomes part of the context. Yet the problem of looking for emotions and social signals in multimedia is challenging, because they can be subtle, complex, and multilayered.

Bridging the Gap

Multimedia research aims to address problems that exist as a result of human-generated multimedia data. Humans are a key part of the process, and their natural ability to communicate affective and social information about themselves is therefore a fundamental part of the process by which data is captured, analyzed, stored, consumed, and distributed. Here, we summarize the bottlenecks in closing the gap.

Unclear Vision

We hope that this article has helped to more clearly specify the links between emotional and social signals and multimedia research by summarizing achievements in the fields so far. In particular, we hope that Figure 3 illustrates more precisely the existing gaps and how research that crosses these boundaries can better support multimedia research.

Misperceptions about the "Other Field"

The research panel discussion, together with post-panel feedback and comments, indicated that multimedia researchers are wary of getting into fuzzy phenomenon such as affective and social signals. The majority of researchers in social and emotional psychology aim to identify statistically generalizable behaviors—and these are often the behaviors that researchers in affective computing and social signal processing derive experiments from first. The panel successfully brought together researchers in both social and affective signals with those working in core multimedia analysis and retrieval domains and enabled a concrete dialogue for the first time, even from skeptical outsiders.

No Time to Talk

Still, many "doubters" exist, and, as with any cross-disciplinary research, more dialogue between the researchers in affective computing and social signal processing and multimedia is needed to concretely reduce the problem space into clearly achievable paths. This will require more dialogue, and it will take time and effort to learn about other areas of research. However, we strongly believe that there are major benefits to closing this gap. Much of the research that could affect multimedia applications revolves around our understanding of how to model humans' emotional, affective, and social needs with respect to multimedia.

Multimedia research aims to address problems that exist as a result of human-generated multimedia data. Humans are a key part of the process.

Future Research Directions

It is clear that social and emotional signals are highly embedded in many forms of multimedia research. As computing becomes more pervasive, where everything can be packaged as a mobile application and storage and processing are in the cloud, the *offline* and the *online* worlds move ever closer together. This, in turn, pushes the fields of social, behavioral, and emotional psychology to become inherently embedded in multimedia systems. To further prepare for this trend, we need to start breaking down the related research questions now. Summarizing the discussion in the panel, we categorize two open areas for which multidisciplinary research is needed.

Exploiting Context

Effective multimedia recommendation from mobile devices is already here. However, it has yet to become seamlessly integrated into people's lives.

Open questions include the following: What if the meaning of the content can be better obtained from the context surrounding the content? Does introducing emotional and social dimensions help? How much of the context can we capture and record? Which aspects of what is recorded actually contribute to a better understanding of the content? How do we represent a highly time-dependent and dynamic context?

Addressing the Real World

Moving away from individuals and applications in controlled environments and toward individuals and groups in natural situations, we need to focus more on real-time emotional and social signal processing, understanding and interpretation of big multimodal data from personal devices (the new multimedia data). Behavioral psychologists can give us indications, often in laboratory conditions, of how someone might behave, but when we extrapolate to settings in the wild and outside of the laboratory to solve real problems for real users in real environments, we end up facing unexpected issues and challenges.[5]

Open questions include the following: How do we train our models to handle the multiplicity of the real world? At the first level, we need benchmarks in the real world. However, collecting data in the wild is a challenging task. The datasets tend to be smaller as we try to devise the right experimental conditions, gather volunteers, and obtain ecologically valid data and ground truth in an accurate and privacy-preserving manner to explore certain behavioral phenomena. To what extent can we transfer knowledge from controlled settings to the real world to improve multimedia systems operating in the wild?

Progress in these areas will largely depend on ongoing communication regarding—and a mutual understanding of—how we engage with each other and with other communities; how we share what has been achieved, failed, and learned; and how we identify areas in which we need help. This will create the fertile soil needed to develop groundbreaking ideas and projects and fruitful collaboration between multimedia computing and affective and social signal processing fields.

Acknowledgments

This panel would not have been made possible without the contributions of a number of researchers. We thank our panel co-organizers Björn Schuller and Lexing Xie for their contributions during the proposal and planning of the proposal. Thanks to Dick Bulterman, Rainer Lienhart, Alex Hauptmann, Nicu Sebe, and Elisabeth Andre for being stimulating panelists. Thanks to Albert Ali Salah and Oya Aran, who helped start the new area on emotional and social signals at ACM Multimedia. Thanks also to Nicu Sebe, Alan Hanjalic, Alessandro Vinciarelli, and Daniel Gatica-Perez for providing structural support in forming the area and getting it started. Finally, thanks to Mohammad Soleymani and Martha Larson for helpful discussions related to the topics covered in this article. (Also note that both authors contributed equally to this article.)

References

1. R. Masters, "Compassionate Wrath: Transpersonal Approaches to Anger," *J. Transpersonal Psychology*, vol. 32, no. 1, pp. 31–51, 2000.
2. K. Scherer, "What Are Emotions? And How Can They Be Measured?" *Social Science Information*, vol. 44, no. 4, 2005, pp. 695–729.
3. P. Ekman and W.V. Friesen, *Unmasking the Face: A Guide to Recognizing Emotions from Facial Clues*, Prentice-Hall, 1975.
4. H. Gunes and B. Schuller, "Categorical and Dimensional Affect Analysis in Continuous Input: Current Trends and Future Directions," *Image and Vision Computing*, vol. 31, no. 2, 2013, pp. 120–136.
5. R.W. Picard, "Affective Media and Wearables: Surprising Findings," *Proc. ACM Int'l Conf. Multimedia* (MM), 2014, pp. 3–4.
6. J.A. Russell, "A Circumplex Model of Affect," *J. Personality and Social Psychology*, vol. 39, no. 6, 1980, pp. 1161–1178.
7. H. Gunes and M. Pantic, "Automatic, Dimensional and Continuous Emotion Recognition," *Int'l J. of Synthetic Emotions*, vol. 1, no. 1, 2010, pp. 68–99.
8. M. Pantic and A. Vinciarelli, "Implicit Human-Centered Tagging," *IEEE Signal Processing*, vol. 26, no. 6, 2009, pp. 173–180.
9. G. Englebienne and H. Hung, "Mining for Motivation: Using a Single Wearable Accelerometer to Detect People's Interests," *Proc. 2nd ACM Int'l Workshop on Interactive Multimedia on Mobile and Portable Devices*, 2012, pp. 23–26.
10. D. McDuff et al., "Automatic Measurement of Ad Preferences from Facial Responses Gathered over the Internet," *Image and Vision Computing*, vol. 32, no. 10, 2014, pp. 630–640.
11. S. Petridis and M. Pantic, "Is This Joke Really Funny? Judging the Mirth by Audiovisual Laughter Analysis," *Proc. Int'l Conf. Multimedia & Expo* (ICME), 2009, pp. 1444–1447.
12. S. Koelstra and I. Patras, "Fusion of Facial Expressions and EEG for Implicit Affective Tagging," *Image and Vision Computing*, vol. 31, no. 2, 2013, pp. 164–174.
13. M. Soleymani et al., "Continuous Emotion Detection in Response to Music Videos," *Proc. Int'l Conf. Automatic Face and Gesture Recognition* (FG), 2011, pp. 803–808.
14. A. Pentland, *Honest Signals: How They Shape Our World*, MIT Press, 2008.
15. D. Gatica-Perez, "Automatic Nonverbal Analysis of Social Interaction in Small Groups: A Review," *Image and Vision Computing*, vol. 27, no. 12, 2009, pp. 1775–1787.
16. T. Kim et al., "Meeting Mediator: Enhancing Group Collaboration Using Sociometric Feedback," *Proc. 2008 ACM Conf. Computer Supported Cooperative Work*, 2008, pp. 457–466.
17. A. Popescu-Belis et al., "The AMIDA Automatic Content Linking Device: Just-in-Time Document

Retrieval in Meetings," *Machine Learning for Multimodal Interaction*, Springer, 2008, pp. 272–283.
18. S. Xia et al., "Understanding Kin Relationships in a Photo," *IEEE Trans. Multimedia*, vol. 14, no. 4, 2012, pp. 1046–1056.
19. G. Mohammadi and A. Vinciarelli, "Automatic Personality Perception: Prediction of Trait Attribution Based on Prosodic Features," *IEEE Trans. Affective Computing*, vol. 3, no. 3, 2012, pp. 273–284.
20. J. Biel and D. Gatica-Perez, "The YouTube Lens: Crowdsourced Personality Impressions and Audiovisual Analysis of Vlogs," *IEEE Trans.* Multimedia, vol. 15, no. 1, 2013, pp. 41–55.
21. J. Joshi, H. Gunes, and R. Goecke, "Automatic Prediction of Perceived Traits Using Visual Cues under Varied Situational Context," *Proc. 22nd Int'l Conf. Pattern Recognition* (ICPR), 2014, pp. 2855–2860.
22. S. Kalayci, H.K. Ekenel, and H. Gunes, "Automatic Analysis of Facial Attractiveness from Video," *Proc. IEEE Int'l Conf. Image Processing* (ICIP), 2014, pp. 4191–4195.
23. O. Aran and D. Gatica-Perez, "Cross-Domain Personality Prediction: From Video Blogs to Small Group Meetings," *Proc. 15th ACM Int'l Conf. Multimodal Interaction*, 2013, pp. 127–130.
24. Y. Zhu, I. Heynderickx, and J.A. Redi, "Alone or Together: Measuring Users' Viewing Experience in Different Social Contexts," *IS&T/SPIE Electronic Imaging*, 2014, p. 90140W.
25. T. Gan et al., "Temporal Encoded F-Formation System for Social Interaction Detection," *Proc. 21st ACM Int'l Conf. Multimedia*, 2013, pp. 937–946.
26. M. Cristani et al., "Unveiling the Multimedia Unconscious: Implicit Cognitive Processes and Multimedia Content Analysis," *Proc. 21st ACM Int'l Conf. Multimedia*, 2013, pp. 213–222.
27. Y. Bachrach et al., "Personality and Patterns of Facebook Usage," *Proc. 3rd Ann. ACM Web Science Conf.*, 2012, pp. 24–32.

Hatice Gunes is a senior lecturer (associate professor), leading the Affective and Human Computing Lab in the School of Electronic Engineering and Computer Science at Queen Mary University of London. Contact her at h.gunes@eecs.qmul.ac.uk; www.eecs.qmul.ac.uk/~hatice.

Hayley Hung is an assistant professor in the Pattern Recognition and Bioinformatics Group in the department of Intelligent Systems at the Delft University of Technology. Her research group focuses on investigating research questions related to the automated interpretation of social signals from multiple sensor modalities. Contact her at h.hung@tudelft.nl; http://homepage.tudelft.nl/3e2t5.