Automating the Extraction of Emotion-Related Multimedia Semantics

ARTHUR G. MONEY[*] and HARRY AGIUS[†]
Brunel University, School of Information Systems, Computing and Mathematics
St John's, Uxbridge, Middlesex, UB8 3PH, UK

ABSTRACT: *We propose a contextual user-emotion-based analysis (CUEBA) approach to analysing multimedia content. This approach enables the automated extraction of emotion-related semantics based on user facial expressions and physiological responses while viewing and capturing multimedia content.*

'User-in-the-loop' multimedia content analysis

The digital multimedia information age has meant that the volume and type of information readily available to us is changing. With increasing frequency, users are accessing multimedia documents, such as video and audio files, which are rich in semantic content, but, unlike text, the meaning is embedded in multiple modes of communication. This poses new challenges in computing, modelling and querying such content. Furthermore, the proliferation of Internet-ready devices with various screen sizes, sound capabilities, control interfaces and bandwidth capabilities has meant that users increasingly require personalised multimedia content, based on individual taste, and usage scenarios [1].

Multimedia content analysis is concerned with developing automated methods to identifying and extracting the range of semantics embedded within a multimedia document. This is normally achieved by analysing the digital media signal for low-level features. In the case of video, these features include colour, texture, shape and object motion. Despite many exceptional and innovative approaches, multimedia content analysis is still challenged by a 'sensory gap' that exists between real world 3D objects and events and their 2D multimedia representations and a 'semantic gap' that between abstracted semantics and the users' interpretations of these [2]. This constitutes a loss of real-world and interpretative detail when determining meaning from the media stream in isolation.

Traditionally, solutions for multimedia content analysis have ruled out user involvement as a viable means of assisting the process in pursuit of the dream of full automation and minimal inconvenience to the user. In recent years, however, research [3, 4] has begun to question the feasibility of purely automated approaches and sought to bring the user 'into the loop'. This research has demonstrated that minimal user input at the point of media capture can provide valuable contextual metadata that goes some way to closing the longstanding challenges of the sensory and semantic gaps.

CUEBA – Contextual User-Emotion-Based Analysis of multimedia content

With a view to exploring the 'user-in-the-loop' approach to multimedia content analysis further, coupled with the goal of extracting high-level semantics well suited to offering personalised content to the end user, we propose a **C**ontextual **U**ser-**E**motion-**B**ased **A**nalysis (CUEBA) approach to multimedia content. CUEBA is inspired by work in the HCI domain that focuses on the development of emotionally-intelligent computer systems [5], as well as technological developments, such as SenseWear® PRO2 Armband from Bodymedia, and automatic facial recognition techniques such as those used by Ekman and Friesen [6].

[*] E-mail: arthurmoney@yahoo.com
[†] Corresponding author. E-mail: harryagius@acm.org

Initially, the CUEBA approach requires the user to wear relevant physiological sensing equipment. In this case, the user's Blood Volume Pump (BVP), Heart Rate (HR), Galvanic Skin Response (GSR) and respiratory patterns are measured unobtrusively, via wireless wearable sensors. In addition, a digital camera captures the user's facial expressions. Emotional responses are recorded digitally whilst the user captures, analyses and views multimedia content. The resulting data is processed to abstract emotional criteria such as valence and arousal ratings and, where possible, compute basic emotional categorisation. The abstracted information is stored in the form of emotion metadata and mapped temporally onto the multimedia content. The result is personalised emotional metadata which represents the high-level semantic content of the multimedia document and can be used to deliver generic and personalised summaries and highlights of multimedia content.

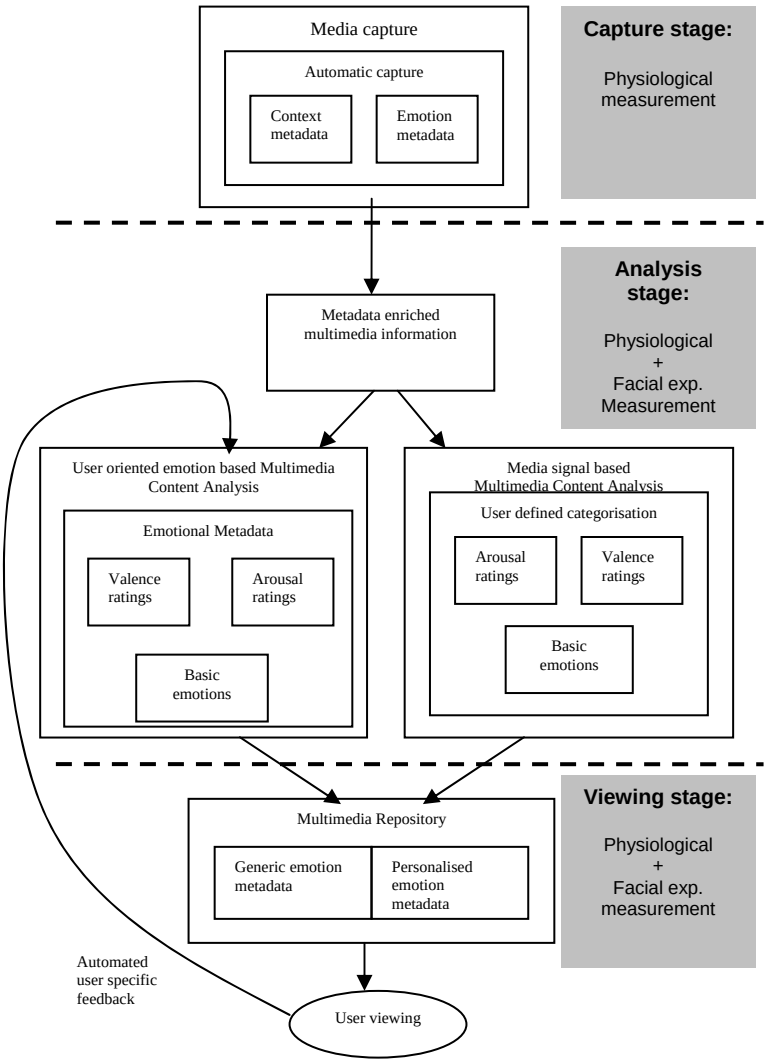

Figure 1: The CUEBA approach.

Figure 1 is an overview of CUEBA, and shows how physiological responses and facial expressions can be applied to generate emotion metadata at the capture, analysis and viewing stages of the multimedia document lifecycle. At the *capture stage*, contextual and physiological data is recorded and bound to the captured content. It is then computed to form metadata-enriched multimedia information and content. The *analysis stage* presents opportunities to collect further emotional metadata in more controlled conditions. Valence

and arousal ratings and basic emotions are then computed using a combination of emotional response data captured at both the capture and analysis stages. These are then used to abstract generic representations of emotional content. Finally, the multimedia content is stored in a multimedia repository together with generic emotional metadata. At the *viewing stage*, multimedia documents are accessed by individual users whose emotional responses are measured and used to develop highly personalised emotion metadata representations of the multimedia content.

Concluding remarks

In pursuit of a fully functional CUEBA approach at every stage of the multimedia life cycle, we initially plan to carry out empirical experimentation to measure physiological aspects of the users' emotional response whilst viewing various popular multimedia content using a ProComp Infiniti ™ 8-channel, multi-modality encoder. Specifically we will be measuring respiration, galvanic skin response, heart rate and blood volume pump. In addition, we will be observe, and record (for later analysis) spontaneous facial responses to multimedia content, and code these using FACS [7]. Whilst developing the CUEBA approach, we aim to:

- Explore the extent to which a variety of common multimedia content, such as film, sports video, and news reports, elicit user responses.
- Develop methods of interpreting user response data, within the various content domains.
- Develop personalised summaries and highlights of multimedia content, based on interpreted user response data.
- Validate empirically via user testing the extent to which user response data can be used to infer personalised summaries, highlights and memorable multimedia content.
- Develop a prescriptive model that enables effective and efficient mapping of user emotional response to multimedia content summaries and highlights.

CUEBA represents a new direction for multimedia content analysis research, adopting a user-in-the-loop approach to content analysis. The approach supports personalised high-level semantic abstraction, whilst requiring no conscious cognitive involvement from the user. This represents a major step to overcoming the semantic and sensory gaps that have challenged multimedia content analysis domain for well over a decade.

References

[1] S. M. Hossain, M. A. Md. Abdur Rahman, and A. El Saddik, "A framework for repurposing multimedia content," *Proc. IEEE Canadian Conference on Electrical and Computer Engineering*, vol. 2, pp. 971-974, 2-5 May 2004.

[2] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1349-1380, 2000.

[3] M. Davis, S. King, N. Good, and R. Sarvas, "From context to content: leveraging context to infer media metadata," *Proc. ACM Multimedia '04*, pp. 188-195, 10-16 October 2004.

[4] R. Nair, "Calculating an aggregated level of interest function for recorded events," *Proc. Proc. ACM Multimedia '04*, pp. 272-275, 10-16 October 2004.

[5] R. W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 10, pp. 1175-1191, 2001.

[6] Ward. R, "An analysis of facial movement tracking in ordinary human-computer interaction," *Interacting with Computers*, vol. 16, no. 15, pp. 879-896, 2004.

[7] P. Ekman, W. V. Friesen and J. C. Hager, *Facial Action Coding System*, A Human Face, Salt Lake City, UT, 2002.