1
Multimedia implicit tagging

Mohammad Soleymani[a] and Maja Pantic[b]

1.1 Introduction

Social and behavioral signals carry invaluable information regarding how audiences perceive the multimedia content. Assessing the responses from the audience, we can generate tags, summaries and other forms of metadata for multimedia representation and indexing. Tags are a form of metadata which enables a retrieval system to find and re-find the content of interest (Larson et al. (2011)). Unlike classic tagging schemes where users direct input is needed, Implicit Human-Centered Tagging (IHCT) was proposed (Pantic and Vinciarelli (2009)) to generate tags without any specific input or effort from users. Translating the behavioral responses into tags results in "implicit" tags since there is no need for users' direct input as reactions to multimedia are displayed spontaneously (Soleymani and Pantic (2012)).

User generated explicit tags are not always assigned with the intention of describing the content and might be given to promote the users themselves (Pantic and Vinciarelli (2009)). Implicit tags have the advantage of being detected for a certain goal relevant to a given application. For example, an online radio is interested in the mood of its songs can assess listeners emotions; a marketing company is interested in assessing in the success of its video advertisements.

It is also worth mentioning that implicit tags can be a complementary

[a] iBUG group, Imperial College London, UK. His work is supported by the European Research Council under the FP7 Marie Curie Intra-European Fellowship: Emotional continuous tagging using spontaneous behavior (EmoTag).
[b] iBUG group, Imperial College London, UK and EEMCS, University of Twente, the Netherlands. Her work is supported in part by the European Community's 7th Framework Programme (FP7/2007-2013) under the grant agreement no 231287 (SSPNet) and ERC Starting Grant agreement no. ERC-2007-StG-203143 (MAHNOB).

source of information in addition to the existing explicit tags. They can be also used to filter out the tags which are non-relevant to the content (Soleymani and Pantic (2013); Soleymani et al. (2013)). A scheme of implicit tagging versus explicit scenario versus explicit tagging is shown in Fig. 1.1. Recently, we are witnessing a growing interest from industry on this topic (Klinghult (2012); McDuff et al. (2012); Fleureau et al. (2013); Silveira et al. (2013)) which is a sign of its significance.

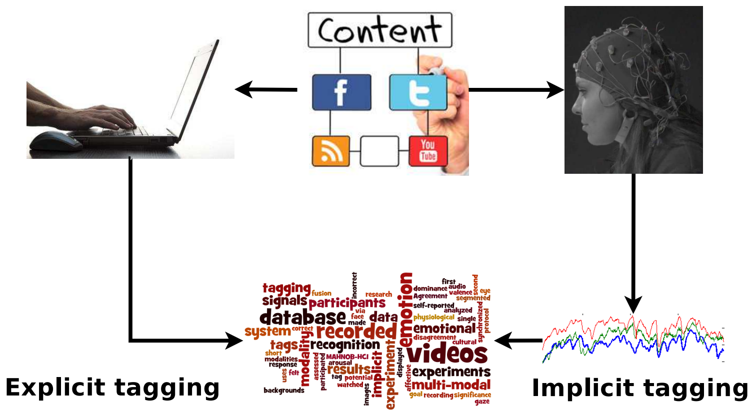

Figure 1.1 Implicit tagging vs. explicit tagging scenarios. The analysis of the bodily reactions to multimedia content replaces the direct input of the tag by users. Thus, users do not have to put any effort into tagging the content.

Analyzing spontaneous reactions to multimedia content can assist multimedia indexing with the following scenarios: (i) direct translation to tags: users spontaneous reactions will be translated into emotions or preference, *e.g.*, interesting, funny, disgusting, scary (Kierkels et al. (2009); Soleymani et al. (2012b); Petridis and Pantic (2009); Koelstra et al. (2010); Silveira et al. (2013); Kurdyukova et al. (2012)); (ii) assessing the correctness of explicit tags or topic relevance, *e.g.*, agreement or disagreement over a displayed tag or the relevance of the retrieved result (Koelstra et al. (2009); Soleymani et al. (2012a); Arapakis et al. (2009b); Jiao and Pantic (2010); Moshfeghi and Jose (2013)); (iii) user profiling: a user's personal preferences can be detected based on her reactions to retrieved data and be used for re-ranking the results; (iv) content summarization: highlight detection is also possible using implicit feedbacks from the users (Fleureau et al. (2013); Joho et al. (2010); Chênes et al. (2012)).

Classic multimedia indexing relies on concepts that characterize its content in terms of events, objects, locations, etc. The indexing that only relies on the concepts depicted in the content is called cognitive indexing. Parallel to this approach to indexing, an alternative has also emerged that take affective aspects into account. Affect, in this context, refers to the intensity and type of emotion that is evoked in the consumer of multimedia content (Hanjalic and Xu (2005); Soleymani et al. (2014)). Multimedia affective content can be presented by relevant emotional tags. Being directly related to the users' reaction, implicit tagging directly translates users' emotions into affective representation of multimedia. Affective tags are shown to help recommendation and retrieval systems to improve their performance (Shan et al. (2009); Tkalčič et al. (2010b); Kierkels et al. (2009)).

Other feedbacks from users, including clickthrough rate, dwell time, have been used extensively for information retrieval and topic relevance applications (Shen et al. (2005); Joachims et al. (2005)). In this chapter, we only cover the implicit feedback which is measurable by sensors and cameras from bodily responses. The reminder of this chapter is organized as follows. Section 1.2 provides a background on the recent developments in this field. Available public databases are introduced in Section 1.3. Current challenges and perspectives are discussed in Section 1.4.

1.2 Background

Implicit tagging have been applied to different problems from emotional tagging and preference detection to topic relevance assessment (Soleymani and Pantic (2012)). Currently, there are three main research trends taking advantage of implicit tagging techniques. The first one deals with using emotional reactions to detect users' emotions and content's mood using the expressed emotion, e.g., laughter detection for hilarity (Petridis and Pantic (2009)); the second group of research is focused detecting interest of the viewers and video highlights; the third group of studies are using the spontaneous reactions for information retrieval or search results re-ranking, e.g., eye gaze for relevance feedback (Hardoon and Pasupa (2010)). In the following we review the existing work categorized by their applications.

1.2.1 Emotional tagging

Emotional tags can be used for indexing the content with their affect as well as improving content recommendation (Shan et al. (2009); Kierkels et al. (2009)). Affective information has been shown to improve image and music recommendation (Tkalčič et al. (2010b); Shan et al. (2009)). Tkalčič et al. used affect detected from facial expression in response to images for an image recommender. Their experimental results showed that the affective implicit tags could improve the explicit tagging as a complementary source of information (Tkalčič et al. (2013)).

Physiological signals have been also used to detect emotions with the goal of implicit emotional tagging. Soleymani et al. (Soleymani et al. (2009)) proposed an affective characterization for movie scenes using peripheral physiological signals. Eight participants watched 64 movie scenes and self-reported their emotions. A linear regression trained by relevance vector machines (RVM) was utilized to estimate each clip's affect from physiological features. Kierkels et al. (Kierkels et al. (2009)) extended these results and analyzed the effectiveness of tags detected by physiological signals for personalized affective tagging of videos. Quantized arousal and valence levels for a clip were then mapped to emotion labels. This mapping enabled the retrieval of video clips based on keyword queries. A similar approach was taken using a linear ridge regression for emotional characterization of music videos. Arousal, valence, dominance, and like/dislike rating was detected from the physiological signals and video content (Soleymani et al. (2011)). Koelstra et al. (Koelstra et al. (2012)) used Electroencephalogram (EEG) and peripheral physiological signals for emotional tagging of music videos. In a similar study (Soleymani et al. (2012b)), a multimodal emotional tagging was conducted using EEG signals and pupillary reflex. Khomami Abadi et al. (Abadi et al. (2013b)) recorded and analyzed MagnetoEncephaloghram (MEG) signals as an alternative to the EEG signals with the ability to monitor brain activities. Although they could obtain comparable results to the results obtained by EEG, the price and apparatus complexity of MEG machines do not make it an apparent candidate for such applications.

In an approach taken for emotional tagging, emotional events, defined as arousing events, were first detected in movies from peripheral physiological responses and then their valence was detected using Gaussian processes classifiers (Fleureau et al. (2012)). Such a strategy can be also justified based on the heart-shaped distribution of emotions on arousal and valence plane (Dietz and Lang (1999)) in which emotions with higher arousal have more extreme pleasantness or unpleasantness.

Engagement of viewers with movie scenes was assessed by physiological signals and facial expressions (Abadi et al. (2013a)). Measuring engagement a system will be able to steer the story in a hyper-narrative movie where different outcomes all possible based on the users' reactions.

Spontaneous audio responses can be also used for implicit emotional tagging. Petridis and Pantic proposed a method for tagging videos for the level of hilarity by analyzing user's laughter (Petridis and Pantic (2009)). Different types of laughter can be an indicator of the level of hilarity of multimedia content. Using audiovisual modalities, they could recognize speech, unvoiced laughter, and voiced laughter.

1.2.2 Highlight and interest detection

Users' interest in content can help recommender systems, content producers and advertisers to better focus their efforts towards higher user satisfaction. Kurdyukova et al. (Kurdyukova et al. (2012)) setup a display that can detect the interest of the passersby by detecting their faces, facial expressions and head pose. In addition, the social context, groups, conversations and gender were recognized which can be used for profiling purposes for advertisements. In a study on estimating movie ratings (Silveira et al. (2013)), Galvanic Skin Response (GSR) were recorded and analyzed from a group of movie audience. Movie ratings on five-point scale were classified into low rating (1-3) and high rating (4-5) classes. Their method could achieve better results incorporating GSR responses along demography information for two out of three studied movies.

Interest in the advertisements was shown to be detectable by analyzing the facial expressions on viewers on the web. McDuff et al. (McDuff et al. (2012, 2013)) measured the level of smile from the video advertisement audience to assess their interest in the content. They collected a large number of samples using crowdsourcing by recording the responses on users' webcams. Ultimately, they were able to detect fairly accurately the desire to watch the video again and whether the viewers liked the videos.

Video highlight detection and summarization is an important application for indexing and visualization purposes. Joho et al. (Joho et al. (2010, 2009)) developed a video summarization tool using facial expressions. A probabilistic emotion recognition based on facial expressions was employed to detect emotions of 10 participants watching eight video clips. The expression change rate between different emotional expressions and the pronounce level of expressed emotions were used as features to detect personal highlights in the videos. The pronounce levels they used was ranging from highly expressive emotions, surprise and happiness, to no expression or neutral. Chênes et al (Chênes et al. (2012)) used physiological linkage between different viewers to detect video highlights. Skin temperature and Galvanic Skin Response (GSR) were found to be informative in detecting video highlights via physiological linkage. They achieved 78.2% of accuracy in detecting highlight by their proposed method. In a more recent study, Fleureau et al. (Fleureau et al. (2013)) used GSR responses of a group of audience simultaneously to create an emotional profile of movies. The profiles generated from the physiological reposes were shown to match the user reported highlight.

1.2.3 Relevance assessment

Users' responses also carry pertinent information regarding the relevance of retrieved content to a query. Relevance of content to the user generated tags or tags detected by content based indexing systems can be also assessed by users' responses Soleymani et al. (2013). Arapakis et al. (Arapakis et al. (2009a)) introduced a method to assess the topical relevance of videos in accordance to a given query using facial expressions showing users' satisfaction or dissatisfaction. Based on facial expressions recognition techniques, basic emotions were detected and compared with the ground truth. They were able to predict with 89% accuracy whether a video was indeed relevant to the query. The same authors later studies the feasibility of using affective responses derived from both facial expressions and physiological signals as implicit indicators of topical relevance. Although the results were above random level and support the feasibility of the approach, there is still room for improvement from the best obtained classification accuracy, 66%, on relevant versus non-relevant classification (Arapakis et al. (2009b)). In the same line Arapakis et al. (Arapakis et al. (2010)) compared the performance of personal versus general affect recognition approaches for topical relevance assessment and found that accounting for personal differences in their emotion recognition method improved their performance. In a more recent study, Moshfeghi and Jose (Moshfeghi and Jose (2013)) showed that physiological responses and facial expressions can be used as complementary source of information in addition to dwell time for relevance assessment. Their study was evaluated with an experiment on a video retrieval platform.

In another information retrieval application, Kelly and Jones (Kelly and Jones (2010)) used physiological responses to rerank the content collected via a lifelogging application. The lifelogging application collects picture, text messages, GSR, skin temperature and the energy that the body of a user consumed using an accelerometer. Using the skin temperature they could improve the Mean Average Precision (MAP) of baseline, retrieval system by 36%.

Koelstra et al. (Koelstra et al. (2009)) investigated the use of electroencephalogram (EEG) signals for implicit tagging of images and videos. They showed short video excerpts and images first without tags and then with a tag. They found significant differences in EEG signals (N400 evoked potential) between responses to relevant and irrelevant tags. These differences were nevertheless not always present; thus precluding classification. Facial expression and eye gaze were used to detect users' agreement or disagreement with the displayed tags on 28 images (Jiao and Pantic (2010); Soleymani et al. (2012a)). The results showed that not all the participants in the experiment were expressing their agreement or disagreement on their faces and their eye gaze were more informative for agreement assessment. Soleymani and Pantic, the authors of this chapter, showed that EEG (Soleymani and Pantic (2013)) signals and N400 while aggregated from multiple participants can reach a high accuracy for detecting the non-relevant content. Soleymani et al. (Soleymani et al. (2013)) further studied the effectiveness of different modalities for relevance assessment on the same dataset. They showed that in a user independent approach eye gaze performs much better than EEG signals and facial expressions to detect tag relevance. Eye gaze responses have been also used to detect interest for image annotation (Haji Mirza et al. (2012)), relevance judgment (Salojärvi et al. (2005)), interactive video search (Vrochidis et al. (2011)), and search personalization (Buscher et al. (2009)).

1.3 Databases

In this section, we introduce the publicly available databases which are developed for the sole purpose of implicit human-centered tagging studies.

The MAHNOB HCI database (Soleymani et al. (2012a)) is developed for experimenting implicit tagging approaches for two different scenarios, namely, emotional tagging, tag relevance assessment. This database consists of two experiments. The responses including, EEG, physiological signals, eye gaze, audio and facial expressions of 30 people were recorded. The first experiment was watching 20 emotional video extracted from movies and online repositories. The second experiment was tag agreement experiment in which images and short videos with human actions were shown the participants first without a tag and then with a displayed tag. The tags were either correct or incorrect and participants' agreement with the displayed tag was assessed. An example of an eye gaze pattern and fixations points on an image with an overlaid label is shown in Fig. 1.2. This database is publicly available on the Internet[1].

Figure 1.2 An example of displayed images is shown with eye gaze fixation and scan path overlaid. The size of the circles represents the time spent staring at each fixation point.

A Database for Emotion Analysis using Physiological Signals (DEAP) (Koelstra et al. (2012)) is a database developed for emotional tagging of music videos. It includes peripheral and central nervous system physiological signals in addition to face videos from 32 participants. The face videos were only recorded from 22 participants. EEG signals were recorded from 32 active electrodes. Peripheral nervous system physiological signals were EMG, electroocologram (EOG), blood volume pulse (BVP) using plethysmograph, skin temperature, and GSR. The spontaneous reactions of participants were recorded in response to music video clips. This database is publicly available on the Internet[2].

[1] http://mahnob-db.eu/hct-tagging/
[2] http://www.eecs.qmul.ac.uk/mmv/datasets/deap/

The Pinview database comprises of eye gaze and interaction data collected in an image retrieval scenario (Auer et al. (2010)). The Pinview database includes explicit relevance feedback interaction from the user, such as pointer clicks and implicit relevance feedback signals, such as eye movements and pointer traces. This database is available online[3].

Tkalčič et al. collected the LDOS-PerAff-1 corpus of face video clips in addition to the participants personality (Tkalčič et al. (2010a)). Participants personalities were assessed by International Personality Item Pool (IPIP) questionnaire (Goldberg et al. (2006)). Participants watched a subset of images extracted from International Affective Picture system (IAPS) (Lang et al. (2005)) and on a five points likert scale rated their preference for choosing the picture for their desktop wallpaper. The LDOS-PerAff-1 database is available online[4].

1.4 Challenges and Perspectives

Reading users' minds and generating the ground-truth for emotion and interest detection is one of the main challenges of implicit tagging studies. It is often easier for the users to compare or rank the content based on their emotion rather than assigning an exact label or absolute ratings (Yannakakis and Hallam (2011)). Although comparing pair or a group of content to each other require a larger number of trials and longer experiments it should be taken into account in future studies.

The other challenge is to have non-intrusive, easy to use and cheap sensors that can be commercially produced. Thanks to the growing interest from the industry portable and wearable sensors and camera are becoming cheaper and more accessible, e.g., Microsoft Kinect, Google glass. In addition to the sensor based methods, there is also a trend in detecting physiological signals and facial expressions through users' webcams (McDuff et al. (2013)). Due to the availability of webcams on almost all the devices, there is a huge potential for its use.

Emotional expressions in natural settings are mostly subtle and person dependent which make them hard to detect. Therefore, large databases and specific machine learning techniques still have to be developed for bringing implicit tagging ideas into practice. So far, the emotional models are limited either to the discrete basic emotions or the dimensional valence-arousal-dominance spaces. Developing new emotional models and

[3] http://www.pinview.eu/databases/
[4] http://slavnik.fe.uni-lj.si/markot/Main/LDOS-PerAff-1

dimensions specific to different applications, such as the one proposed by Eggink and Bland (Eggink and Bland (2012)) and Benini et al. (Benini et al. (2011)), should be also explored.

There are also contextual factors such as time, environment, cultural background, mood and personality which are not necessarily easy to assess or consider (Soleymani et al. (2014)). The important contextual factors for each application need to be carefully identified and their effect has to be incorporated into the final tagging or retrieval process.

Some people might also find such systems intrusive, and they have legitimate privacy concerns. For example, such technologies can be used for surveillance and marketing purposes without users' consent. These concerns need to be addressed by researchers in collaborations with ethics and law experts.

Implicit tagging is showing its potentials by attracting interest from the industrial entities. The proliferation of commercially produced sensors, such as handheld devices, equipped with RGB-D cameras, will help the emergence of the new techniques for multimedia implicit tagging.

References

Abadi, Mojtaba Khomami, Staiano, Jacopo, Cappelletti, Alessandro, Zancanaro, Massimo, and Sebe, Nicu. 2013a (September). Multimodal Engagement Classification for Affective Cinema. Pages 411–416 of: *Affective Computing and Intelligent Interaction and Workshops, 2013. ACII 2013. 3rd International Conference on.*

Abadi, Mojtaba Khomami, Kia, Seyed Mostafa, Subramanian, Ramanathan, Avesani, Paolo, and Sebe, Nicu. 2013b (September). User-centric Affective Video Tagging from MEG and Peripheral Physiological Responses. Pages 582–587 of: *Affective Computing and Intelligent Interaction and Workshops, 2013. ACII 2013. 3rd International Conference on.*

Arapakis, I., Moshfeghi, Y., Joho, H., Ren, R., Hannah, D., and Jose, J. M. 2009a (July). Integrating facial expressions into user profiling for the improvement of a multimodal recommender system. Pages 1440–1443 of: *Multimedia and Expo, 2009. ICME 2009. IEEE International Conference on.*

Arapakis, Ioannis, Konstas, Ioannis, and Jose, Joemon M. 2009b. Using facial expressions and peripheral physiological signals as implicit indicators of topical relevance. Pages 461–470 of: *Proceedings of the seventeen ACM international conference on Multimedia.* MM '09. New York, NY, USA: ACM.

Arapakis, Ioannis, Athanasakos, Konstantinos, and Jose, Joemon M. 2010. A comparison of general vs personalised affective models for the prediction of topical relevance. Pages 371–378 of: *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval.* SIGIR '10. New York, NY, USA: ACM.

Auer, P., Hussain, Z., Kaski, S., Klami, A., Kujala, J., Laaksonen, J., Leung, A., Pasupa, K., and Shawe-Taylor, J. 2010. Pinview: Implicit feedback in content-based image retrieval. Pages 51–57 of: *JMLR: Workshop on Applications of Pattern Analysis.*

Benini, S., Canini, L., and Leonardi, R. 2011. A Connotative Space for Supporting Movie Affective Recommendation. *IEEE Trans. Multimedia*, **13**(6), 1356–1370.

Buscher, Georg, van Elst, Ludger, and Dengel, Andreas. 2009. Segment-level display time as implicit feedback: a comparison to eye tracking. Pages 67–74 of: *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*. SIGIR '09. New York, NY, USA: ACM.

Chênes, Christophe, Chanel, Guillaume, Soleymani, Mohammad, and Pun, Thierry. 2012. Highlights detection in movie scenes through inter-users physiological linkage. In: Ramazan, Naeem, van Zwol, Roelof, Lee, Jong-Seok, Cluver, Kai, and Hua, Xian-Sheng (eds), *Social Media Retrieval*. Computer Communications and Networks Series. Berlin, Heidelberg: Springer. in press.

Dietz, Richard B., and Lang, Annie. 1999. Aefective agents: Effects of agent affect on arousal, attention, liking and learning. In: *Cognitive Technology Conference*.

Eggink, Jana, and Bland, Denise. 2012. A large scale experiment for mood-based classification of tv programmes. Pages 140–145 of: *Multimedia and Expo (ICME), 2012 IEEE International Conference on*. IEEE.

Fleureau, J., Guillotel, P., and Huynh-Thu, Q. 2012. Physiological-Based Affect Event Detector for Entertainment Video Applications. *Affective Computing, IEEE Transactions on*. in press.

Fleureau, Julien, Guillotel, Philippe, and Orlac, Izabela. 2013 (September). Affective Benchmarking of Movies Based on the Physiological Responses of a Real Audience. Pages 73–77 of: *Affective Computing and Intelligent Interaction and Workshops, 2013. ACII 2013. 3rd International Conference on*.

Goldberg, Lewis R., Johnson, John A., Eber, Herbert W., Hogan, Robert, Ashton, Michael C., Cloninger, C. Robert, and Gough, Harrison G. 2006. The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, **40**(1), 84–96.

Haji Mirza, S., Proulx, M., and Izquierdo, E. 2012. Reading Users' Minds from Their Eyes: A Method for Implicit Image Annotation. *Multimedia, IEEE Transactions on*. in press.

Hanjalic, A., and Xu, Li-Qun. 2005. Affective video content representation and modeling. *Multimedia, IEEE Transactions on*, **7**(1), 143–154.

Hardoon, David R., and Pasupa, Kitsuchart. 2010. Image ranking with implicit feedback from eye movements. Pages 291–298 of: *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*. ETRA '10. New York, NY, USA: ACM.

Jiao, J., and Pantic, M. 2010. Implicit image tagging via facial information. Pages 59–64 of: *Proceedings of the 2nd international workshop on Social signal processing*. ACM.

Joachims, Thorsten, Granka, Laura, Pan, Bing, Hembrooke, Helene, and Gay, Geri. 2005. Accurately interpreting clickthrough data as implicit feedback. Pages 154–161 of: *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*. SIGIR '05. New York, NY, USA: ACM.

Joho, Hideo, Jose, Joemon M., Valenti, Roberto, and Sebe, Nicu. 2009. Exploiting facial expressions for affective video summarisation. In: *Proceeding of the ACM International Conference on Image and Video Retrieval*. CIVR '09. New York, NY, USA: ACM.

Joho, Hideo, Staiano, Jacopo, Sebe, Nicu, and Jose, Joemon. 2010. Looking at the viewer: analysing facial activity to detect personal highlights of multimedia contents. *Multimedia Tools and Applications*, **51**(2), 505–523.

Kelly, Liadh, and Jones, Gareth. 2010. Biometric Response as a Source of Query Independent Scoring in Lifelog Retrieval. Pages 520–531 of: Gurrin, Cathal, He, Yulan, Kazai, Gabriella, Kruschwitz, Udo, Little, Suzanne, Roelleke, Thomas, Rger, Stefan, and van Rijsbergen, Keith (eds), *Advances in Information Retrieval*. Lecture Notes in Computer Science, vol. 5993. Springer Berlin / Heidelberg.

Kierkels, Joep J. M., Soleymani, Mohammad, and Pun, Thierry. 2009. Queries and tags in affect-based multimedia retrieval. Pages 1436–1439 of: *ICME'09: Proceedings of the 2009 IEEE international conference on Multimedia and Expo*. Piscataway, NJ, USA: IEEE Press.

Klinghult, Gunnar. 2012 (nov). *Camera button with integrated sensors*. US Patent App. 13/677,517.

Koelstra, S., Muhl, C., and Patras, I. 2009. EEG analysis for implicit tagging of video data. Pages 1–6 of: *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*. IEEE.

Koelstra, Sander, Yazdani, Ashkan, Soleymani, Mohammad, Mühl, Christian, Lee, Jong-Seok, Nijholt, Anton, Pun, Thierry, Ebrahimi, Touradj, and Patras, Ioannis. 2010. Single Trial Classification of EEG and Peripheral Physiological Signals for Recognition of Emotions Induced by Music Videos. Chap. 9, pages 89–100 of: Yao et al (ed), *Brain Informatics*. Lecture Notes in Computer Science, vol. 6334. Berlin, Heidelberg: Springer.

Koelstra, Sander, Muhl, Christian, Soleymani, Mohammad, Lee, Jong-Seok, Yazdani, Ashkan, Ebrahimi, Touradj, Pun, Thierry, Nijholt, Anton, and Patras, Ioannis (Yiannis). 2012. DEAP: A Database for Emotion Analysis Using Physiological Signals. *IEEE Transactions on Affective Computing*, **3**, 18–31.

Kurdyukova, Ekaterina, Hammer, Stephan, and Andr, Elisabeth. 2012. Personalization of Content on Public Displays Driven by the Recognition of Group Context. Pages 272–287 of: Patern, Fabio, Ruyter, Boris, Markopoulos, Panos, Santoro, Carmen, Loenen, Evert, and Luyten, Kris (eds), *Ambient Intelligence*. Lecture Notes in Computer Science, vol. 7683. Springer Berlin Heidelberg.

Lang, Pj, Bradley, Mm, and Cuthbert, Bn. 2005. *International affective picture system (IAPS): Affective ratings of pictures and instruction manual*. Tech. rept. A-8. University of Florida, Gainesville, Florida, US.

Larson, Martha, Soleymani, Mohammad, Serdyukov, Pavel, Rudinac, Stevan, Wartena, Christian, Murdock, Vanessa, Friedland, Gerald, Ordelman, Roeland, and Jones, Gareth J. F. 2011. Automatic tagging and geotagging in video collections and communities. Pages 51:1–51:8 of: *Proceedings of the 1st ACM International Conference on Multimedia Retrieval*. ICMR '11. New York, NY, USA: ACM.

McDuff, Daniel, El Kaliouby, Rana, and Picard, Rosalind W. 2012. Crowdsourcing Facial Responses to Online Videos. *Affective Computing, IEEE Transactions on*, **3**(4), 456–468.

McDuff, Daniel, el Kaliouby, Rana, Demirdjian, David, and Picard, Rosalind. 2013. Predicting online media effectiveness based on smile responses gathered over the Internet. Pages 1–7 of: *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*.

Moshfeghi, Yashar, and Jose, Joemon M. 2013. An effective implicit relevance feedback technique using affective, physiological and behavioural features. Pages 133–142 of: *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*. ACM.

Pantic, Maja, and Vinciarelli, Alessandro. 2009. Implicit Human-Centered Tagging. *IEEE Signal Processing Magazine*, **26**(6), 173–180.

Petridis, S., and Pantic, M. 2009. Is this joke really funny? judging the mirth by audiovisual laughter analysis. Pages 1444 –1447 of: *Multimedia and Expo, 2009. ICME 2009. IEEE International Conference on*.

Salojärvi, Jarkko, Puolamäki, Kai, and Kaski, Samuel. 2005. Implicit Relevance Feedback from Eye Movements. Pages 513–518 of: Duch, Wlodzislaw, Kacprzyk, Janusz, Oja, Erkki, and Zadrozny, Slawomir (eds), *Artificial Neural Networks: Biological Inspirations ICANN 2005*. Lecture Notes in Computer Science, vol. 3696. Springer Berlin / Heidelberg.

Shan, Man K., Kuo, Fang F., Chiang, Meng F., and Lee, Suh Y. 2009. Emotion-based music recommendation by affinity discovery from film music. *Expert Syst. Appl.*, **36**(4), 7666–7674.

Shen, Xuehua, Tan, Bin, and Zhai, ChengXiang. 2005. Context-sensitive information retrieval using implicit feedback. Pages 43–50 of: *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*. SIGIR '05. New York, NY, USA: ACM.

Silveira, Fernando, Eriksson, Brian, Sheth, Anmol, and Sheppard, Adam. 2013. Predicting Audience Responses to Movie Content from Electro-Dermal Activity Signals. In: *Proceedings of the 2013 ACM Conference on Ubiquitous Computing*. ACM. in press.

Soleymani, M., and Pantic, M. 2012. Human-centered implicit tagging: Overview and perspectives. Pages 3304–3309 of: *IEEE SMC*.

Soleymani, M., and Pantic, M. 2013. Multimedia Implicit Tagging using EEG Signals. In: *Multimedia and Expo, 2013. ICME 2013. IEEE International Conference on,*IEEE, for IEEE.

Soleymani, M., Koelstra, S., Patras, I., and Pun, T. 2011 (march). Continuous emotion detection in response to music videos. Pages 803–808 of: *Automatic Face Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on.*

Soleymani, M., Kaltwang, S., and Pantic, M. 2013. Human Behavior Sensing for Tag Relevance Assessment. In: *Proceedings of the 21st ACM international conference on Multimedia.* MMACM, for ACM.

Soleymani, Mohammad, Chanel, Guillaume, Kierkels, Joep J. M., and Pun, Thierry. 2009. Affective Characterization of Movie Scenes Based on Content Analysis and Physiological Changes. *International Journal of Semantic Computing*, **3**(2), 235–254.

Soleymani, Mohammad, Lichtenauer, Jeroen, Pun, Thierry, and Pantic, Maja. 2012a. A Multimodal Database for Affect Recognition and Implicit Tagging. *IEEE Transactions on Affective Computing*, **3**, 42–55.

Soleymani, Mohammad, Pantic, Maja, and Pun, Thierry. 2012b. Multimodal Emotion Recognition in Response to Videos. *IEEE Transactions on Affective Computing.* in press.

Soleymani, Mohammad, Larson, Martha, Pun, Thierry, and Hanjalic, Alan. 2014. Corpus development for affective video indexing. *IEEE Transactions on Multimedia*, **16**(4), 1075–1089.

Tkalčič, Marco, Odic, Ante, Košir, Andrej, and Tasic, Jurij. 2013. Affective labeling in a content-based recommender system for images. *IEEE Transactions on Multimedia*, **15**(2), 391–400.

Tkalčič, Marko, Tasič, Jurij, and Košir, Andrej. 2010a. The LDOS-PerAff-1 Corpus of Face Video Clips with Affective and Personality Metadata. In: *Proceedings of Multimodal Corpora Advances in Capturing Coding and Analyzing Multimodality, LREC.*

Tkalčič, Marko, Burnik, Urban, and Košir, Andrej. 2010b. Using affective parameters in a content-based recommender system for images. *User Modeling and User-Adapted Interaction*, **20**(4), 279–311.

Vrochidis, Stefanos, Patras, Ioannis, and Kompatsiaris, Ioannis. 2011. An eye-tracking-based approach to facilitate interactive video search. Pages 43:1–43:8 of: *Proceedings of the 1st ACM International Conference on Multimedia Retrieval.* ICMR '11. New York, NY, USA: ACM.

Yannakakis, Georgios N., and Hallam, John. 2011. Ranking vs. Preference: A Comparative Study of Self-reporting. Pages 437–446 of: DMello, Sidney, Graesser, Arthur, Schuller, Bjrn, and Martin, Jean-Claude (eds), *Affective Computing and Intelligent Interaction.* Lecture Notes in Computer Science, vol. 6974. Springer Berlin Heidelberg.