2012 IEEE International Conference on Systems, Man, and Cybernetics
October 14-17, 2012, COEX, Seoul, Korea

Human-Centered Implicit Tagging: Overview and Perspectives

Mohammad Soleymani
Department of Computing
Imperial College London
London SW7 2AZ, UK
m.soleymani@imperial.ac.uk

Maja Pantic
Department of Computing
Imperial College London
Faculty of EEMCS
University of Twente, the Netherlands
m.pantic@imperial.ac.uk

***Abstract*—Tags are an effective form of metadata which help users to locate and browse multimedia content of interest. Tags can be generated by users (user-generated explicit tags), automatically from the content (content-based tags), or assigned automatically based on non-verbal behavioral reactions of users to multimedia content (implicit human-centered tags). This paper discusses the definition and applications of implicit human-centered tagging. Implicit tagging is an effortless process by which content is tagged based on users' spontaneous reactions. It is a novel but growing research topic which is attracting more attention with the growing availability of built-in sensors. This paper discusses the state of the art in this novel field of research and provides an overview of publicly available relevant databases and annotation tools. We finally discuss in detail challenges and opportunities in the field.**

***Index Terms*—tagging, implicit tagging, emotion recognition, multimedia indexing.**

I. Introduction

Information management systems use tags as an effective form of metadata to support users in finding and re-finding multimedia content of interest. Tags can come in different form including semantic tags and geotags [1]. In contrast to classic tagging schemes where users direct input is mandatory, Implicit Human-Centered Tagging (IHCT) was proposed [2] to gather tags and annotations without any effort from users. The main idea behind IHCT is that nonverbal behaviors displayed when interacting with multimedia data (e.g., facial expressions, head nods, eye gaze, physiological responses, etc) provide information useful for improving the tag sets associated with the data. The resulting tags are called "implicit" since there is no need for users' direct input as reactions to multimedia are displayed spontaneously. Currently, social media websites encourage users to tag the multimedia content. However, the users' intent when tagging multimedia content does not always match the information retrieval goals. A large portion of user-defined tags are either motivated by the goal of increasing the popularity and reputation of a user in an online community or based on individual judgments and goals [2]. For example, a user might tag content to increase the popularity and visibility of himself or his content. In contrast to the standard "explicit" tagging, implicit tagging does not prompt the users for tags while they listen to or watch a multimedia content. Moreover, if implicit tagging is done reliably resulting tags carry less irrelevant and inaccurate information compared to the case with "explicit" tagging. Tags obtained through IHCT are expected to be more robust than tags associated with the data explicitly, at least in terms of: generality (they make sense to everybody) and statistical reliability (all tags will be sufficiently represented). A scheme of implicit tagging versus explicit scenario versus explicit tagging is shown in Fig. 1.

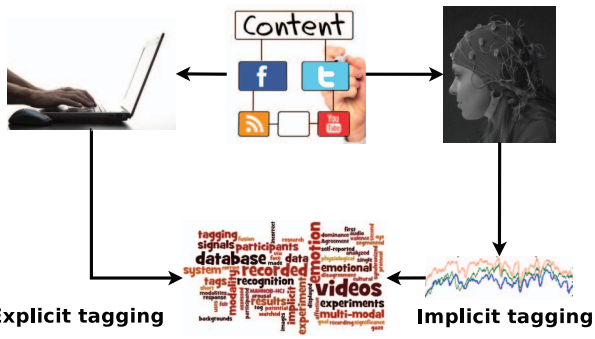

Fig. 1. Implicit tagging vs. explicit tagging scenarios. The analysis of the bodily reactions to multimedia content replace the direct interaction between user and the computer. Therefore, user do not have to put any effort into tagging the content.

The users' behavior and spontaneous reactions to multimedia data can provide useful information for multimedia indexing with the following scenarios: (i) direct assessment of tags: users spontaneous reactions will be translated into emotional keywords, *e.g.*, funny, disgusting, scary [3], [4], [5], [6]; (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 [7], [8], [9], [10]; (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 [11], [12].

Multimedia indexing has focused on generating characterizations of content in terms of events, objects, etc. The judgment relies on cognitive processing combined with general world knowledge and is considered to be objective due to its reproducibility by users with a wide variety of backgrounds. Parallel to this approach to indexing, an alternative has also emerged that also take affective aspects into account. Here,

978-1-4673-1714-6/12/$31.00 ©2012 IEEE

affect refers to the intensity and type of emotion that is evoked in a user while watching/listening to multimedia content [13]. Affective characteristics of multimedia are important features for describing multimedia content and can be presented by relevant emotional tags. Implicit tagging, dealing directly with users' reactions, can be used directly to find affective tags for a given content. These tags can, in return, help recommendation and retrieval systems to improve their performance [14], [15], [3]. Challenges and difficulties in using self reported emotions [16] make implicit tagging a suitable alternative for recognizing emotional tags.

Users do not evaluate media content on the same emotional criteria for affective tagging. Some might tag multimedia content with words to express their emotion while others might use tags to describe the content. For example, a picture receive different tags based on the objects in the image, the camera by which the picture was taken or the emotion a user felt looking at the picture. Scherer defines this by intrinsic and extrinsic appraisal [17]. Intrinsic appraisal is independent from the current goals and values of the viewer while extrinsic or transactional appraisal leads to feeling emotions in response to the stimuli. For example, the content's intrinsic emotion of a picture with a smiling face is happiness whereas this person might be a hatred figure to the viewer and the extrinsic appraisal leads to unpleasant emotions. This classification between intrinsic and extrinsic emotions should be taken into account while dealing with affective tags. The emotional tags that can be generated by implicit tagging are the extrinsic emotional tags.

Other feedbacks from users, including clickthrough data have been used extensively for information retrieval and topic relevance applications [18], [19]. In this paper, we only cover the implicit feedbacks which are measurable with sensors and cameras from bodily responses. The rest of the paper is organized as follows. Section II provides a background on the recent developments in this relatively young topic. Available resources including tools and databases are introduced in Section III. Current challenges and perspectives are discussed in Section IV.

II. State of the art

Pantic and Vinciarelli define implicit tagging as using non-verbal spontaneous behavior to find relevant keyword or tags for multimedia content [2]. Implicit tagging research has recently attracted researchers' attention, and number of studies have been published [20], [11], [21]. Implicit tagging has been used in image annotation, video highlight detection, topical relevance detection and retrieval result re-ranking. The existing literature can be divided into two categories, one dealing with using emotional reactions to tag the content with the expressed emotion, e.g., laughter detection for hilarity [5], and the second group of studies using the spontaneous reactions for information retrieval or search results, e.g., eye gaze for relevance feedback [22]. A summary of the recent relevant literature on this topic is given in Table I.

There have been also studies using unnimodal or multimodal approaches for detecting the behavioral or emotional responses to multimedia [21], [23], [24], [20], [6]. There is currently a research trend towards estimating emotions from multimedia content automatically [13], [15], [14]. The emotion recognition has been also used in applications such as detecting topical relevance, or summarizing videos [11], [21], [9].

Emotional characteristics of videos have improved music and image recommendation. Shan et al. [14] used affective characterization using content analysis to improve film music recommendation. Tkalčič et al. showed how affective information can improve image recommendation [15]. In their image recommendation scenario, affective scores of images from the international affective picture system (IAPS) [35] were used as features for an image recommender. They conducted an experiment with 52 participants to study the effect of using affective scores. The image recommender using affective scores showed a significant improvement in the performance of their image recommendation system.

An affective characterization for movie scenes using peripheral physiological signals was proposed by Soleymani et al. [20]. 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. A similar approach was taken using a linear ridge regression for emotional characterization of music videos [31]. Arousal, valence, dominance, and like/dislike rating was detected from the physiological signals and video content.

Kierkels et al. [3] proposed a method for personalized affective tagging of multimedia using peripheral physiological signals. Valence and arousal levels of participants' emotion when watching videos were computed from physiological responses using linear regression [20]. 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. So far this novel method achieved low precision.

Koelstra et al. [6] recorded EEG and peripheral physiological signals of six participants in response to music videos. Participants rated their felt emotions by means of arousal, valence and like/dislike rating rating. The emotional responses of each participant was classified into two classes of low/high arousal, low/high like/dislike, and low/high valence. The average classification rates varied between 55% and 58% which is slightly above random level.

Joho et al. [11], [21] 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.

Chêne et al [27] used physiological linkage between differ-

TABLE I
THE SUMMARY OF IMPLICIT TAGGING LITERATURE IN DIFFERENT SCENARIOS. SENSORS ACRONYMS: GALVANIC SKIN RESPONSE (GSR), SKIN TEMPERATURE (TEMP), RESPIRATION AMPLITUDE (RESP.), ELECTROMIOGRAM (EMG), NEAR BODY AMBIENT TEMPERATURE (NB-TEMP.), HEAT FLUX (HF), ACCELEROMETER (ACC.), CAMERA (CAM.), ELECTROENCEPHALOGRAM (EEG), EYE GAZE TRACKER (EGT), MICROPHONE (MICRO.), BLOOD VOLUME PULSE (BVP) WITH PLETHYSMOGRAPHY, MEAN AVERAGE PRECISION (MAP), CONTINUOUS (CONT.)

Study	Sensor	Modality	Application/Method	Content	# classes	Best result
Arapakis et al. [9]	Cam. & GSR, Temp., NB-Temp., HF, Acc.	visual & physiological signals	topic relevance assessment, facial expression analysis	video search results	2	66.5%
Arapakis et al. [25]	Cam.	visual	topic relevance assessment, facial expression analysis	search results	2	72.5%
Buscher et al. [26]	EGT	eye gaze and scrolling	implicit feedback, search personalization	search results	2	MAP=0.83
Chêne et al. [27]	GSR, Temp., EMG, Resp., BVP	physiological signals	video summarization with physiological linkage	video	2	78.2%
Fleureau et al. [27]	GSR, EMG, , BVP	physiological signals	video emotional event detection	video	2	86.1%
Haji Mirza et al. [28]	EGT	eye gaze	relevance judgment by attention assessment	images	2	recall=0.53
Jiao & Pantic [10], [8]	Cam. & EGT	visual and eye gaze	image tagging with agreement assessment	image	2	72.1%
Joho et al. [11], [21]	Cam.	visual	video summarization by emotion detection	video	2	MAP=0.4
Kelly and Jones [29]	GSR, Temp., Acc.	physiological signals	retrieval reranking, arousal assessment	search results	2	MAP improvement 0.35
Kierkels et al. [3]	GSR, Temp., EMG, Resp., BVP	physiological signals	emotional tagging, emotion detection	video	2	-
Koelstra et al. [6]	GSR, Temp., EMG, Resp., BVP, EEG	physiological signals	emotional tagging, emotion detection	video	2	85.5%
Koelstra et al. [7]	EEG	EEG	video tagging, agreement assessment	video	2	-
Petridis & Pantic [5]	Cam, Micro.	audiovisual	hilarity detection, laughter detection	video	3	74.7%
Salojärvi et al. [30]	EGT	eye gaze	relevance judgment, attention assessment	search results	2	65.8%
Soleymani et al. [20]	GSR, Temp., EMG, Resp., BVP	physiological signals	emotional tagging, emotion detection	video	cont.	-
Soleymani et al. [31]	GSR, Temp., EMG, Resp., BVP, EEG	physiological signals	emotional tagging, emotion detection	video	cont.	-
Soleymani et al. [4]	EEG & EGT	EEG, pupil	emotional tagging, multimodal emotion detection	video	3	76.4%
Tkalčič et al. [32]	Cam.	visual	image tagging, emotion detection	image	2	F1=0.59
Vrochidis et al. [33], [34]	EGT	eye gaze	relevance judgment, attention assessment	video	2	95.1%

ent viewers to detect video highlights. Skin temperature and Galvanic Skin Response (GSR) were found to be informative in detecting video highlights via physiological linkage. The achieved 78.2% of accuracy in detecting highlight by their proposed method.

Petridis and Pantic proposed a method for tagging videos for the level of hilarity by analyzing user's laughter [5]. 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 with the accuracy of 74.7%.

Koelstra et al. 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 [7]. These differences were nevertheless not always present; thus precluding classification.

Arapakis et al. [36] 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. In a more recent study, the feasibility of using affective responses derived from both facial expressions and physiological signals as implicit indicators of topical relevance was investigated. Although the results are 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 [9]. In the same line Arapakis et al. 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 [25].

In another information retrieval application, Kelly and Jones [29] 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%.

Facial expression and eye gaze were used to detect users' agreement or disagreement with the displayed tags on 28 images [10], [8]. 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. Eye gaze responses have been also used to detect interest for image annotation [28], relevance judgment [30], interactive video search [34], and search personalization [26].

III. Resources

In this section, we discuss affective representation and techniques for gathering self-reported annotations on affective reactions.

A. Emotional representation and self reporting

Building the ground truth has been always a major challenge for emotion detection studies. Emotional self-reporting provides users' feedback on their felt emotions, and is an important part in emotion recognition studies which are the essential components of IHCT processes. There are different emotional representations including, discrete, continuous and component process model. Discrete emotions theories are inspired by Darwin and support the idea of the existence of the certain number of basic and universal emotions [17], [37]. Multiple emotional self-reporting methods have been created and used so far [38], [39], [17], [40], [41]. However, none of them give a generalized, simple and accurate mean for emotional self-reporting.

Emotional self-reporting can be done either in free-response or forced-choice formats. In the free-response format, the participants are free to express their emotions by words. In the forced-choice, participants are asked to answer specific questions and indicate their emotion. Forced-choice self-reports on affective experiments use either discrete or dimensional approaches. Based on discrete emotions, self-reporting tools were developed in which users are asked to report their emotions with emotional words on nominal, and ordinal scales. Dimensional approaches of emotional self-reporting are based on bipolar dimensions of emotions. Emotions can be reported on every dimension using ordinal or continuous scales [40].

Russell [42] introduced the circumplex model of affects for emotion representation. The advantage of this circumplex over either discrete or dimensional models is that all the emotions can be mapped on the circumplex only with the angle. Therefore, all emotions are presented on a circular and one dimensional model. Self Assessment Manikins (SAM) is one of the most famous emotional self-reporting tools. It consists of manikins expressing emotions. The emotions are varying on three different dimensions; namely, arousal, valence, and dominance [40].

Scherer [17] positioned 16 emotions around a circle to combine both dimensional and discrete emotional approaches to create the Geneva emotion wheel. For each emotion around the wheel five circles with increasing size from the center to the sides are displayed. The size of the circle is an indicator of the intensity of felt emotion (see Fig. 2). In an experiment, a participant can pick up to two emotions which were the closest to his/her experience from 20 emotions and report their intensities with the size of the marked circles. No emotion or other emotion can be indicated in the center. The emotions are sorted on the circle in a way to have, high control emotions on the top and low control emotions in the bottom whereas the horizontal axis which is not visible on the wheel represent valence or pleasantness.

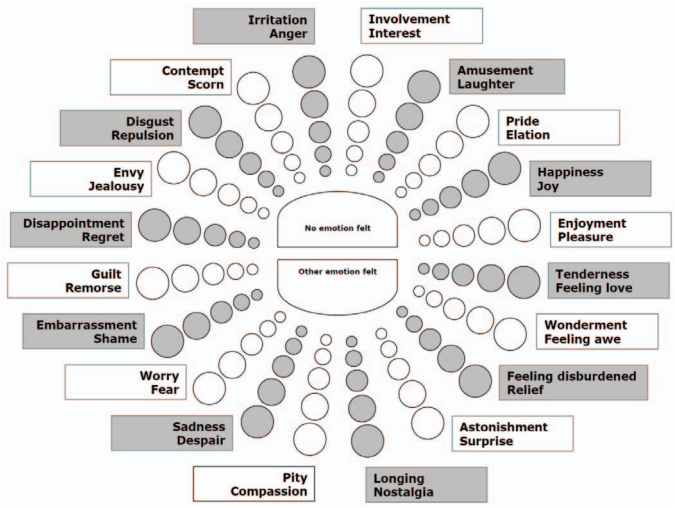

Fig. 2. A participant can indicate his emotion on Geneva emotion wheel by clicking or choosing the circles.

B. 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 [8] 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 a diplayes label is shown in Fig. 3. This database is publicly available on the Internet[1].

[1]http://mahnob-db.eu/hct-tagging/

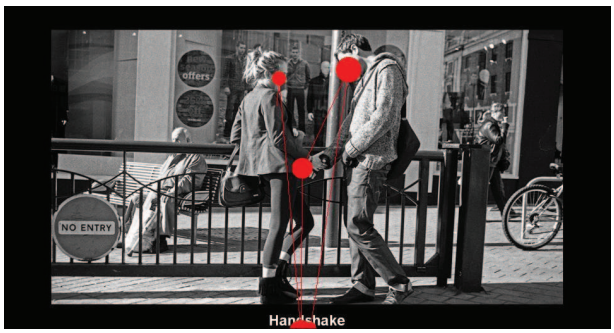

Fig. 3. An example of displayed images is shown with eye gaze fixation and scan path overlaid. The size of the circles represents the time spent staring at each fixation point.

A Database for Emotion Analysis using Physiological Signals (DEAP) [43] is a recent database that 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].

The Pinview database comprises of eye gaze and interaction data collected in an image retrieval scenario [44]. The Pinview databases includes explicit relevance feedback interaction from the user, such as pointer clicks and implicit relevance feedback signals, such as eye movements and pointer traces. These databases are available online[3].

Tkalčič et al. collected the LDOS-PerAff-1 corpus of face video clips in addition to the participants personality [45]. Participants personalities were assessed by International Personality Item Pool (IPIP) questionnaire [46]. Participants watched a subset of images extracted from International Affective Picture system (IAPS) [35] 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].

IV. Challenges and Perspectives

With the growing interest in commercially produced sensors and cameras, e.g., Microsoft Kinect, implicit tagging and interactive multimedia content delivery systems are going to emerge. However, the research in incorporating spontaneous reactions of viewers or listeners is still in its early stage. One of the main challenges of such studies is to create a ground truth by looking into the users mind. Therefore, large annotated dataset are a key development which should be followed by researchers. There are also contextual factors such as time, environment, cultural background, mood and personality which are not necessarily easy to assess or consider. 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.

Despite its challenges, we believe that applications related to entertainment and future media will recognize the value of implicit tagging and will deploy it as one of their core components. The following challenges can be identified as open issues that will need to be addressed.

Emotional and spontaneous reactions can vary from person to person. One key challenge will be to build machine learning techniques which can automatically consider these differences to improve the reliability of emotion recognition components.

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.

Existing sensors, e.g. Microsoft Kinect, Affectiva Q-sensor should be further developed or adapted to such applications. Not everybody is comfortable with the idea of wearing a wristband in order to sense their bodily changes. Thus, less intrusive sensors should be considered and their technologies should be further developed.

[2]http://www.eecs.qmul.ac.uk/mmv/datasets/deap/

[3]http://www.pinview.eu/databases/

[4]http://slavnik.fe.uni-lj.si/markot/Main/LDOS-PerAff-1

Acknowledgment

This work of Soleymani is supported by the European Research Council under the FP7 Marie Curie Intra-European Fellowship: Emotional continuous tagging using spontaneous behavior (EmoTag). The work of Pantic 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).

References

[1] M. Larson, M. Soleymani, P. Serdyukov, S. Rudinac, C. Wartena, V. Murdock, G. Friedland, R. Ordelman, and G. J. F. Jones, "Automatic tagging and geotagging in video collections and communities," in *Proceedings of the 1st ACM International Conference on Multimedia Retrieval*, ser. ICMR '11. New York, NY, USA: ACM, 2011, pp. 51:1–51:8.

[2] M. Pantic and A. Vinciarelli, "Implicit human-centered tagging," *IEEE Signal Processing Magazine*, vol. 26, no. 6, pp. 173–180, November 2009.

[3] J. J. M. Kierkels, M. Soleymani, and T. Pun, "Queries and tags in affect-based multimedia retrieval," in *ICME'09: Proceedings of the 2009 IEEE international conference on Multimedia and Expo*. Piscataway, NJ, USA: IEEE Press, 2009, pp. 1436–1439.

[4] M. Soleymani, M. Pantic, and T. Pun, "Multimodal emotion recognition in response to videos," *IEEE Transactions on Affective Computing*, 2012, in press.

[5] S. Petridis and M. Pantic, "Is this joke really funny? judging the mirth by audiovisual laughter analysis," in *Multimedia and Expo, 2009. ICME 2009. IEEE International Conference on*, 2009, pp. 1444 –1447.

[6] S. Koelstra, A. Yazdani, M. Soleymani, C. Mühl, J.-S. Lee, A. Nijholt, T. Pun, T. Ebrahimi, and I. Patras, "Single Trial Classification of EEG and Peripheral Physiological Signals for Recognition of Emotions Induced by Music Videos," in *Brain Informatics*, ser. Lecture Notes in Computer Science, Yao et al, Ed. Berlin, Heidelberg: Springer, 2010, vol. 6334, ch. 9, pp. 89–100.

[7] S. Koelstra, C. Muhl, and I. Patras, "Eeg analysis for implicit tagging of video data," in *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*. IEEE, 2009, pp. 1–6.

[8] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE Transactions on Affective Computing*, vol. 3, pp. 42–55, 2012.

[9] I. Arapakis, I. Konstas, and J. M. Jose, "Using facial expressions and peripheral physiological signals as implicit indicators of topical relevance," in *Proceedings of the seventeen ACM international conference on Multimedia*, ser. MM '09. New York, NY, USA: ACM, 2009, pp. 461–470.

[10] J. Jiao and M. Pantic, "Implicit image tagging via facial information," in *Proceedings of the 2nd international workshop on Social signal processing*. ACM, 2010, pp. 59–64.

[11] H. Joho, J. Staiano, N. Sebe, and J. Jose, "Looking at the viewer: analysing facial activity to detect personal highlights of multimedia contents," *Multimedia Tools and Applications*, vol. 51, no. 2, pp. 505–523, October 2010.

[12] C. Chênes, G. Chanel, M. Soleymani, and T. Pun, "Highlights detection in movie scenes through inter-users physiological linkage," in *Social Media Retrieval*, ser. Computer Communications and Networks Series, N. Ramazan, R. van Zwol, J.-S. Lee, K. Cluver, and X.-S. Hua, Eds. Berlin, Heidelberg: Springer, 2012, in press.

[13] A. Hanjalic and L.-Q. Xu, "Affective video content representation and modeling," *Multimedia, IEEE Transactions on*, vol. 7, no. 1, pp. 143–154, 2005.

[14] M. K. Shan, F. F. Kuo, M. F. Chiang, and S. Y. Lee, "Emotion-based music recommendation by affinity discovery from film music," *Expert Syst. Appl.*, vol. 36, no. 4, pp. 7666–7674, September 2009.

[15] M. Tkalčič, U. Burnik, and A. Košir, "Using affective parameters in a content-based recommender system for images," *User Modeling and User-Adapted Interaction*, vol. 20, no. 4, pp. 279–311, September 2010.

[16] R. W. Picard and S. B. Daily, "Evaluating Affective Interactions: Alternatives to Asking What Users Feel," in *CHI Workshop on Evaluating Affective Interfaces: Innovative Approaches*, 2005.

[17] K. R. Scherer, "What are emotions? And how can they be measured?" *Social Science Information*, vol. 44, no. 4, pp. 695–729, December 2005.

[18] X. Shen, B. Tan, and C. Zhai, "Context-sensitive information retrieval using implicit feedback," in *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, ser. SIGIR '05. New York, NY, USA: ACM, 2005, pp. 43–50.

[19] T. Joachims, L. Granka, B. Pan, H. Hembrooke, and G. Gay, "Accurately interpreting clickthrough data as implicit feedback," in *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, ser. SIGIR '05. New York, NY, USA: ACM, 2005, pp. 154–161.

[20] M. Soleymani, G. Chanel, J. J. M. Kierkels, and T. Pun, "Affective Characterization of Movie Scenes Based on Content Analysis and Physiological Changes," *International Journal of Semantic Computing*, vol. 3, no. 2, pp. 235–254, June 2009.

[21] H. Joho, J. M. Jose, R. Valenti, and N. Sebe, "Exploiting facial expressions for affective video summarisation," in *Proceeding of the ACM International Conference on Image and Video Retrieval*, ser. CIVR '09. New York, NY, USA: ACM, 2009.

[22] D. R. Hardoon and K. Pasupa, "Image ranking with implicit feedback from eye movements," in *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, ser. ETRA '10. New York, NY, USA: ACM, 2010, pp. 291–298.

[23] C. L. Lisetti and F. Nasoz, "Using noninvasive wearable computers to recognize human emotions from physiological signals," *EURASIP J. Appl. Signal Process.*, vol. 2004, no. 1, pp. 1672–1687, January 2004.

[24] K. Takahashi, "Remarks on Emotion Recognition from BioPotential Signals," in *in 2nd Int. Conf. on Autonomous Robots and Agents, 2004*, 2005.

[25] I. Arapakis, K. Athanasakos, and J. M. Jose, "A comparison of general vs personalised affective models for the prediction of topical relevance," in *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, ser. SIGIR '10. New York, NY, USA: ACM, 2010, pp. 371–378.

[26] G. Buscher, L. van Elst, and A. Dengel, "Segment-level display time as implicit feedback: a comparison to eye tracking," in *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, ser. SIGIR '09. New York, NY, USA: ACM, 2009, pp. 67–74.

[27] J. Fleureau, P. Guillotel, and Q. Huynh-Thu, "Physiological-based affect event detector for entertainment video applications," *Affective Computing, IEEE Transactions on*, 2012, in press.

[28] S. Haji Mirza, M. Proulx, and E. Izquierdo, "Reading users' minds from their eyes: A method for implicit image annotation," *Multimedia, IEEE Transactions on*, 2012, in press.

[29] L. Kelly and G. Jones, "Biometric response as a source of query independent scoring in lifelog retrieval," in *Advances in Information Retrieval*, ser. Lecture Notes in Computer Science, C. Gurrin, Y. He, G. Kazai, U. Kruschwitz, S. Little, T. Roelleke, S. Rger, and K. van Rijsbergen, Eds. Springer Berlin / Heidelberg, 2010, vol. 5993, pp. 520–531.

[30] J. Salojärvi, K. Puolamäki, and S. Kaski, "Implicit relevance feedback from eye movements," in *Artificial Neural Networks: Biological Inspirations ICANN 2005*, ser. Lecture Notes in Computer Science, W. Duch, J. Kacprzyk, E. Oja, and S. Zadrozny, Eds. Springer Berlin / Heidelberg, 2005, vol. 3696, pp. 513–518.

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

[32] M. Tkalčič, A. Odič, A. Košir, and J. Tasič, "Impact of Implicit and Explicit Affective Labeling on a Recommender System's Performance," in *Advances in User Modeling*, ser. Lecture Notes in Computer Science, L. Ardissono and T. Kuflik, Eds. Springer Berlin / Heidelberg, 2012, vol. 7138, pp. 342–354.

[33] S. Vrochidis, I. Kompatsiaris, and I. Patras, "Utilizing implicit user feedback to improve interactive video retrieval," *Advances in Multimedia*, vol. 2011, pp. 1–18, Jan. 2011.

[34] S. Vrochidis, I. Patras, and I. Kompatsiaris, "An eye-tracking-based approach to facilitate interactive video search," in *Proceedings of the 1st ACM International Conference on Multimedia Retrieval*, ser. ICMR '11. New York, NY, USA: ACM, 2011, pp. 43:1–43:8.

[35] P. Lang, M. Bradley, and B. Cuthbert, "International affective picture system (IAPS): Affective ratings of pictures and instruction manual," University of Florida, Gainesville, Florida, US, Tech. Rep. A-8, 2005.

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

[37] P. Ekman, *Basic Emotions*. John Wiley & Sons, Ltd, 2005, pp. 45–60.

[38] P. Desmet, *Measuring emotion: development and application of an instrument to measure emotional responses to products*. Norwell, MA, USA: Kluwer Academic Publishers, 2003, ch. 9, pp. 111–123.

[39] P. Winoto and T. Y. Tang, "The role of user mood in movie recommendations," *Expert Systems with Applications*, vol. 37, no. 8, pp. 6086–6092, 2010.

[40] M. M. Bradley and P. J. Lang, "Measuring emotion: the Self-Assessment Manikin and the Semantic Differential." *J Behav Ther Exp Psychiatry*, vol. 25, no. 1, pp. 49–59, March 1994.

[41] J. A. Russell, A. Weiss, and G. A. Mendelsohn, "Affect Grid: A single-item scale of pleasure and arousal," *Journal of Personality and Social Psychology*, vol. 57, no. 3, pp. 493–502, September 1989.

[42] J. A. Russell, "A circumplex model of affect," *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, December 1980.

[43] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Y. Patras, "Deap: A database for emotion analysis using physiological signals," *IEEE Transactions on Affective Computing*, vol. 3, pp. 18–31, 2012.

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

[45] M. Tkalčič, J. Tasič, and A. Košir, "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*, 2010.

[46] L. R. Goldberg, J. A. Johnson, H. W. Eber, R. Hogan, M. C. Ashton, C. R. Cloninger, and H. G. Gough, "The international personality item pool and the future of public-domain personality measures," *Journal of Research in Personality*, vol. 40, no. 1, pp. 84–96, 2006.