

Towards enhanced video access and recommendation through emotions

Eva Oliveira

LaSIGE, University of Lisbon
FCUL, 1749-016 Lisbon, Portugal,
IPCA, 4750-117 Arcozelo BCL,
Portugal
+351 2175000533
eoliveira@ipca.pt

Nuno Magalhães Ribeiro

CEREM – Centro de Estudos e
Recursos Multimidiáticos
Universidade Fernando Pessoa
Praça 9 de Abril, 349
4249-004 Porto – Portugal
nribeiro@ufp.edu.pt

Teresa Chambel

LaSIGE, University of Lisbon
FCUL, 1749-016 Lisbon
Portugal
+351 2175000533
tc@di.fc.ul.pt

Abstract

In this work we discuss an emotional classification of videos based on users physiological signals and video low-level processing. This kind of automatic user classification has the potential to increase the naturalness of interaction. Everyday the number of online videos and films is increasingly available. The emotional dimensions of videos require specific tools to enable video access and search through emotions. We will discuss video classification based on emotions and how user physiological signal from users can contribute to automatically classify videos and enhance a recommendation system based in emotions.

Keywords

video affective classification, user physiological signals, facial expression recognition, video access, video search, video recommendation

ACM Classification Keywords

H3.3 [Information Storage and Retrieval]: Information Search and Retrieval; H.5.2 [Information Interfaces and Presentation]: User Interfaces;

Introduction

One of the greatest strengths of video is its power to generate attitudes and emotions as no other medium can. It is also an excellent medium for displaying affective information. Watching films on the web is also fast becoming a pastime, requiring the need for new and improved search tools that allow viewers to

Copyright is held by the author/owner(s).

CHI 2009, April 4 – 9, 2009, Boston, MA, USA

ACM 978-1-60558-246-7/09/04.

quickly find a specific subject, cast member or content in a film or movie scene. Given that videos are emotionally rich, it is also required tools to find video scenes with specific emotions interacting according to our emotional states. First, we review significant works that study video emotion classification, both by user-centered analysis and content-based classification. Then we propose a set of considerations that contribute to enhance the access and discovery of video emotions in the context of web video delivery.

Emotional Classification of videos

Emotionally classifying a media item, be it a video, text or an image, requires the analysis of the item to determine its affective properties, followed by the choice of an appropriate affective model to better translate the item emotional properties into an emotional representation. Videos have several affective dimensions, including 1) the director's intention when making a scene, the objective emotional content of a scene, and 2) video characteristics informed by the cinematographic techniques used to recreate some emotional environment or induce an intended emotion in the viewer 3) the emotional impact of the scene on viewers. There are already some emotional perspectives for movie classification. From the director's point of view there are a number of cinematographic techniques (Arijon, 1976) to induce a specific emotional environment, like shots duration, lightning conditions, color and movements. From the user and content point of view there are three main models of emotion. The Categorical perspective, also known as Darwinist perspective, where emotions are classified into classes: Ekman's list of basic emotions – Anger, Disgust, Fear, Sadness, Surprise and Happy – are commonly accepted as being the face of this

emotional model (Ekman & Oster, 1979). The second is the Dimensional model, proposed in (Russell, 1980) classifies every emotion in a 2D space by valence (positive/negative) and arousal (calm/excited). The third perspective is the Appraisal Theory, also categorical, constitutes of a set of words describing emotional descriptors. Each emotional descriptor is defined as the evaluation of the interaction between someone and their goals, beliefs and environment: Scherer (2005) defined emotional descriptors based on these assumptions to create the Geneva Affect Label Code (Scherer, 2005).

Video Affective Classification by User's Physiological Signals

The classification of movies by its affective dimensions has recently been the focus of some notable studies. In (Money & Agius, 2009) is reported an exhaustive experimentation on how user physiological responses vary when elicited by different genres of video content in order to validate the development of personalized video summaries. They showed specific video segments to a group of viewers monitored with biometric artifacts (electro dermal response, respiration amplitude, respiration rate, blood volume pulse and heart rate) and concluded that there are significant differences between users when watching the same video segments. Money et al. consider the dimensional theory to classify the affective results in this study. Another conclusion from this study is that there is a strong relation between some movie genres and some biometric artifacts, being horror/thriller films and comedy those who have the most impact on users. One of their output classifications - the aggregated responses to video genres - is a percentage of all user

biometric responses that were considered significant. However, their outputs do not fall into a specific emotional category. In another experiment (Smeaton & Rothwell, 2009), the authors developed a cinema context experimentation by measuring physiological signals from users in order to classify the emotional impact of films. With the objective of testing if video highlights can be detected from physiological signals, they tested whether the film experience is different in a group context or in an individual context, and found out that music is correlated with users emotional highlights. Furthermore, using the categorical perspective of emotions, they manually classified the supposed evoked emotions of the prepared films with 21 emotional categories (Salway & Graham, 2003). They also analyzed low-level audio features of the film in order to distinguish speech from music and from silent frames, so as to improve the accuracy of emotional classifications. The output classification is obtained by comparing the detected biometric peaks with the manually introduced emotional categories for each movie segment. Another study (Arapakis et al., 2009) proposes a video retrieval system based on user feedback that derives from real-time facial expression recognition. Here, the affective classification is automatically processed by eMotion: a facial expression recognition system (Valenti, Sebe & Gevers, 2007), that outputs Ekman's basic emotions, decides, depending on the result, if a video is relevant or irrelevant, which interferes in retrieval ranking positioning. Although these works feature advanced research in the affective classification of videos, there is still a lack of emotion categorization output.

Video Affective Classification by its Content

Algorithms for video content analysis allow the capture of low-level features that can provide semantic meaning for that video data stream. Kang (2003) proposes a content-based affective classification based in Hidden Markov Models to analyze low-level features such as color, motion and shot cut rate information, and output their results into a 2D dimensional emotional space. They focused only in fear, sadness and joy due to the difficulty of discriminating fear from anger by using just color or motion and shot cut rate. From this work results a correspondence between these emotion categories and the low-level features analyzed (Kang, 2003). In another work, authors developed a framework to represent and model the affective content in videos (Hanjalic & Xu, 2005). They based their content analysis in cinematographic techniques, such as motion analysis, vocal effects, shot length and sound and rhythm analysis, being perfectly aware that the emotional result stems from the director's intention - the *expected* emotion, and not necessarily correspond to the viewer actual feelings. They used the dimensional approach of a 2D space (valence, arousal). One of the relevant results of this work is the creation of links between the 2D dimensional emotion space and the low-level features of the video. More recently, Soleymani et al. (2008) propose an emotional classification of movies, which they have called *complementary approach* of emotional models, to analyze affective content. The authors use a cinematographic perspective to analyze low-level features of the film from both audio and visual and map them into basic emotions categories, and then used the dimensional perspective to correlate each with other emotional descriptors to give more meaning to the

selected category. After carrying out a survey to investigate the relevance of Ekman's list of basic emotions, where 9 users were asked to propose a word for every scene of two random films. The result was the adoption of only five of the six basic emotions and add the neutral state to assort their set of emotional classes, because *Disgust* turned out to be inexistent, when finding matches in genres or viewers feelings (Soleymani, Chanel, Kierkels & Pun, 2008). From our perspective, this work is one of the most complete analyzer of emotions based on content, given that all the other content analysis studies are somehow incomplete in terms of the outputted emotional categories, making it difficult to assert the affective comprehension of movies. The reason why most of these works are very recent is the fact that only in the past few years has the affective computing area brought emotions to a wider range of applications (Picard, 1999).

Requirements for Video Classification based on Affective Information to collaborative recommendation systems

This paper builds on our previous work (Oliveira 2008) as it discusses the classification requirements needed to represent emotions, oriented towards movies properties. In this section we present a set of classification considerations related with video affective properties representation in order to help creating a mechanism for obtaining meaningful affective information from the perspective of viewers, movies and director's *expected* affective impact. In our opinion, the following is the most relevant aspects that must be included in classification procedure in order to enable the creation of emotion-oriented applications that include video, users and their affective relationship.

- The choice of the semantic representation to emotional information is the crucial aspect when dealing with affective classification. This representation must have the following characteristics: 1) it must be simple enough so that emotions can be captured from the diversity of existing emotional data gathering methods; 2) it should be readable by any module of an emotion processing system; 3) it should be in accordance with the W3C Emotion Markup Language guidelines (Schroder, Wilson, Jarrold & Evans, 2008) so as to enable the communication between future web services. In our opinion, the semantic representation can use EARL (Emotion Annotation and Representation Language), a language based in XML proposed in (Schröder, Pirker & Lamolle, 2006), given that it supports the representation of the three main models of emotion, respects all the characteristics indicated above, and allows the integration of new elements in its specification.
- It is important to collect a set of films covering each and every basic high intensity emotion, in order to ensure that we have, at least, one recognized category. Additionally, a neutral state must also be included in the list of basic emotions, which corresponds to the occurrence of an emotionless moment in the video.
- Recognition of user affective information should be handled by physiological signals and recognition of facial expressions. Both should be used complementarily to reinforce the accuracy of the recognition.

The classification of user physiological signals along with the analysis of cinematographic techniques of videos can create the rules of an emotional

recommendation system. The following are some requirements for such a mechanism:

1. Every video must be classified affectively from the user and from the content perspective.
2. Every scene must have at least one basic emotion associated, in order to have a complete affective description of the video along its duration.
3. Every user has an emotional profile constituted by all his implicit (physiological) classification of every classified movie.
4. Every video has an emotional profile constituted by the low-level analysis of scene and by its genre.
5. User emotional profiles are constructed over time by collecting and analyzing all emotional user data detected for each video scene;
6. The three major emotional models should be represented, if possible, in every classification of every scene, improving the recommendation due to the wider range of data to relate.
7. User physiological signals when captured in real-time can be used to recommend videos based on the emotion the user is feeling, and put in headline a relaxed video if the user is stressed.
8. User physiological signals can change the way we watch the videos, adapting the interface to the main emotion presented, changing interface colors, movements, and even the lightning of it elements.

These classification requirements and developments derive from our ongoing work (Oliveira 2008) on the classification and access of videos based on emotions in the way we watch our videos and the way we organize, search and interact with them. They are also reflected in the way we chose to organize the videos and with

which we tend to develop an affective connection. At the individual video level, the video can be presented with an emotional timeline, representing the video's emotions along time, either from a more objective or more subjective perspective. Users can use this information to gain more awareness of the emotions involved, and of how different the two perspectives are, and also to access scenes based on their dominant emotions.

Conclusion

It has been made clear in this paper that the purpose of this study is to explore the emotional classification of videos, either by user physiological signals and video low-level analysis. Measuring physiological signals of users is a natural input mechanism that can be used to automatically classify information with emotional semantic. This classification must have valuable information to recommend videos emotionally compatible with the user, in certain moments or anytime he asks for an emotion specific video. The recognition of these emotional states also allows the adaptation of interfaces based on videos main emotion. Our system is in development and will empower the discovery of interesting emotional information in unknown or unseen movies, compare reactions to the same movies among other users, compare directors intentions with users effective impact, analyze over time our reactions or directors tendencies. We also believe it would be interesting to analyze the differences between the user physiological signals in response to known movies with newer ones, and see the importance of classifying the unfamiliar ones.

Eva Oliveira is currently a lecturer in the School Technology at the Polytechnic Institute of Cávado and Ave, Portugal. She is taken her Phd at Human-Computer Interaction and Multimedia Group at LaSIGE/University of Lisbon, since 2008.

Teresa Chambel is a professor at the University of Lisbon, Portugal, where she teaches since 1988, and received a Ph.D. in Informatics, on video, hypermedia and learning technologies, distributed hypermedia. More information may be found at: <http://www.di.fc.ul.pt/~tc>

Nuno Magalhães Ribeiro is Associate Professor at the Faculty of Science and Technology of Fernando Pessoa University (Porto, Portugal). He holds a Ph.D. on Computer Science by the University of York (U.K.).

Acknowledgements

This work was partially supported by LaSIGE through the FCT Pluriannual Funding Programme.

Bibliography

Arapakis, I., Moshfeghi, Y., Joho, H., Ren, R., Hannah, D., & Jose, J. M. (2009). Enriching user profiling with affective features for the improvement of a multimodal recommender system. In *Conference on image and video retrieval*.

Arijon, D. (1976). *Grammar of the film language* (illustrated ed.). New York: Hastings House.

Ekman, P. & Oster, H. (1979). Facial expressions of emotion. *Annual Review of Psychology*.

Fasel, B. & Luetttin, J. (2003). Automatic facial expression analysis: A survey. *Pattern Recognition*, 36(1), 259-275.

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

Kang, H. B. (2003). Affective content detection using hmms. *Proceedings of the Eleventh ACM International Conference on Multimedia*, 259-262.

Money, A. G. & Agius, H. (2009). Analysing user physiological responses for affective video summarisation. *Displays*, 30(2), 59-70.

Oliveira, E. & Chambel T. (2008). Emotional Video Album: getting emotions into the picture. In *emotion-in-hci'2008, The 4th Workshop on Emotion in Human-Computer Interaction: Designing for People, at HCI'2008, the 22nd BCS HCI Group conference on HCI: Culture, Creativity, Interaction*, Liverpool, UK, September 1-5, 2008.

Picard, R. W. (1999). Affective computing for HCI. In *Proceedings of HCI international (the 8th international conference on human-computer interaction) on human-computer interaction: Ergonomics and user interfaces*.

Russell, J. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*.

Salway, A. & Graham, M. (2003). Extracting information about emotions in films. *Proceedings of the Eleventh ACM International Conference on Multimedia*, 299-302.

Scherer, K. R. (2005). What are emotions? And how can they be measured?. *Social Science Information*, 44(4), 695.

Schroder, M., Wilson, I., Jarrold, W., & Evans, D. (2008). What is most important for an emotion markup language?. *Proc. Third Workshop Emotion and Computing*.

Schröder, M., Pirker, H., & Lamolle, M. (2006). First suggestions for an emotion annotation and representation language. In *Proceedings of LREC*.

Smeaton, A. F. & Rothwell, S. (2009). Biometric responses to music-rich segments in films: The cdvplex.

Soleymani, M., Chanel, G., Kierkels, J. J. M., & Pun, T. (2008). Affective ranking of movie scenes using physiological signals and content analysis.

Valenti, R., Sebe, N., & Gevers, T. (2007). Facial expression recognition: A fully integrated approach. In *Int. Workshop on visual and multimedia digital libraries*.