Chapitre d'actes   2009   Accepted version   Open Access

This is an author manuscript post-peer-reviewing (accepted version) of the original publication. The layout of the published version may differ .

- -

A Collaborative Personalized Affective Video Retrieval System

- -

Soleymani, Mohammad; Davis, Jeremy; Pun, Thierry

A collaborative Personalized Affective Video Retrieval System

Mohammad Soleymani   Jeremy Davis   Thierry Pun

Computer Vision and Multimedia Laboratory, Computer Science Department

University of Geneva

Battelle Campus, Building A, Rte. De Drize 7,

CH - 1227 Carouge, Geneva, Switzerland

mohammad.soleymani@unige.ch, davis4@etu.unige.ch, thierry.pun@unige.ch

http://cvml.unige.ch

Abstract

In this demonstration, a collaborative personalized affective video retrieval is introduced. A dataset of 155 video clips extracted from Hollywood movies were annotated by the emotion felt by participants. More than 1300 annotations from 40 participants were gathered in a database to be used for affective retrieval system. The retrieval system is able to retrieve videos based on emotional keyword query as well as arousal and valence query. The user's personal profile (gender, age, cultural background) was employed to improve the collaborative filtering in retrieval.

1. Introduction

Multimedia information retrieval is attracting significant interest due to the exponential growth of digital multimedia content. For example the open access online video database, YouTube, had a watching rate of more than 100 millions videos per day in 2006 [1].

In this paper we use a valence-arousal and label based approach for affect representation and annotation. Russell [2] proposed this 3D continuous space called the valence-arousal-dominance space which was based on a self-representation of emotions from multiple subjects. In this demonstration only arousal and valence dimensions were used. Hanjalic et al. [3] introduced the idea of "personalized content delivery" as a valuable tool in affective indexing and retrieval systems and used arousal and valence for video affective representation.

Collaborative filtering has been used extensively in recommendation systems [4]. In recommendation systems, the rating of a product is estimated based on the ratings of other users. Most of the collaborating filtering systems use K-nearest neighbor (KNN) classifier to choose and rank the similar products.

In this system, we used a weighted KNN to find the closest clips to an affective query. During the retrieval process, the user first registers him- or herself by giving his information to make a new profile. Then the user will be able to give a query both by emotional keywords and arousal and valence. The arousal and valence query is manipulated by clicking on a valence-arousal plane to indicate the arousal and valence values.

1.1. Technical requirements

This demonstration is web-based and it only needs a broadband internet connection. In the case of internet connectivity absence, it is also possible to create an offline, stand alone system. The demo is accessible via the following URL:

http://cvml.unige.ch/videoAnnotation/private/retrieve.php

2. Methods

2.1. Assessments

A web based annotation system has been launched to assess participants' felt emotion. In this system, a user signs up giving his/her personal information including gender, age, and email address. The system also asked optional information like, cultural background and origin which will help the system to form a profile of the user.

Figure 1 shows a snapshot of the assessment interface where a video clip is being shown. After watching each video clip, the participant expressed his/her felt emotion using arousal and valence, quantized in nine levels. The participants also chose the emotional label manifesting his/her felt emotion. The emotion labels are afraid, amused, anxious, disgusted, joyful, neutral, and sad. These labels have been chosen based on the labels assessed in our pilot experiments. During our pilot experiments we asked 10 participants to freely express their emotions, elicited by movie scenes with words. These emotional keywords were the ones which appeared more frequently [5]. Note that they roughly correspond to the six basic "Ekman's emotions" [6].
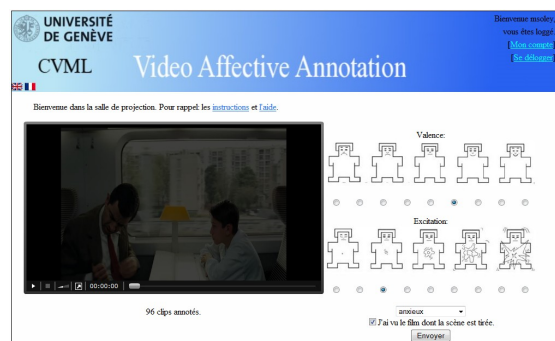

Figure 1. A snapshot of the affective annotation platform.

Table1. List of the movies in the dataset.

Drama Movies	Comedy Movies
The pianist, Hotel Rwanda, Apocalypse now, American history X, Hannibal	Man on the moon, Mr. Bean's holiday, Love actually
Horror Movies	**Action Movies**
Silent hill, 28 days later, The shining	Kill Bill Vol. 1, Kill Bill Vol. 2, Platoon, The thin red line, Gangs of New York

2.2. Dataset

The dataset is extracted from 16 full length Hollywood movies (mostly popular movies). The majority of movies were selected either because they were used in similar studies (e.g. [7]), or because they were recent and popular. 155 short clips which are about one to two minutes long were chosen from these movies to form the dataset. The choice has been made by the authors manually from the emotional scenes of the movies.
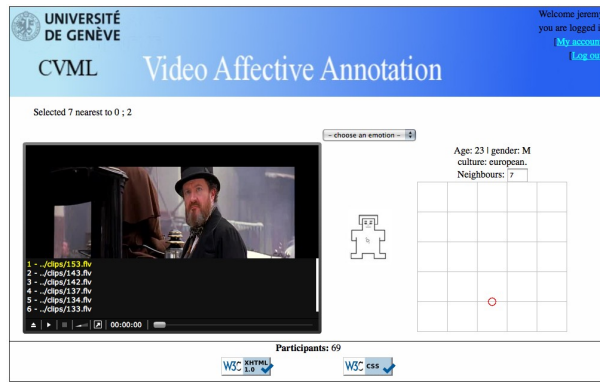

Figure 2. A snapshot of the affective retrieval platform.

2.3. Retrieval process

2.3.1   Query by emotion label

The implemented retrieval system is able to retrieve video clips in response to a query by label, t_q, (see Figure 2). The label or emotional keyword can be selected from the list of seven emotions which has been used for tagging. For each of n users who annotated this clip, the tag from similar users will be more valuable in this process. The tags, t_i, which were selected by the same gender/age and social background group, G_j, will count twice and the number of the rest of tags will remain the same. After computing the query tag score of each clip, the clips are sorted based on each clip's score, s_j.

$$s_j = \sum_i \alpha.t_i, \alpha = \begin{cases} 2, i \in G_j \\ 1, i \notin G_j \end{cases}, t_i = \begin{cases} 0, t_i \neq t_q \\ 1, t_i = t_q \end{cases}, i=1..n \quad (1)$$

2.3.2   Query by arousal and valence

The second option to query this system is by arousal and valence. This has been done by means of an interface which receives arousal and valence by clicking on a two dimensional valence-arousal plane. The system records the coordinates and ranks the clips by their arousal and valence using the Euclidean distance. The video clips arousal and valence (a_j, v_j of j-th clip) were computed based on the user's profile and participants' evaluations (see equations (2) and (3)). Again, the samples from similar profiles were considered twice and the median of arousal and valence for video clips were computed. Finally using a K nearest neighbor approach the system retrieves and ranks results.

$$a_j = median\left\{ a_j^i \right\}, i=1..n \quad (2)$$

$$v_j = median\left\{ v_j^i \right\}, i=1..n \quad (3)$$

3. Summary

A collaborative affective video retrieval system has been introduced. The ability of the user profiling and using the tags from the users with similar taste is considered to improve the retrieval results.

4. Acknowledgement

References

[1] "YouTube serves up 100 million videos a day online," http://www.usatoday.com/tech/news/2006-07-16-youtube-views_x.htm, 2006.

[2] J. A. Russell and A. Mehrabian, "Evidence for A 3-Factor Theory of Emotions," *Journal of Research in Personality*, vol. 11, no. 3, pp. 273-294, 1977.

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

[4] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 6, pp. 734-749, June 2005.

[5] M. Soleymani, G. Chanel, J. J. M. Kierkels, and T. un, "Affective Characterization of Movie Scenes Based on Multimedia Content Analysis and User's Physiological Emotional Responses,", ISM '08: Proceedings of the 2008 Tenth IEEE International Symposium on Multimedia IEEE Computer Society, pp. 228-235, 2008.

[6] P.Ekman, W.V.Friesen, M.O'Sullivan, A.Chan, I.Diacoyanni-Tarlatzis, K.Heider, R.Krause, W.A.LeCompte, T.Pitcairn, and P.E.Ricci-Bitti, "Universals and cultural differences in the judgments of facial expressions of emotion," *Journal of Personality and Social Psychology*, vol. 53, no. 4 1987.

[7] J. Rottenberg, R.D.Ray, and J.J.Gross, "Emotion elicitation using films," in *The handbook of emotion elicitation and assessment*. A.Coan and J.J.B.Allen, Eds. London: Oxford University Press, 2007.