Exploiting Evidential Theory in the Fusion of Textual, Audio, and Visual Modalities for Affective Music Video Retrieval

Shahla Nemati,
Department of Computer Engineering, Faculty of Engineering,
Shahrekord University, Iran
nemati@eng.ui.ac.ir

Ahmad Reza Naghsh-Nilchi,
Department of Artifital Intelligent, Faculty of Computer Engineering, University of Isfahan,
Isfahan, Iran
nilichi@eng.ui.ac.ir

Abstract—**Developing techniques to retrieve video contents with regard to their impact on viewers' emotions is the main goal of affective video retrieval systems. Existing systems mainly apply a multimodal approach that fuses information from different modalities to specify the affect category. In this paper, the effect of exploiting two types of textual information to enrich the audio–visual content of music video is evaluated; subtitles or songs' lyrics and texts obtained from viewers' comments in video sharing websites. In order to specify the emotional content of texts, an unsupervised lexicon–based method is applied. This method does not need any human-coded corpus for training and is much faster than supervised approach. In order to integrate these modalities, a new information fusion method is proposed based on the Dempster-Shafer theory of evidence. Experiments are conducted on the video clips of DEAP dataset and their associated viewers' comments on YouTube. Results show that incorporating songs' lyrics with the audio–visual content has no positive effect on the retrieval performance, whereas exploiting viewers' comments significantly improves the affective retrieval system. This could be justified by the fact that viewers' affective responses depend not only on the video itself but also on its context.**

Keywords—*Affective music video retrieval; Lexicon-based sentiment analysis; Information fusion; Emotion detection*

I. INTRODUCTION

Affective video content analysis systems are designed to estimate the emotional response of viewers when watching a video. Due to the growing awareness of their potential for personalized multimedia information retrieval, the importance of these systems has been increased in recent years. Furthermore, affective video analysis is believed to lend to numerous multimedia applications such as video production, video summarization, and more importantly, video recommendation [1].

With the availability and growth of video sharing websites from one hand, and the ease of capturing and sharing videos from another hand, more and more video contents are being published on the Internet. For example, YouTube, the third most popular website globally, has more than 4 billion video views per day and receives 300 hours of new video every minute. Therefore, finding relevant videos matching the desires and needs of users is a challenging task [2].

In this paper, music video is considered as the case study for evaluating the retrieval system. Music is the most popular category of videos and was the most searched-for topic on YouTube in 2014. Moreover, more than 38% of YouTube video views were music videos [3]. However, there are several differences between music video and other genres of videos making the problem of affective music video retrieval more challenging. For example, due to the frequent change of content and lighting, shot length is usually shorter in music videos while, their frame differences are larger. Another difference is that although face features are quite important, in many music videos it is impossible, or at least very difficult, to apply these features [4]. This has several reasons such as the use of special effects or colorful lighting and the diagonal or horizontal position of face (e.g. when the artist is dancing or sleeping) [4]. Furthermore, unlike most genres, face features are less useful in music videos because there is not necessarily a correlation between artist's face and the affect evoked in a viewer while watching a music video.

Due to the importance of visual and auditory modalities, most of existing affective video analysis methods apply audio–visual features [1, 5]. Some studies apply user's physiological responses as well. However, such methods are of little use because they need specific measurement devices and expect too much users' involvement [6]. Most recently, viewers' comments on a video have been applied as a new modality for affective content detection [2]. These comments produce a textual modality with a different nature in comparison with that of the audio-visual modalities. This is due to the fact that the interactivity of video sharing websites such as YouTube provides such an external modality that may be applied to compensate the internal contents of video clips. This interactivity is obtained through the use of various facilities that allow viewers to share their opinion after watching a video. [2].

Data fusion algorithms are the main components applied in integrating affective information from various modalities

978-1-5090-6454-0/17/$31.00© IEEE

in multimodal approaches. The fusion can be performed in either feature–level or decision–level. Both levels are applicable for combining audio–visual features (as they are frame–level features), but only decision–level fusion may be applied to combine the audio–visual decisions with textual information (since textual features belong to the clip–level). In fact every music video consists of several frames from which audio and visual features are extracted, while each textual comment describes the music video clip as a whole.

Fusion algorithms such as product of confidence measures, voting, and sum are applied in most studies [7, 8]. Although these simple methods are common in data fusion, it is revealed that more theoretical approaches such as Dempster-Shafer (DS) of evidence obtain more accurate results for affective video retrieval [2].

The DS-based fusion method is an evidential approach with certain advantages compared to similar probabilistic methods such as Bayesian method [9, 10]. However, the main drawback of the existing DS-based method suggested for affective video retrieval systems is the fact that it do not deal with the so-called combination paradox problem that leads to obtaining a counterintuitive conclusion. This problem was first highlighted by Zadeh [11] and it occurs when the evidential system is highly conflicting. The combination paradox may occur in affective video retrieval systems because as pointed earlier, these systems usually apply modalities of different natures. In this study, audio, visual, and textual modalities that were recently proposed in [2] are applied. These modalities are of conflicting nature and thus applying the ordinary DS-based fusion may decrease the performance and reliability of the retrieval system.

To overcome the above-mentioned drawback of the original DS method, in this paper, an evidential approach for data fusion based on the modified DS theory of evidence proposed by Yang et al. is suggested [10]. In this method, decisions made based on auditory, visual, and textual modalities are considered as evidence for the overall affect category of the video. Experiments are conducted on the DEAP dataset, a multimodal dataset for the analysis of human affective states [12]. In this paper, the following research questions are addressed:

- Is it useful to incorporate lyrics as textual modality with the audio–visual modalities for affective music video retrieval?
- Do the existing fusion methods perform well on affective music video retrieval?
- At which level the auditory, visual, and textual modalities should be combined?

II. Related Work

Affective video analysis has been receiving growing attention in recent years from both academia and industry [6]. A common two–dimensional model is called Valence–Arousal space (V–A space) [8]. Valence represents positive versus negative affect. Arousal, on the other hand, shows the level of activation caused by the affect (i.e., low versus high). The main benefit of this model is that basic emotions can be represented as different areas on V–A coordinates [13]. In this paper, the V–A affective model are applied as well.

Most affective video analysis approaches are multimodal. These systems considered different combinations of modalities including visual–speech, face–physiological signal, facial–vocal–body movements, and speech–physiological signal [14]. However, the auditory and visual modalities can be considered as the most common modalities. Various kinds of audio–visual features including pitch, energy, motion, and color are extracted from videos [13].

Apart from the above-mentioned combinations, some recent studies have also exploited textual modality. For example, Xu et al. analyzed the subtitle files of DVD/DivX videos to assist affective content detection [6]. More recently, Poria et al. proposed a novel methodology for multimodal sentiment analysis that uses audio, visual and textual modalities [15]. They applied text transcription of the video clip along with audio–visual content to extract information from user generated multimodal data in contexts such as e–learning and e–health [16]. As pointed out earlier, in this paper we aim to show that incorporating textual modality in affective video retrieval can improve the performance of system. However, despite the above–mentioned existing studies, the transcripts (lyrics in our case) as textual modality will not applied. Instead, textual viewers' comments will be applied as an external source of knowledge to compensate the auditory and visual modalities.

III. Methodology

The overall structure of the proposed system is shown in Figure 1. In the first part of the figure, part (A), the auditory, visual, and textual contents of the input music video are extracted using the feature extraction modules. Parts (B) and (C) of Figure 1 show two possible ways of fusing different modalities. In the first form, part (B), the audio and visual features are first combined into an audio–visual feature vector using an early fusion method (the Feature–Level Fusion module) and then, an audio–visual classifier is applied to specify the affect category of the input music video. Finally, using a late fusion method (the Decision–Level Fusion module), this decision is aggregated with the result of textual classifier.

An alternative method of fusing the auditory, visual, and textual modalities is depicted in part (C) of Figure 1. In this method, three independent classifiers are applied to the audio, visual, and textual feature vectors and their results are then combined into an overall affect class using a decision–level fusion method. It should be noted that the decision–level fusion is necessary for fusing textual modality because textual features belong to the clip–level while, audio–visual features are related to the frame–level. In other words, a typical music video consists of several frames from which audio and visual features are extracted, whereas each text describes the video clip as a whole.

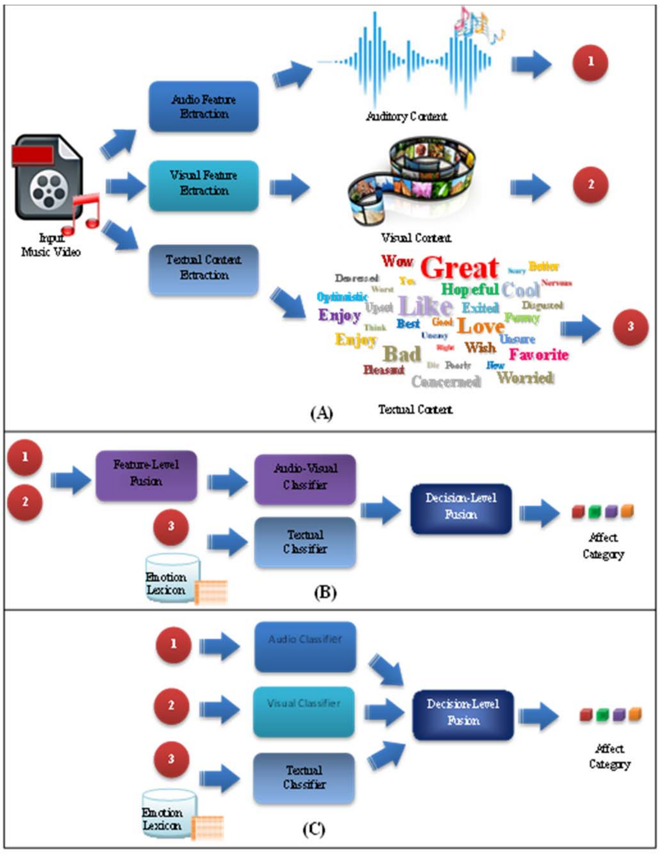

Fig. 1. The Overall structure of the proposed method: (A) Feature extraction, (B) Feature-level fusion, (C) Decision-level fusion.

A. Audio–Visual Feature Extraction

The visual content of a music video is usually designed in a way that coordinates with the music to have the intended impact on the viewers [17]. Therefore, both the auditory and visual contents of music videos are useful for affective analysis. In this study, low–level audio–visual features are applied because it has been shown that there is a relationship between these features and the emotional states of viewers [8]. Specifically, previous studies showed that zero–crossing rate (ZCR), energy, Mel–frequency cepstral coefficients (MFCC), and pitch features from the auditory channel and color, motion, and lighting key among visual features have direct correlation with the affective category [8, 18]. This has motivated our choice of such audio–visual features for affective music video retrieval.

B. Classification

In this study, classifying input music videos with respect to their affective content is addressed. As pointed out earlier, there are several psychological models to explain human emotion. In the current study, the two–dimensional V–A model is applied. This model divides the affective space into four quadrants in the V–A space were labeled as “negative–high” (NH), “negative–low” (NL), “positive–high” (PH), and “positive–low” (PL), respectively.

In this study, two possible ways of classifying music videos based on their audio–visual content are evaluated. In the first method (two–step fusion), as shown in Figure 1 (B), auditory and visual features are first extracted and then they are merged into one feature vector. Finally, the audio–visual feature vector is fed into an audio–visual classifier and the result is applied with the result of classifying the associated texts. More details about the fusion methods will be presented in the following section. In the second method (decision–level fusion), as shown in Figure 1 (C), textual, auditory, and visual contents are first classified separately and then, the results are combined using a decision–level fusion method.

In order to classify music videos using the above–mentioned methods, Support Vector Machine (SVM) is applied in the current study. SVM is a popular supervised classifier in machine learning and data mining.

In order to identify the affective tone in texts, sentiment analysis approaches are usually applied. Sentiment analysis or opinion mining is a sub–field of data mining that aims to extract people’s opinions, evaluations, and emotions from their writings [19]. In this study, the effectiveness of exploiting two kinds of textual content is compared, namely internal textual content obtained from songs’ lyrics, and external textual content obtained from users’ comments in video sharing websites. Although these textual contents have different natures, a same method may be employed to classify them with respect to their affective content.

Existing approaches for affective classification of textual content may be classified into two main categories: machine learning approaches and lexicon–based methods [19, 20]. Although machine learning approaches have some advantages, they have obvious disadvantages such as domain dependency and needing a corpus of human–coded texts for training [19]. Lexicon–based approaches, on the other hand, use a list of affective words and their associated emotions. It has been shown that lexicon–based approaches generalize well and provide significant gains. Moreover, recent studies show that lexicon-based approach is faster than supervised methods while producing similar results [21]. In the current study, the NRC word–emotion lexicon (NRC–10) is applied [22]. This lexicon has annotations for about 14,000 words.

Each word in the lexicon is associated with eight binary affect labels: anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. For classifying an input text, integer–valued affect features are applied in an unsupervised scheme. These features show the number of word in the input text associated with different affect labels in the lexicon [22].

C. Fusion

In this study, a fusion method based on the modified Dempster–Shafer (DS) theory of evidence proposed by Yang [10] is suggested which reduces the combination paradox problem.

The first step for applying the DS theory is to identify the problem domain by a finite set ϕ of mutually exclusive hypotheses, called the frame of discernment. The next step is defining a mass function, $m(A)$, that is a basic probability assignment (BPA) to characterize the strength of evidence supporting each subset $A \subseteq \phi$. The final step is utilizing the Dempster's rule of combination to aggregate two independent bodies of evidence into one body of evidence as follows [23]:

$$(m_1 \oplus m_2)(A) = \frac{\sum_{X \cap Y = A} m_1(X) m_2(Y)}{1 - K_{1,2}} \quad (1)$$

$$K_{1,2} = \sum_{X \cap Y = \emptyset} m_1(X) m_2(Y)$$

where, $K_{1,2}$ is applied as a normalization factor to ensure that the combination $m_1 \oplus m_2$ remains a BPA. This factor is also named conflict coefficient since it reflects the degree of conflict between the evidences. If $K_{1,2} = 0$, then the evidences have no conflict; while, $K_{1,2} = 1$ represents a completely conflicting evidences and $0 < K_{1,2} < 1$ represents that two pieces of evidences have partial conflict in supporting evidence.

In the current study, classifiers' outputs are considered as evidence for the final affect category of music videos and the normalized probability function is suggested as follows:

$$m_d(\{c_i\}) = P(c_i | x_d) \quad (2)$$

where $m_d(\{c_i\})$ is the associated mass function of each modality d, $P(c_i | x_d)$ denotes the probability of a music video belonging to class c_i given the feature vector x_d, $c \in [1, \cdots, C]$ is the final affect category to which the music video is assigned, and C is the total number of categories.

As pointed out earlier, the main drawback of the existing DS-based method suggested for affective video retrieval systems is the fact that it do not deal with the so-called combination paradox problem that leads to obtaining a counterintuitive conclusion. To deal with this drawback we suggested to combine the evidences m_1, m_2 through a combination rule as follows [10]:

$$mFinal(X_k) = RelDiff_{1,2}(X_k) * K_{1,2} + \sum_{X \cap Y = X_k} m_1(X) * m_2(Y) \quad (3)$$

where, $RelDiff_{1,2}(X_k)$ is the relative difference factor of two pieces of evidence and is calculated as follows:

$$RelDiff_{1,2}(X_k) = \frac{|m_1(X_k) - m_2(X_k)|}{DegDiff_{1,2}}, k = 1,2,\ldots,2^N \quad (4)$$

The denominator $DegDiff_{1,2}$ indicates the degree of difference between the sources of evidence and is defined by:

$$DegDiff_{1,2} = \sum_k |m_1(X_k) - m_2(X_k)|, k = 1,2,\ldots,2^N \quad (5)$$

Having calculated these combined BPAs, $mFinal$, the final affect category to which the video is assigned is calculated through:

$$cFinal = argmax_i(mFinal(X_k)) \quad (6)$$

IV. RESULTS AND DISCUSSION

This study aims to show the utility of using textual modality for augmenting the audio–visual content of music videos. Specifically, the following research questions will be answered.

- Is the performance of affective video retrieval based on music video lyrics and viewers' comment enough high to consider them as textual modality? (1)
- Do the existing fusion methods perform well on affective music video retrieval?
- At which level the auditory, visual, and textual modalities should be combined?

In order to answer the first research question, the performance of the affective classification of music videos based on their associated lyrics and viewers' comments are compared with a random algorithm in Figure 2. The random algorithm, as its name suggests, randomly assigns each music video to one of four affect categories of the V–A space. This method was run 10 times and results were averaged.

As can be seen in Figure 2, the performance of (3) the affective classification of music videos based on their associated lyrics is much lower than the performance of using viewers' comments. Moreover, it can be seen that the results of employing lyrics are similar to the result of applying the random algorithm. This can be applied to answer the first research question: lyrics, in contrast to viewers' comments, are not appropriate to be applied for the affective classification of music videos.

In order to assess the utility of using viewers' comments for affective music video retrieval, the performance of classifying video clips with and without considering viewers' comments are compared. Figures 3 and 4 show the accuracy and F–Measure of classifying music videos based on the fusion of audio–visual (AV) and the fusion of audio, visual and textual modalities (AVT). In these figures, three common fusion methods namely, product rule, sum, max and ordinary DS (DS) are compared to the proposed Dempster–Shafer-based method (Mod_DS).

It can be seen that in all cases, adding viewers' comments as textual modality to the auditory and visual modalities improves the system performance. These results show that exploiting viewers' comments and combining them with the audio–visual content, significantly improves the performance of affective music video retrieval. Another notable result in these figures is that, in almost all cases, the performance of using only textual features are lower than that of visual features. Hence, it is necessary to apply the multimedia content of music videos to effectively classify them. Moreover, for all fusion methods it is obvious that almost always the fusion of audio–visual modalities lead to higher performance than using each modality independently. This emphasizes the fact that, although the computational cost of using textual viewers' comments is much lower than using the audio–visual content of video clips, the multimedia content should not be ignored for affective video retrieval. In other words, to achieve the best performance, the audio–visual content of video clips should be combined with the associated textual viewers' comments.

As shown in Figures 3 and 4, using all three modalities (i.e., darker bars in the figures), with respect to both measures, the proposed DS−based (Mod_DS) fusion method outperforms the sum, max, product rules and ordinary DS. This shows that the proposed method is a better choice than the existing fusion methods for combining the auditory, visual, and textual modalities. Therefore, the second research question is also answered successfully.

The final research question can be answered by comparing the system performance for two possible ways of incorporating textual modality with the audio–visual content of music videos (i.e., two–step and decision–level fusion approaches).

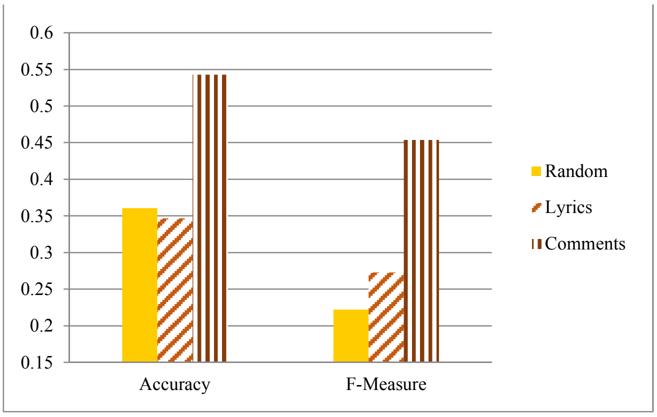

Fig. 2. Comparison of the accuracy and F–measure of classifying video clips based on random algorithm, music video lyrics and viewers' comments.

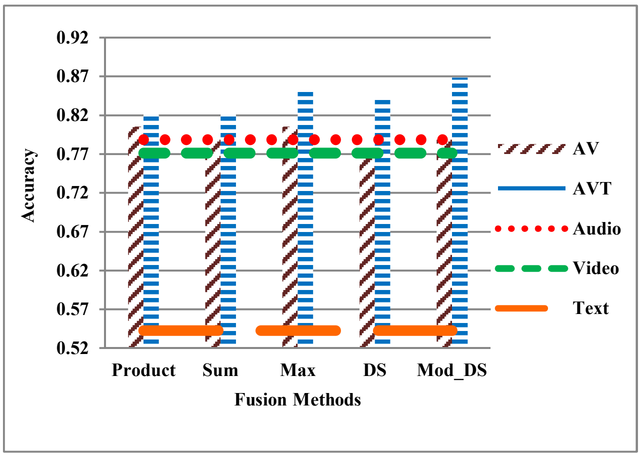

Fig. 3. Comparison of the accuracy of using product, sum, max, DS, and proposed method (Mod_DS) based on the fusion of audio–visual (AV) and audio–visual–textual (AVT) content.

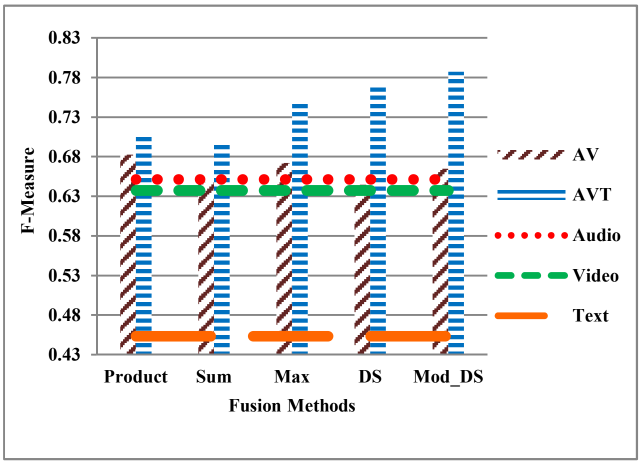

Fig. 4. Comparison of the F–Measure of using product, sum, max, DS, and proposed method (Mod_DS) based on the fusion of audio–visual (AV) and audio–visual–textual (AVT) content.

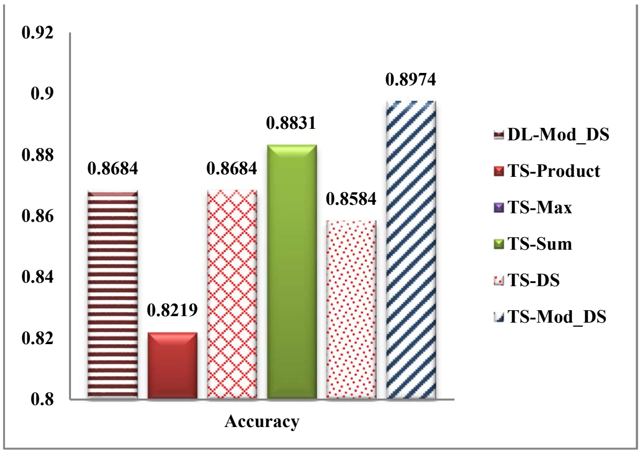

Fig. 5. Comparison of the accuracy of classification based on the two–step (TS) and the decision–level (DL) approach using the product, max, sum, DS, and the proposed fusion methods (Mod_DS).

To this aim, the two–step fusion approach (TS) using four above–mentioned fusion methods are compared with the best result of the decision–level approach (i.e., the proposed method (Mod_DS) using the decision–level approach, DL–DST) in Figures 5 and 6.

As shown in Figures 5 and 6, with respect to both measures, using the two–step fusion, the proposed method outperforms other fusion methods. Moreover, using the proposed fusion method, the performance of the two–step approach is higher than the performance of the decision–level approach. This shows that in order to obtain the best results, the auditory and visual modalities should be fused in feature–level and then the result should be fused with the textual modality at the decision–level.

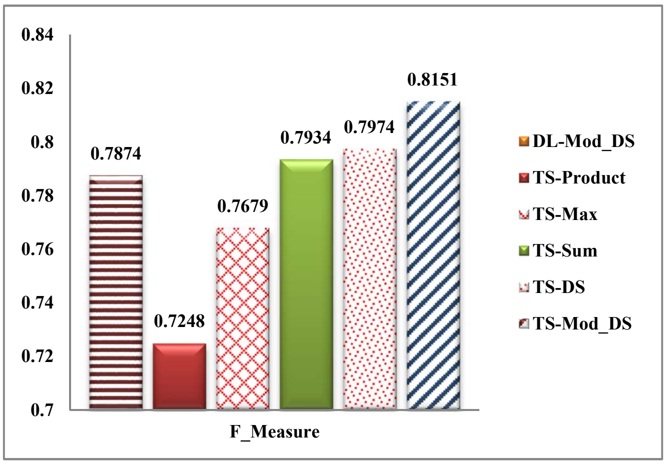

Fig. 6. Comparison of the F–Measure of classification based on the two–step (TS) and the decision–level (DL) approach using the product, max, sum, DS, and the proposed fusion methods (Mod_DS).

V. CONCLUSIONS

In this paper, the impact of combining textual modality with the audio–visual content for affective music video retrieval was investigated. Two different types of textual information were exploited separately for augmenting the audio–visual content, namely songs' lyrics and viewers' comments on the Web. The output of the proposed system (i.e., the affect category of each music video) is one of the four quadrants of the V–A (valence–arousal) affect space.

For fusing the auditory, visual, and textual modalities a two–step hybrid fusion method was proposed. In this method, a feature–level fusion is performed to combine decisions based on the auditory and visual modalities and a decision–level fusion is applied for combining the result with the decision based on the textual modality. The performance of this method was compared with the performance of a decision–level fusion method in which all three modalities provide decisions about the affect category of music video separately. For combining decisions, an adopted version of the Dempster's rule of combination was suggested in this paper. This method was compared to three frequently applied fusion methods namely, the product rule, the sum, and the max method.

Experiments were carried out on the DEAP dataset, a multimodal dataset for analyzing human affective states. All comments associated with the video clips of the DEAP dataset are collected from YouTube website. Experimental results showed that considering songs' lyrics as textual modality does not improve the retrieval performance while, combining viewers' comments with the multimedia contents significantly improves the performance of the proposed affective retrieval system. Moreover, results indicated that the proposed two–step fusion method outperforms the common decision–level approach.

There are several applications for the proposed system. For example, the proposed system helps music video delivery websites in providing more precise and convincing recommendations by taking advantage of users' preferences and the affective content of music videos. Moreover, with the aim of the proposed system music videos can be effectively produced to evoke the intended emotion of viewers.

As a direction for future research, the proposed unsupervised lexicon–based method can be enhanced by considering linguistic approaches. Another line for future research could be investigating other mathematical theories of fusion (e.g. Bayesian data fusion). Finally, in order to devise more efficient affective video retrieval systems, larger datasets incorporating textual modality is needed. Hence, developing a dataset containing audio, visual, and textual modalities is also considered as future work.

REFERENCES

[1] R. M. A. Teixeira, T. Yamasaki, and K. Aizawa, "Determination of emotional content of video clips by low-level audiovisual features," *Multimedia Tools and Applications,* vol. 61, pp. 21-49, 2012.

[2] S. Nemati and A. R. Naghsh-Nilchi, "Incorporating social media comments in affective video retrieval," *Journal of Information Science,* p. 0165551515593689, 2015.

[3] Amazing YouTube Statistics. (2015). http://expandedramblings.com/index.php/youtube-statistics.

[4] L. Agnihotri, N. Dimitrova, and J. Kender, "Method and apparatus for summarizing a music video using content analysis," ed: Google Patents, 2009.

[5] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 31, pp. 39-58, 2009.

[6] M. Xu, J. Wang, X. He, J. S. Jin, S. Luo, and H. Lu, "A three-level framework for affective content analysis and its case studies," *Multimedia tools and applications,* vol. 70, pp. 757-779, 2014.

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

[8] A. Yazdani, E. Skodras, N. Fakotakis, and T. Ebrahimi, "Multimedia content analysis for emotional characterization of music video clips," *EURASIP Journal on Image and Video Processing,* vol. 2013, pp. 1-10, 2013.

[9] M. E. Basiri, A. R. Naghsh-Nilchi, and N. Ghasem-Aghaee, "Sentiment Prediction Based on Dempster-Shafer Theory of Evidence," *Mathematical Problems in Engineering,* vol. 2014, 2014.

[10] J. Yang, H.-Z. Huang, Q. Miao, and R. Sun, "A novel information fusion method based on Dempster-Shafer evidence theory for conflict resolution," *Intelligent Data Analysis,* vol. 15, pp. 399-411, 2011.

[11] L. A. Zadeh, "A simple view of the Dempster-Shafer theory of evidence and its implication for the rule of combination," *AI magazine,* vol. 7, p. 85, 1986.

[12] S. Koelstra, C. Mühl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, *et al.*, "Deap: A database for emotion analysis; using physiological signals," *Affective Computing, IEEE Transactions on,* vol. 3, pp. 18-31, 2012.

[13] S. Wang, Y. Zhu, G. Wu, and Q. Ji, "Hybrid video emotional tagging using users' EEG and video content," *Multimedia Tools and Applications,* vol. 72, pp. 1257-1283, 2014.

[14] X. Chen, A. O. Hero, and S. Savarese, "Multimodal video indexing and retrieval using directed information," *Multimedia, IEEE Transactions on,* vol. 14, pp. 3-16, 2012.

[15] S. Poriaa, E. Cambria, N. Howardc, G.-B. Huangd, and A. Hussaina, "Fusing Audio, Visual and Textual Clues for Sentiment Analysis from Multimodal Content," *Neurocomputing,* vol. 174, pp. 50-59, 2016.

[16] S. Poria, E. Cambria, A. Hussain, and G.-B. Huang, "Towards an intelligent framework for multimodal affective data analysis," *Neural Networks,* vol. 63, pp. 104-116, 2015.

[17] S. Zhang, Q. Huang, S. Jiang, W. Gao, and Q. Tian, "Affective visualization and retrieval for music video," *Multimedia, IEEE Transactions on,* vol. 12, pp. 510-522, 2010.

[18] C. Huang, T. Fu, and H. Chen, "Text□based video content classification for online video□sharing sites," *Journal of the American Society for Information Science and Technology,* vol. 61, pp. 891-906, 2010.

[19] M. E. Basiri, N. Ghasem-Aghaee, and A. R. Naghsh-Nilchi, "Exploiting reviewers' comment histories for sentiment analysis," *Journal of Information Science,* vol. 40, pp. 313-328, 2014.

[20] K. Ravi and V. Ravi, "A survey on opinion mining and sentiment analysis: Tasks, approaches and applications," *Knowledge-Based Systems,* vol. 89, pp. 14-46, 2015.

[21] L. Augustyniak, T. Kajdanowicz, P. Szymanski, W. Tuliglowicz, P. Kazienko, R. Alhajj*, et al.*, "Simpler is better? Lexicon-based ensemble sentiment classification beats supervised methods," in *Advances in Social Networks Analysis and Mining (ASONAM), 2014 IEEE/ACM International Conference on*, 2014, pp. 924-929.

[22] S. M. Mohammad and P. D. Turney, "NRC Emotion Lexicon," NRC Technical Report 2013.

[23] Y. Bi, "The impact of diversity on the accuracy of evidential classifier ensembles," *International Journal of Approximate Reasoning,* vol. 53, pp. 584-607, 2012.