

# Ranking highlight level of movie clips: A template based adaptive kernel SVM method

Zheng Wang<sup>a</sup>, Gaojun Ren<sup>a</sup>, Meijun Sun<sup>b</sup>, Jinchang Ren<sup>c</sup>, Jesse S. Jin<sup>a</sup>

wzheng@tju.edu.cn, gaojunren@tju.edu.cn, sunmeijun@tju.edu.cn, jinchang.ren@strath.ac.uk, jinsheng@tju.edu.cn

<sup>a</sup> School of Computer Software, Tianjin University, Tianjin, China

<sup>b</sup> School of Computer Science and Technology, Tianjin University, Tianjin, China

<sup>c</sup> Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, United Kingdom

**Abstract:** This paper looks into a new direction in movie clips analysis – model based ranking of highlight level. A movie clip, containing a short story, is composed of several continuous shots, which is much simpler than the whole movie. As a result, clip based analysis provides a feasible way for movie analysis and interpretation. In this paper, clip-based ranking of highlight level is proposed, where the challenging problem in detecting and recognizing events within clips is not required. Due to the lack of publicly available datasets, we firstly construct a database of movie clips, where each clip is associated with manually derived highlight level as ground truth. From each clip a number of effective visual cues are then extracted. To bridge the gap between low-level features and highlight level semantics, a holistic method of highlight ranking model is introduced. According to the distance between testing clips and selected templates, appropriate kernel function of support vector machine (SVM) is adaptively selected. Promising results are reported in automatic ranking of movie highlight levels.

**Keywords:** Video analysis; Highlight level; Movie clip; Template based method; Adaptive kernel SVM.

## 1. Introduction

Nowadays, we can easily access to thousands of new video or movie resources from the Internet. To draw eyes of movie consumers, conventional film directors need to put great efforts to produce movie trailers by picking up the most vivid and representative highlighting clips of a movie. The whole process is very challenging and trivial and has raised a question as whether we can automatically determine and extract clips. This is the question we attempt to address in this paper.

Generally, the existing affective models for video highlights analysis can be summarized into two categories [2]: i.e. categorical affective content analysis and dimensional affective content analysis. In categorical affective content analysis, emotions commonly belong to a few pre-defined basic categories, such as “fear”, “anger”, “sad” or “surprise” [3]. Dimensional affective content analysis applies the well-known psychological Arousal–Valence emotion space (A–V space) proposed by Hanjalic [4], which is characterized by the dimensions of arousal (intensity of affect) and valence (type of affect) providing a solid basis to represent the video affective content. The highlight level addressed in this paper is referred to the degree that draws audience's attention. It is labeling a video clip neither with one emotion type nor picking up the video highlights by choosing emotion spaces.

In general, considering the complicated structure and variations in rich content, related works in highlights extraction from generic movies remain rare. Due to the relative simple structure and clear semantics contained, highlights extraction from sports videos [14] and music videos have been intensively investigated [24,25]. Existing work has mainly focused on event-based approaches, where modeling from features to events is required [25]. Consequently, the difficulty in extracting highlights has been converted to another challenging problem, i.e. event detection and recognition.

In order to identify the highlights contents from movies, classification tool is employed to distinguish affective contents from others. Support vector machine (SVM) [1], as an efficient classification tool, has been widely applied in many research fields. David [8] has summarized several applications using SVMs in text categorization, computer vision, and bioinformatics. In addition,

SVM is also applied in the fields of medical diagnosis [9], financial engineering [10], and information processing [11].

Most of the research is mainly focused on single kernel function based SVM, which has inevitably limited its performance [5]. In Ref. [7], it is found that the performance of SVM is greatly affected by the choice of a kernel function. To overcome this problem, mixing or combining multiple kernels by certain mathematical operators instead of using a single one for higher accuracy of SVMs is introduced [6]. Lu [12] proposes to optimize the combined kernel function by particle swarm optimization (PSO) based on the large margin learning theory of SVM. However, there is few and sparse research investigating into incorporating with individual good performance of kernels of SVMs.

The novel contribution of this paper is the clip with user vote database and new conception of template based adaptive kernel SVM framework, which incorporates the good performance of individual kernels in SVMs without creating new combined kernels. The rest of this paper is organized as follows. Section 2 illustrates the overall framework of the proposed approach. Feature extraction and feature selection are presented in Sections 3 and 4, respectively. Section 5 discusses how the novel template based adaptive kernel SVM works. The user survey experiment, data set establishment, and the designed experiments are described in Section 6. Finally, some concluding remarks are drawn in Section 7 along with a brief discussion of possible future improvements.

## 2. Overall methodology

Fig. 1 illustrates the framework of proposed solution. With movie clips as input, the whole procedure is divided into three parts: clip with user vote database; visual cues extraction and template based adaptive kernel SVM classification.

### 2.1. Clip with user vote database

Many movies are manually segmented into clips to ensure each clip contains a relative complete story in several continuous shots. The movie split tool [26] is shown in Fig. 2. Each movie clip has been voted according to the audience's feeling whether the clip is exciting or not. The related votes are regarded as the ground truth to the associated movie clips.

### 2.2. Cues extraction and selection

The feature cues employed in this paper present the essential visual factors of movie clips and can be regarded as indicators between the highlight space and the low-level feature space. Audio events have strong hints to movie affective content and many works [3,13] have been done to analyze audio features. Although the results could be further improved with the audio analysis, this paper does not take audio features into account, but focus on the analysis of visual cues.

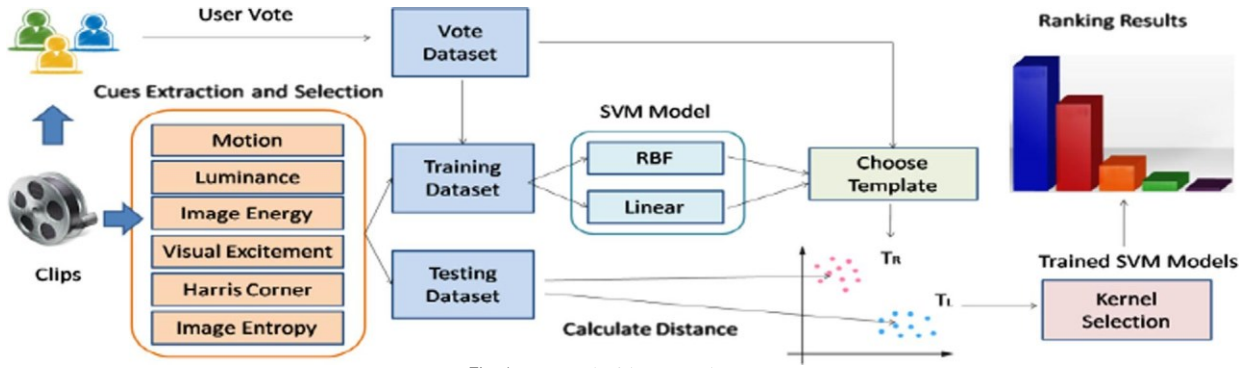


Fig. 1. Framework of the proposed system.

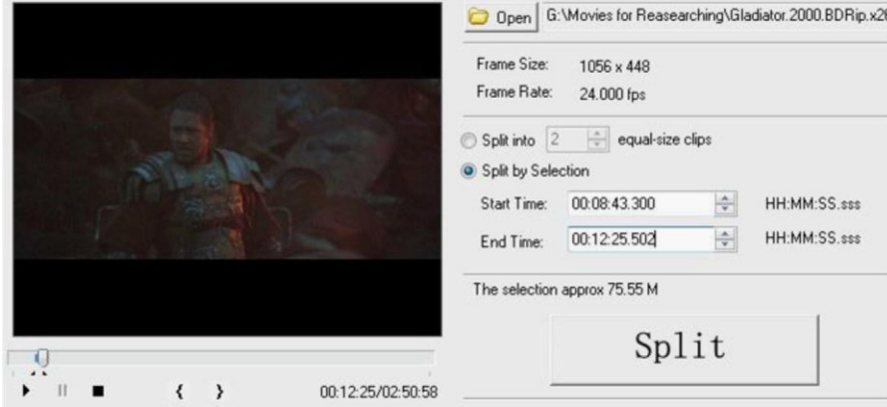


Fig. 2. Movie split interface: example of a movie to be split according to the set up start and end time.

The size difference of clips results in different length of feature cues which makes it difficult to train the model. A time warping method [23] is adopted in this paper to overcome this problem; Finally the principal component analysis method is applied for dimensionality reduction.

### 2.3. Template based adaptive kernel SVMs

To evaluate the highlight levels of clips, instead of using single kernel or creating a new kernel this paper proposes a template based adaptive kernel SVMs by applying single RBF kernel and single linear kernel SVM on the dataset, and select top 10 best result clips respectively as templates.

## 3. Cues extraction

As shown in Fig. 3, we extract six visual cues from movie clips and discuss in detail below.

### 3.1. Motion cue

Motions in a video include both the object motion and camera motion, which play important roles in highlights analysis [15]. Fig. 4 shows an example that a clip which are more attractive to users appear to have high peak value and tremendous changed curves (the blue) than the other (the red). Normally these movie clips with remarkable motion features are more attractive to viewers.

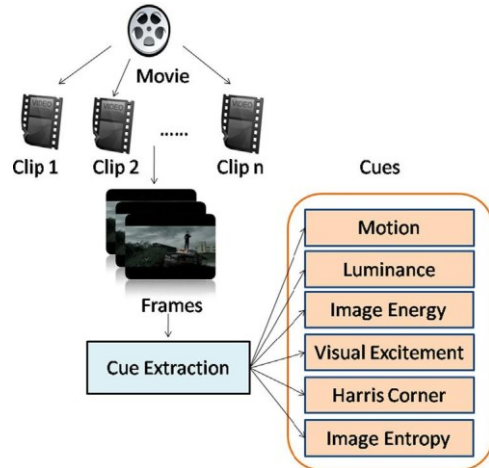


Fig. 3. Cues extraction.

Let the motion feature at frame  $k$  be  $m_{\delta k P}$  with the average magnitudes of  $N$  motion vectors in the frame with its width  $w$ . In the concept of optical flow, motion vectors are processed in 8 directions which are divided by  $2\pi$ .

$$m(k) = \sum_{i=1}^N \frac{|\vec{v}_i| \cdot |p_{ix} - q_{ix}|}{8 \cdot w \cdot N} \quad (1)$$

where  $|\vec{v}_i|$  denote the length of the  $i$ th optical flow vector and  $p_{ix}, q_{ix}$  are  $x$  values of two points in  $v_i$ .

### 3.2. Luminance cue

From the cinematographic perspective, lighting techniques play

an extremely important role in movies. Generally two major esthetic lighting techniques are frequently employed [16]: low-key lighting and high-key lighting. Low-key lighting, or chiaroscuro lighting, characterized by a contrast between light and shadow areas, is quite usual in scenes like horror movies to driven atmosphere. On the contrary high-key lighting, or flat lighting, deemphasizing the light/dark contrast, is usually used to express the atmosphere of cheerful, warm or magnificent scenes.

Luminance cue simulates the lighting in cinematography well. For instance, the luminance curve waves left up and down when the low-key lighting is applied in a movie clip and the luminance curve waves smoothly when the high-key lighting is applied. In Fig. 5 the clip with high level of highlight shows a curve (the blue) with high peak value and drastic changes. On the contrary, the curve of a low voted clip is smoother (the red).

Luminance feature  $l(k)$  at frame  $k$  is obtained in the hue, saturation and intensity (HSI) color space.

$$l(k) = \frac{h+s+i}{3} \quad (2)$$

The parameters  $h$ ;  $s$  and  $i$  are the average value of hue, saturation and intensity, respectively.

### 3.3. Image energy

Image energy describes the distribution of the gray levels in the image and texture coarseness, which in some way reflects the style of the video frame, like cartoon style with simple texture or realistically styles with complex texture. Movie clips are often more attractive to viewers than others, if they are full of beautiful scenery and inviting views and whose average image energy is lower than others. As verified in Fig. 6, the image energy of high voted movie clips (the blue) is lower than a less voted clip (the red). The image energy introduced in this paper is obtained by computing the sum of the squares of each element in gray-level co-occurrence matrix (GLCM) pro-posed by Haralick [18] in the 1970s.

### 3.4. Visual excitement

This feature is proposed by Wang [3] which reflecting the relation between the low-level feature and visual excitement. As a measure of visual excitement, this feature works for clips with different highlight levels.

Fig. 7 shows that the movie clip with higher votes have higher visual excitement value in average (the blue) than those with lower votes (the red), which can be concluded both from the time series curves and the histogram.

### 3.5. Harris Corner

Normally, movie clips with rich content can be regarded as a scene with more interest points than others; these clips usually attract the eyes of the viewer's more easily. The Harris Corner detection [19] is suitable for such a case. Fig. 8 shows that clips with higher votes have an average higher number of interest points (the blue) than those with lower votes (the red).

### 3.6. Image entropy

Two dimensional image entropy is a measurement of the disorder or randomness in video frame [17]. It describes the spatial characteristics of the image gray. Movie clips with some abnormal activities attract users' attention than those with normal activities as shown in Fig. 9 that the movie clips with high votes have higher image entropy (the blue). 2D image entropy  $e(k)$  at frame  $k$  with width  $w$  and height  $h$  can be described by

$$e(k) = - \sum_{i=0}^{255} \sum_{j=0}^{255} P_{ij} \ln P_{ij} \quad (3)$$

$$P_{ij} = \frac{f(i,j)}{wh} \quad (4)$$

The  $f(i,j)$  is the number of pairs denoted by  $(i,j)$ , which represents the pixel's gray value and its neighborhood gray value are  $i$ ;  $j$  respectively.

### 4. Feature selection

This section elaborates the feature selection stage of the proposed framework. The purpose of this stage is dimensionally reduction and feature length aligning.

Given an input movie clip  $V$ , we denote each of these features by  $f_i$  and the feature set by  $F$  below, where  $k$  is the number of feature cues and equals six in this paper.

$$F = \{f_i; i = 1, 2, \dots, k\} \quad (5)$$

Due to different numbers of frames contained in these clips, the size of the raw feature  $f_i$  acquired directly from a movie clip differs from each other. This requires the features to be resized into the same length. The feature selection method is formulated in formula (6)

$$F' = \text{PCA} (T(f_1), T(f_2), \dots, T(f_k)) \quad (6)$$

The  $F'$  is the final integrated feature vector after selection, the function  $\text{PCA}$  [21] and  $T$  represent the principal components analysis and time warping method, respectively. This proposed method converts different size of feature series to a given size and reduces data redundancy.

Fig. 10 shows how the time warping method with hyperbolic tangent curve works. For instance, consider a 25 s clip with 1000 frames, mapping the 1000 dimensions feature to the  $x$  axis. The 1000 dimensions range from  $-500$  to  $500$  on the  $x$  axis will be mapped to a result ranges from  $-25$  to  $25$  on the  $y$  axis. This mapping keeps the details of the middle part dimensions and reduces the noised of both ends at the same time, which is suitable for movie clips since the middle part of a movie clip is more likely to have the effective content.

Each feature time series  $f_i$  is warped to the size of 50, and then we obtain a total 300 dimension integrated feature of every clip. Finally a PCA process is applied to get a 100 dimensions integrated feature for the following stage.

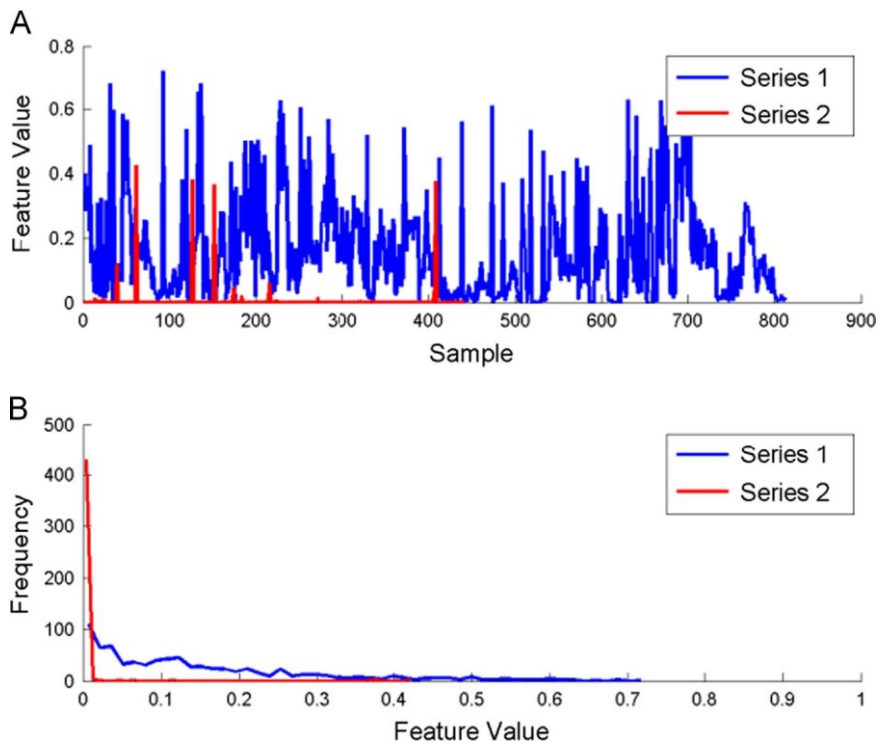


Fig. 4. (A) Motion features time curve of two clips with different number of user votes and (B) histogram of feature values. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

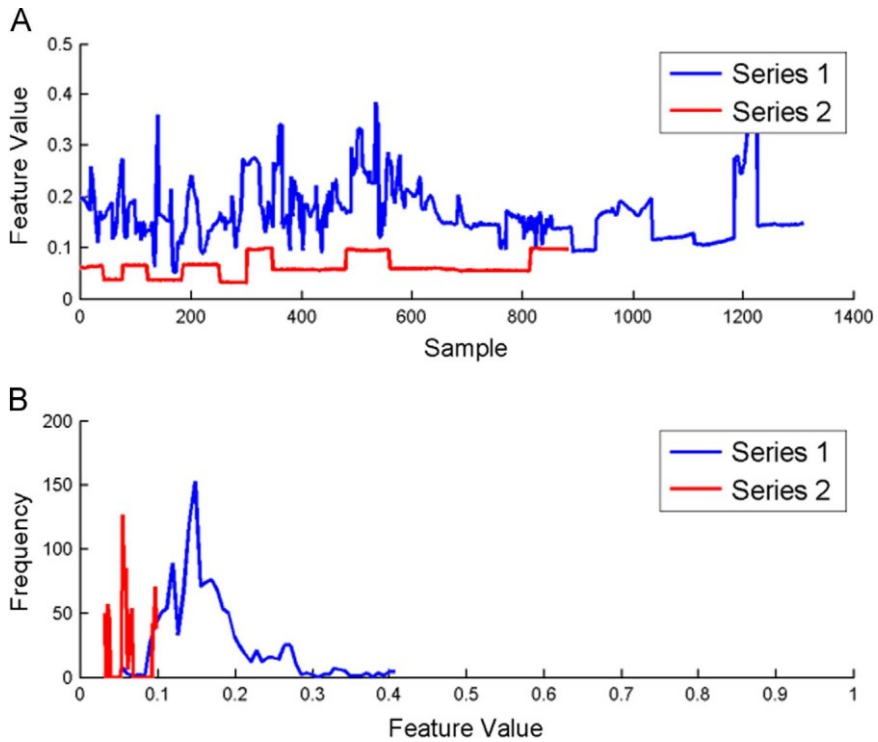


Fig. 5. (A) Luminance features time curve of two clips with different number of user votes and (B) histogram of feature values. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

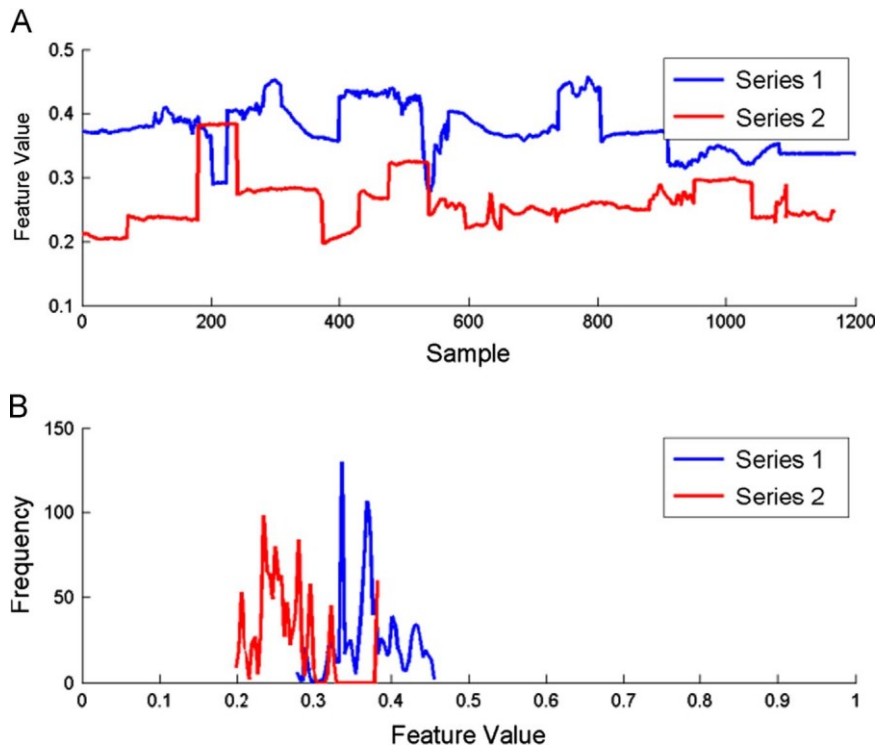


Fig. 6. (A) Energy features time curve of two clips with different number of user votes and (B) histogram of feature values. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

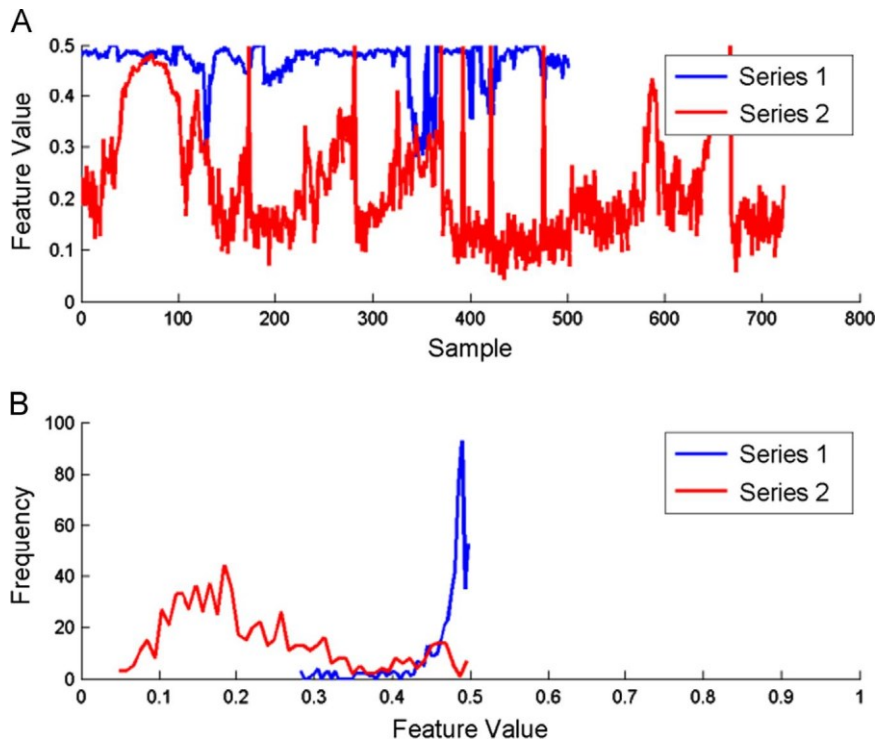


Fig. 7. (A) Visual excitement features time curve of two clips with different number of user votes and (B) histogram of feature values. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

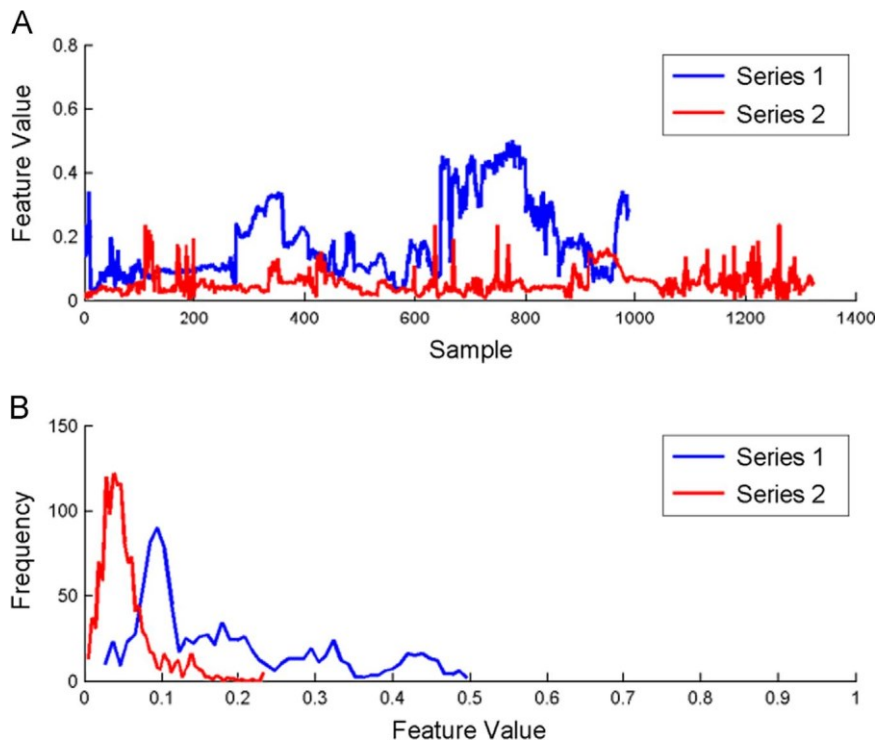


Fig. 8. (A) Harris Corner features time curve of two clips with different number of user votes and (B) Histogram of feature values. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

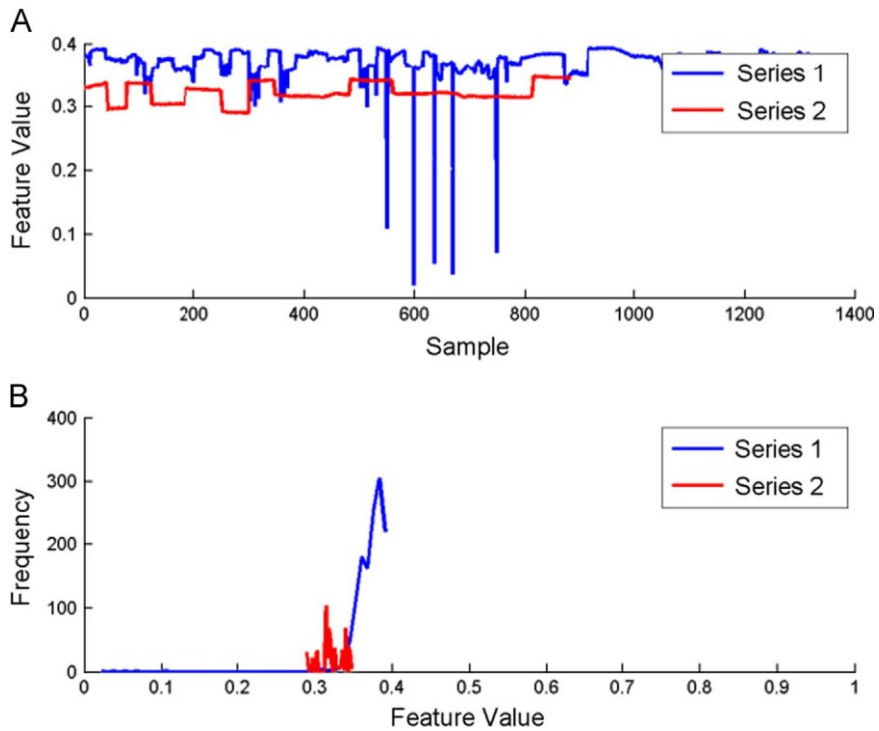


Fig. 9. (A) Image entropy features time curve of two clips with different number of user votes and (B) histogram of feature values. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

### 5. Template based adaptive kernel SVM

In the database, we have obtained the votes  $v$  for each video clip, which stands for the evaluation value of affection, and the relationship between the low-level feature and the affection value is given as follows:

$$W \propto F^0; v$$

677

where  $F^0$  is the integrated feature from Eq. (6). With the relationship  $W$ , we propose SVM based framework which adaptively select RBF or linear kernel based on the comparison results between testing clip and templates, as shown in Fig. 11.

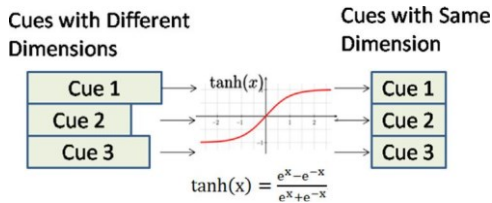


Fig. 10. Time warping method with the hyperbolic tangent curve.

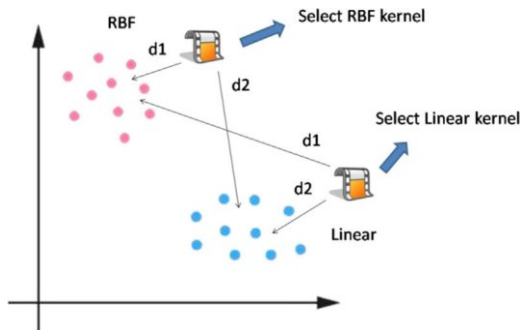


Fig. 11. Template based adaptive kernel SVM.

Table 1 Movies in the database.

No.	Movie names	Genres
1	Red (2010)	Action/Comedy
2	Mission: Impossible III (2006)	Action/Thriller
3	Live Free or Die Hard (2007)	Action/Thriller
4	Gladiator (2000)	Action/Drama
5	Terminator 2: Judgment Day (1991)	Action/Sci-Fi
6	The Day After Tomorrow (2004)	Disaster/Sci-Fi
7	Twister (1996)	Disaster/Drama
8	Titanic (1997)	Disaster/Romance
9	The Perfect Storm (2000)	Disaster/Drama
10	2012 (2009)	Disaster/Family
11	The Silence of the Lambs (1991)	Horror/Crime
12	Silent Hill (2006)	Horror/Thriller
13	The Shining (1997)	Horror
14	Final Destination (2000)	Horror/Thriller
15	Alien: Resurrection (1997)	Horror/Sci-Fi
16	Black Hawk Down (2001)	War/History
17	The Lord of the Rings: The Return of the King (2003)	War/Action
18	Troy (2004)	War/Romance
19	Brave Heart (1995)	War/History
20	Saving Private Ryan (1998)	War/History

The whole method can be described as follows:

- 1) Prepare training and testing dataset from  $W$ .
- 2) Training two SVM model with linear and RBF kernel, denoted as  $M_L$  and  $M_R$ , respectively.

Table 2

Comparison results of three methods.

Error range	Linear		RBF		Proposed	
	Clips amount	Average error	Clips amount	Average error	Clips amount	Average error
(0, 10]	80	4.509	86	5.033	114	4.365
(10, 20]	49	15.137	52	14.559	38	14.756
(20, 30]	27	24.537	30	23.165	17	22.923
(30, 40]	9	33.310	1	31.371	0	—
(40, 50]	2	41.472	1	41.378	1	40.014
(50, 60]	2	55.796	0	—	0	—
(60, 70]	0	—	0	—	0	—
(70, 80]	1	86.172	0	—	0	—
Total	170	13.797	170	11.514	170	8.753

- (3) Based on the testing results, select top 10 clips' features set for each model and take them as templates sets, denoted as  $T_L$  and  $T_R$ .
- (4) For a testing clip feature  $F^0$ , calculate the distance  $d_L$  and  $d_R$  to the two templates sets respectively. Then  $d_L \frac{1}{4}$  Distance  $\delta F^0; T_L \mathcal{P}$  and  $d_R \frac{1}{4}$  Distance  $\delta F^0; T_R \mathcal{P}$ , where the Distance() function is the Hamming distance method [22].
- (5) If  $\delta d_L < 4d_R \mathcal{P}$ , select  $M_L$  as predicting model for the clip; else select  $M_R$ .

## 6. Experiments

This section introduces some experiments to evaluate the proposed framework and show the advantage of the template based adaptive kernel SVM method. For the convenience of comparative assessment, a user experiment was conducted to establish the ground truth which each movie clip is labeled with user votes as the measurement of highlight level.

### 6.1. Proposed database

The user experiment takes place in a room with about 50 computers in two days. The audiences are 70 undergraduate students with 30 females and 40 males who have received detailed instructions of how the experiment is conducted.

During the experiment, each user is asked to watch these 350 movie clips on a computer screen with a noise canceling headphone. When a user has watched a movie clip, he/she is asked to vote for those clips on a given questionnaire, based on whether he/she enjoys it. We make statistics and analysis according to the results of the questionnaire. As a result, votes of movie clips range from 0 (nobody votes for it) to 70 (everybody votes for it).

The clips mentioned above are manually chosen from a total of 20 movies listed in Table 1 with four major movie genres including action, horror, war and disaster. Each genre contains 5–6 movies and there are about 15–18 clips are chosen from each movie. The movie clips varies from a minimum of 40 s to a maximum of 2 min in length. Fig. 12 shows the key frames of these movies.

The criterion of choosing clip is selecting those clips whose highlight level rang from low to high averagely. Finally, there are 596 clips. Fig. 13 shows the sorted user vote's histogram, it can be observed that the database contains clips with different highlight levels. The clip database will be published very soon.

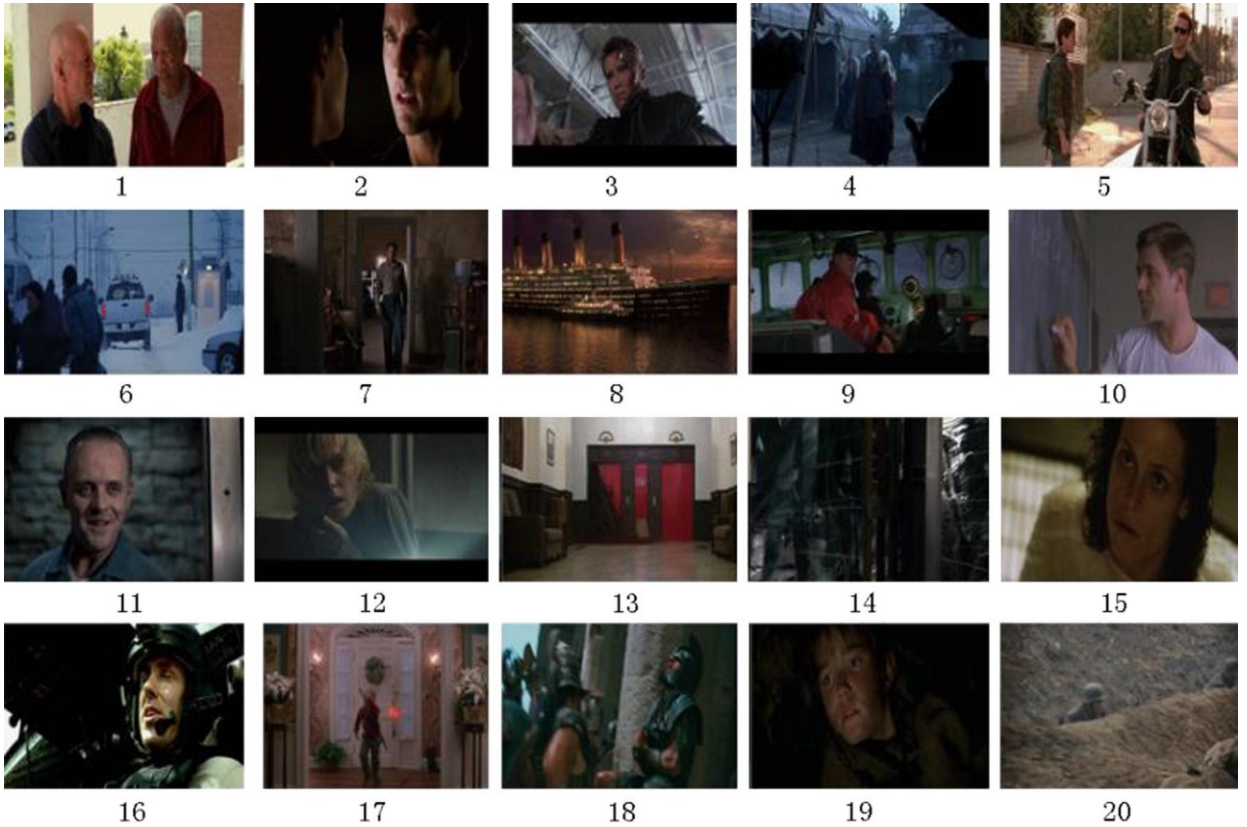


Fig. 12. Key-frames from the testing database.

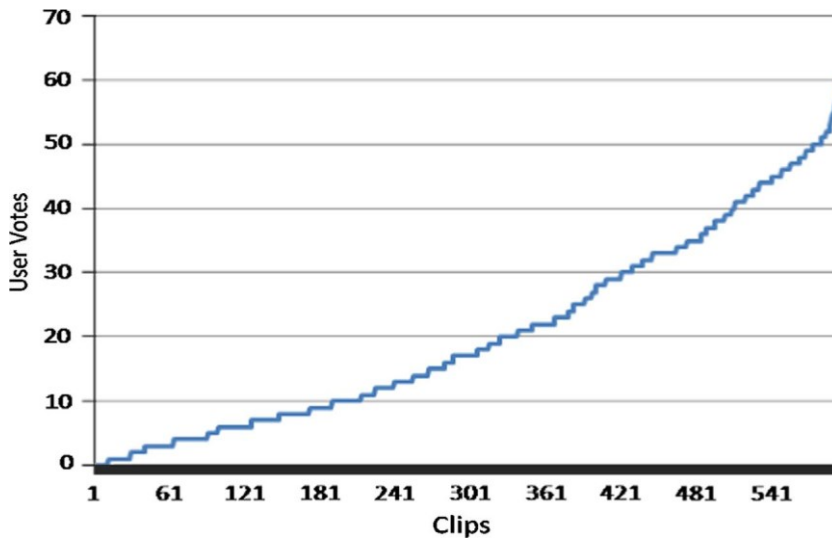


Fig. 13. Statistic histogram of user votes which have been sorted.

## 6.2. Experiments on the proposed database

To show the advantage of the proposed method in this paper, experiments with single kernel SVMs and proposed method are compared.

To train the single kernel SVM, 426 clips are selected randomly from database, half of them are training data and others are testing data. According to the results, the top 10 best result clips are selected as templates. Then the other 170 clips are tested by using

the RBF kernel SVM, linear kernel SVM and the proposed method. Table 2 shows the statistic error results of linear kernel SVM, RBF kernel SVM and proposed method. The average error between predicted value and the ground truth is employed as evaluation criterion.

As shown in Table 2, it is obvious to see that proposed performs better than linear kernel and RBF kernel SVM in most cases. When the error value lies in the ranges from 0 to 20, the results shows that the proposed method result contains more clips than the other two



methods, which indicates the higher performance of proposed method.

A snapshot of the results comparison between the ground truth and the proposed method is shown in Fig. 14. The result of the

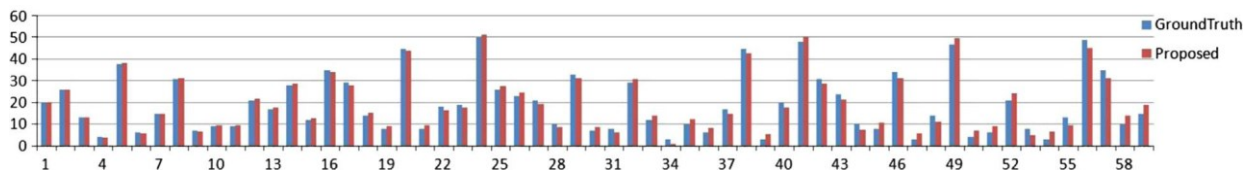


Fig. 14. Result comparison between ground truth and the proposed method

### 6.3. Other case study

To demonstrate the framework proposed in this paper, half part (about 55 min in length) of the movie *Wanted* (2008) with a total of 110 min in length is recruited as an example of movie highlights summarization.

First of all, the selected movie is sliced into clips and re-connected by the shot boundary detection method [20]. Those movie clips whose length is less than 3 s are neglected automatically. Second, the mentioned integrated feature cues are extracted from each of the clips. Finally, for each feature vector, the predicted result is obtained with the template based adaptive kernel SVM model. The top 10 clips with highest predicted results make up the final movie highlights summarization, as shown in Table 3.

Table 3  
Movie trailer with 53 s length containing 10 shots generated for the movie *Wanted* (2008).

No.	Snap Shot	Description
1		A Gun is flying from a man towards the hero.
2		The hero is running after a man in the train station.
3		The hero is trying to calm down the scared passengers. The scene is chaotic as people shout loudly.
4		A train is running through a gorge and suddenly the train stops and falls down from pathway.
5		The hero is seeking for his enemy in a train with a gun in his hand. Passengers are scared.
6		The hero is firing around with a man's body as shield. Many enemies die.
7		A train is running through a city with beautiful night urban landscape.
8		A man is driving a car trying to catch up the train, shooting to the train at the time.
9		Someone gets shot and falls into the lake, bleeding.
10		The hero gets back to the base for a final battle.

## 7. Conclusions and future work

In this paper, we built a clip database with user votes for their highlight level, and then a novel template based adaptive

proposed method is very close to the ground truth.

kernel SVM framework is introduced to ranking movie clip highlight level by using six visual feature cues. These features are combined together via time warping and PCA to represent the characteristics of a given movie clip. Experimental results have demonstrated that the proposed framework is effective in ranking movie clips and producing movie highlight recommendation.

Further studies will be conducted to enlarge the proposed clip database and improve the performance of proposed algorithm by using deep learning method.

## Acknowledgment

The authors wish to acknowledge the support from the National Natural Science Foundation of China, under the Grants 61003201 and 61202165, and a joint project funded by the Royal Society of Edinburgh and NSFC 61211130125.

## References

- [1] K.-R. Muller, S. Mika, G. Ratsch, K. Tsuda, B. Scholkopf, An introduction to kernel-based learning algorithms, *IEEE Trans. Neural Netw.* 12 (2) (2001) 181–202.
- [2] S. Zhang, Q. Huang, S. Jiang, W. Gao, Q. Tian, Affective visualization and retrieval for music video, *IEEE Trans. Multimed.* 12 (6) (2010) 510–522.
- [3] H.L. Wang, L.F. Cheong, Affective understanding in film, *IEEE Trans. Circuits Syst. Video Technol.* 16 (6) (2006) 689–704.
- [4] A. Hanjalic, L.Q. Xu, Affective video content representation and modeling, *IEEE Trans. Multimed.* 7 (1) (2005) 143–154.
- [5] O. Chapelle, V.N. Vapnik, Choosing multiple parameters for support vector machines, *Mach. Learn.* 46 (1–3) (2002) 131–159.
- [6] Mehmet Gönen, Ethem Alpaydm, Multiple kernel learning algorithms, *J. Mach. Learn. Res.* 12 (2011) 2211–2268.
- [7] M. Hussain, S.K. Wajid, A. Elzaat, M. Berbar, A comparison of SVM kernel functions for breast cancer detection, in: *Proceedings of the Eighth International Conference on Computer Graphics, Imaging and Visualization (CGIV)*, 2011, pp. 145–150.
- [8] V. David, A. Sánchez, Advanced support vector machines and kernel methods, *Neurocomputing* 55 (1–2) (2003) 5–20.
- [9] D. Conforti, R. Guido, Kernel based support vector machine via semidefinite programming: application to medical diagnosis, *Comput. Oper. Res.* 37 (2010) 1389–1394.
- [10] T. Van Gestel, et al., Financial time series prediction using least squares support vector machines within the evidence framework, *IEEE Trans. Neural Netw.* 12 (4) (2001) 809–821. (Special Issue on Neural Networks in Financial Engineering).
- [11] V. Vapnik, et al., Support vector method for function approximation, regression estimation, and signal processing, *The Advances in Neural Information Processing Systems*, MIT Press, Cambridge, MA, 1997, 281–287.
- [12] Ming-Zhu Lu, C.L.P. Chen, Jian-Bing Huo, Optimization of combined kernel function for SVM by particle swarm optimization, in: *Proceedings of the 2009 International Conference on Machine Learning and Cybernetics*, 12–15 July 2009.
- [13] M. Xu, L.-T. Chia, J. Jin, Affective content analysis in comedy and horror videos by audio emotional event detection, in: *Proceedings of the IEEE International Conference on Multimedia and Expo*, vol. 61 (2), July 2005, pp. 2–5.
- [14] A. Ekin, A.M. Tekalp, R. Mehrotra, Automatic soccer video analysis and summarization, *IEEE Trans. Image Process.* 12 (7) (2003) 796–807.
- [15] J.W. Hsieh, S.L. Yu, Y.S. Chen, Motion-based video retrieval by trajectory matching, *IEEE Trans. Circuits Syst. Video Technol.* 16 (3) (2006) 396–409.

- [16] D. Bordwell, K. Thompson, *Film Art: An Introduction*, 7th ed. McGraw-Hill, New York, 2004.
- [17] Md. Haidar Sharif, Chabane Djeraba, An entropy approach for abnormal activities detection in video streams, *Pattern Recognit.* 45 (7) (2012) 2543–2561.
- [18] R.M. Haralick, K. Shanmugam, I. Dinstein, Textural features for image classification, *IEEE Trans. Syst. Man Cybern. SMC-3* (6) (1973) 610–621.
- [19] Li Yi-bo, Li Jun-Jun, Harris Corner detection algorithm based on improved contourlet transform, *Procedia Eng.* 15 (2011) 2239–2243.
- [20] Jinhui Yuan, Huiyi Wang, et al., A formal study of shot boundary detection, *IEEE Trans. Circuits Syst. Video Technol.* 17 (2) (2007) 168–186.
- [21] E. Sahouria, A. Zakhor, Content analysis of video using principal components, *IEEE Trans. Circuits Syst. Video Technol.* 9 (8) (1999) 1290–1298.
- [22] Lei Zhang, Yongdong Zhang, et al., Binary code ranking with weighted Hamming distance, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2013, pp. 1586–1593.
- [23] F.K.-P. Chan, Haar wavelets for efficient similarity search of time-series: with and without time warping, *IEEE Trans. Knowl. Data Eng.* 15 (3) (2003) 686–705.
- [24] Shiliang Zhang, Qingming Huang, Shuqiang Jiang, Wen Gao, Qi Tian, Affective visualization and retrieval for music video, *IEEE Trans. Multimed.* 12 (6) (2010) 510–522.
- [25] Guangyu Zhu, Changsheng Xu, Event tactic analysis based on broadcast sports video, *IEEE Trans. Multimed.* (2009) 49–67.
- [26] (<http://www.boilsoft.com>).