MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

Subjective Evaluation Criterion for Selecting Affective Features and Modeling Highlights

Liyuan Xing, Hua Yu, Qingming Huang, Ajay Divakaran

TR2006-027 January 2006

Abstract

In this paper, we propose a subjective evaluation criterion which is a guide for selecting affective features and modeling highlights. After the database of highlight ground truth is established and commonly used effective features are extracted, evaluation experiments are designed on tennis and table tennis as examples. Based on the experiments, we conclude that: 1) the commonly used affective features are correlative; 2) the effective combination of affective features is motion vector (MV) average, cheer duration, excited speech duration and event duration; 3) the highlights model is approximately linear.

SPIE Conference Multimedia Content Analysis, Management and Retrieval

Subjective Evaluation Criterion for Selecting Affective Features and Modeling Highlights

Liyuan Xing [a*], Hua Yu [a], Qingming Huang [a], Ajay Divakaran [b]
[a] Graduate School of the Chinese Academy of Sciences, Yuquan Road 19A, Beijing
[b] Mitsubishi Electric Research Laboratories, Cambridge, MA 02139

ABSTRACT

In this paper, we propose a subjective evaluation criterion which is a guide for selecting affective features and modeling highlights. After the database of highlight ground truth is established and commonly used affective features are extracted, evaluation experiments are designed on tennis and table tennis as examples. Based on the experiments, we conclude that: 1) the commonly used affective features are correlative; 2) the effective combination of affective features is motion vector (*MV*) average, *cheer* duration, *excited speech* duration and *event* duration; 3) the highlights model is approximately linear.

Keywords: Subjective Evaluation Criterion, Affective Features, Highlights, Modeling

1. INTRODUCTION

Nowadays, an increasing amount of digitized sports video is produced day by day. More and more people access sports video using TV set-top box, PC, PDA, or even mobile phone. It is important to extract the valuable content in a sports video for saving both the user's time and downloading costs. Therefore sports video analysis using highlights detection for content summarization has been a hot research topic in recent years. Many research on highlights detection so far, but uniform evaluation criterion is absent.

Highlights in sports video is usually classified into two categories, that is, event oriented [1,2] and exciting degree oriented [3,4]. For the former highlights is defined as fixed exciting events such as "home run" events in baseball and "score" events in tennis, which are evaluated by the precision and recall. There is scarcity of exciting degree of each event. Sometimes it is not easy to decide whether one type of the event is more exciting than the other. For the latter there is no criterion. And researchers usually evaluate the experimental results by observing the local maxima of the highlights curve to see whether it is an exciting event or not.

In this paper, we propose a subjective evaluation criterion, which can be used both in event oriented and exciting degree oriented highlights detection systems. Because any highlights calculated by computer is compared with that marked by human, and we can know how much extent the highlights reflects human perception. In order to make the subjective evaluation criterion works, the first thing is to establish the database of highlight ground truth. Highlight rank marked by more than 6 people can guarantee it is the ground truth. Secondly, affective features are needed to express the highlight, so commonly used affective features are extracted. If these two steps are done, you can use the proposed subjective evaluation criterion to selecting affective and modeling highlights. In the following, we will provide a detailed description of the process by taking tennis and table tennis as examples.

The rest of the paper is organized as follows. In section 2, we select several commonly used affective features. In section 3 a subjective evaluation criteria is proposed. Evaluation experiments are designed in section 4, we can conclude

[*] Further author information: send correspondence to Liyuan Xing, Qingming Huang, email: {lyxing, qmhuang}@jdl.ac.cn; Hua Yu, email: yuh@gscas.ac.cn; Ajay Divakaran, email: ajayd@merl.com

that these affective features are correlative and effective combination of them can be found and the highlight model is approximately linear. Section 5 concludes the paper.

2. AFFECTIVE FEATURES

In [3,4,6], the commonly used affective visual feature is motion. It captures the pace of action [4] and shows a significant impact on individual affective response [6]. The commonly used affective audio features are cheers and pitch-related features [3,4,6], which are considered as having strong relationship to the affective content of the video. For example, in sports video, *cheer* of the audience and the pitch improvement of the commentator imply highlight scenes. The longer duration of *cheer* and higher average energy of *cheer*, as well as the longer duration of *excited speech* and higher average pitch of *excited speech* are, the more exciting the event is. Finally, *event* duration is another affective feature that has relation to affective content of video [6].

By these observations, we extract a total of 6 affective features on each event for highlight ranking. They are

a. *MV* average
b. *cheer* duration
c. *cheer* average energy
d. *excited speech* duration
e. *excited speech* average pitch
f. *event* duration

3. SUBJECTIVE EVALUATION CRITERION

It is believed that good evaluation criterion is a guide for the right development direction of the thing. Here, the subjective evaluation criterion is proposed for selecting affective features and modeling highlight.

Firstly, we define the highlight rank of each event as an integer r which is between 0 and R. Supposing that the total segmented event number is M, six people are asked to endow each of the event m a highlights rank $H_m(r)$ in terms of their subjective perception on the event. By these measures, the ground truth for highlight evaluation is established. Based on this ground truth, we can evaluate the highlights ranking result $C_m(r)$ by computer as

$$\text{Affective accuracy} = \frac{1}{M}\sum_{m=0}^{M}\frac{R-\left|H_m(r)-C_m(r)\right|}{R} = \frac{1}{M}\sum_{m=0}^{M}1-\left|H_m^{'}(r)-\mathrm{C}_m^{'}(r)\right| \quad (1)$$

where $\left|H_m^{'}(r)-\mathrm{C}_m^{'}(r)\right|$ represents the relative bias between highlight ranked by human and computer, so the change of R which is selected according to users' requirement will not affect the accuracy. The difference of 1% in accuracy means a difference of 1% in relative bias. If the accuracy is 80%, there is 20% difference between human rank and computer rank relatively. Function (1) shows that the accuracy is obtained by averaging the human-computer rank bias.

4. EVALUATION EXPERIMENTS DESIGN

Experimental data are composed of 4 different broadcast tennis video of French Open 2005 and 4 different broadcast table tennis video of Athens Olympic 2004. These videos are compressed by MPEG-2, digitized at 25 frames/s, and have a resolution of 352×288. The audio signal is sampled at 44100Hz and 16 bits/sample. The details of each video are listed in Table 1 and Table 2.

Table 1: Tennis videos

video	e	f	g	h
whole duration	29:27	26:53	9:40	26:32
total event	83	79	30	98

Table 2: Table tennis videos

video	E	F	G	H
whole duration	27:40	33:57	11:08	34:46
total event	82	61	21	70

In general, the signal system in real world is often non-linear, so the human perception system is. Then we wondered whether the linear highlight model is simple and effective. So we use a nonlinear model (SVM regression) to see whether the affective features and the highlight exciting degree are approximately linear. The first reason we adopt SVM regression to rank the events is that it has the advantages of kernel-based learning method, such as requiring fewer training samples and having better generalization ability. The second reason is that SVM regression provides superior robustness and prediction accuracy for sparse and nonlinear data distribution [7,8]. In our application, the ground truth is sparse since the subjectivity of different person.

4.1 Selecting affective feature

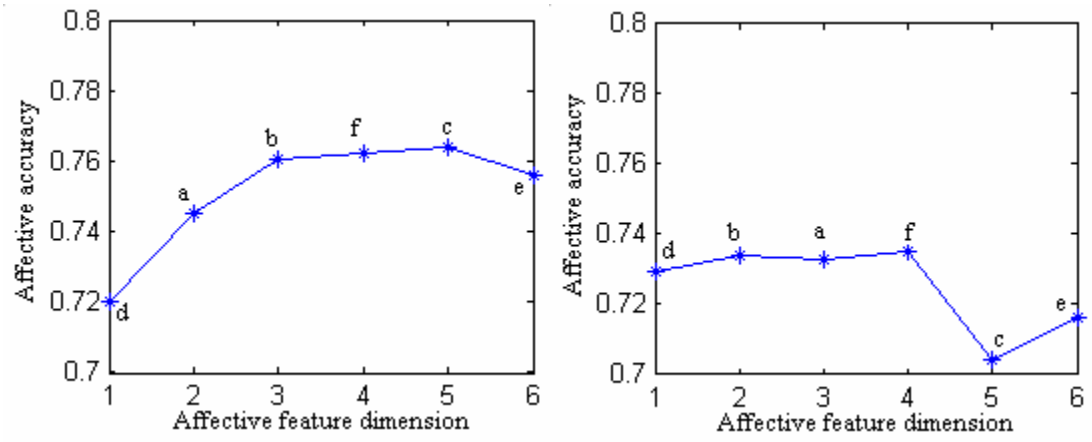
Based on the proposed evaluation criterion (1), we use a forward search algorithm [9] to evaluate the affective features. And the SVM cross validation is performed with three sets of randomly selected training data and testing data. Kernel function for SVM is RBF (Radial Basis Function) with parameter gamma=1/dim. The value of R in function (1) is 10.

Figure 1 shows the accuracy on different feature number. As shown in Figure1, the differences of minimum and maximum are 4% in tennis and 3% in table tennis, respectively. It means that these affective features are quite correlative. As shown in Figure2, one feature alone is able to reflect the exciting degree to large extent. Furthermore, feature **d** is the domain feature in Figure2, but the combination of **a**, **b**, **d** and **f** gets maximum affective accuracy in Figure1.

Based on this experiment, we can have the conclusions as follows

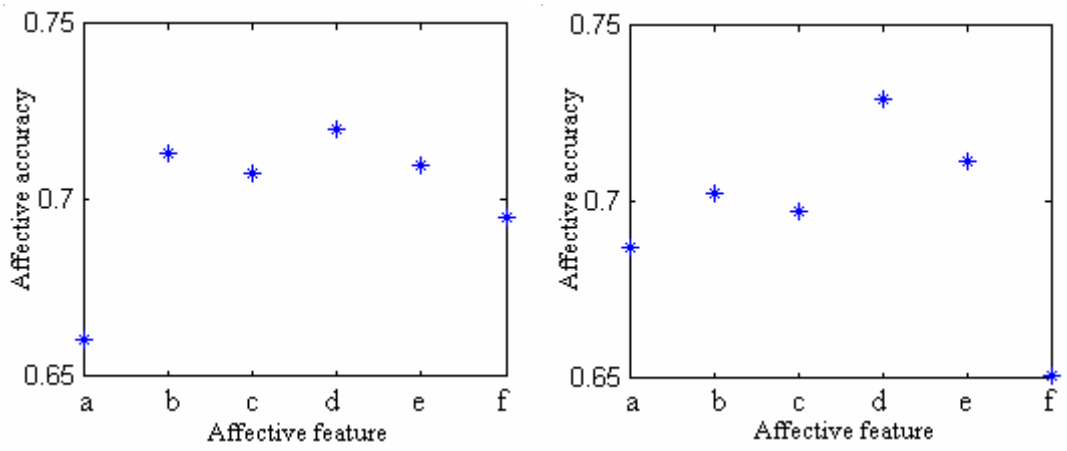
Conclusion 1: The commonly used affective features are correlative;

Conclusion 2: The combination of a, b, d and f is effective;



(a: tennis) (b: table tennis)

Figure 1: The affective feature selection process



(a: tennis) (b: table tennis)

Figure 2: Single affective feature

4.2 Modeling highlight

With the affective features **a**, **b**, **d** and **f** are fed into the regression model, the comparison results of nonlinear and linear regression are listed in Table 3 and Table 4. It can be seen that there is nearly no improvement by using nonlinear regression (SVM regression), so the past work using the linear highlight model is reasonable, which means that it is simple and effective. It also can be seen that the affective accuracy reaches around 80.0% in terms of the ground truth and evaluation criteria. We must make it clear that 82.0% (79.3%) affective accuracy is a marvelous highlights ranking result since the result is obtained fully automatically by computer. This result shows that the determined affective features can reflect human perception to a large extent. Furthermore, it shows that in some special conditions computer can learn from human perception for automatic video content understanding.

Based on this experiment, we can have the conclusions as follows

Conclusion 3: The highlight model is approximately linear;

Table 3: Tennis videos

Train data	Test data	Affective accuracy (%)	
		SVM regression	Linear regression
e	f,g,h	83.3	81.8
h	e,f,g	82.5	83.4
e,f	g,h	79.5	79.0
g,h	e,f	84.2	83.7
average		**82.4**	**82.0**

Table 4: Table tennis video

Train data	Test data	Affective accuracy (%)	
		SVM regression	Linear regression
E	F,G,H	79.3	77.9
H	E,F,G	78.9	77.0
E,F	G,H	85.2	83.1
G,H	E,F	80.6	79.0
average		**81.0**	**79.3**

5. CONCLUSIONS

Our proposed subjective evaluation criterion is a guide for highlight detecting and ranking. It proves that past works on affective features and highlight model are effective. Although the experiments have only been carried on the tennis and table tennis, it is believed that the subjective evaluation criterion is the same with other sports, so the affective features and linear highlight model are.

6. REFERENCES

1. Yihong Gong , Mei Han , Wei Hua , Wei Xu, “Maximum entropy model-based baseball highlight detection and classification,” Computer Vision and Image Understanding, v.96 n.2, p.181-199, November 2004
2. M. Xu, LY. Duan, CS. Xu, and Q. Tian, “A fusion scheme of. visual and auditory modalities for event detection in sports video,” In Proc. of ICASSP, 2003.
3. Alan Hanjalic, “Generic approach to highlights extraction from a sports video,” IEEE ICIP03
4. Zixiang Xiong, Regunathan Radhakrishnan, Ajay Divakaran, “Generation of sports highlights using motion activity in combination with a common audio feature extraction framework,” ICIP 2003.
5. Radhakrishan, R.; Xiong, Z.; Divakaran, A.; Ishikawa, Y, “Generation of Sports Highlights Using a Combination of Supervised & Unsupervised Learning in Audio Domain,” International Conference on Pacific Rim Conference on Multimedia, Vol. 2, pp. 935-939, December 2003
6. A. Hanjalic, L.-Q. Xu, “Affective Video Content Representation and Modeling,” IEEE Transactions on Multimedia, February, 2005.
7. V. Cherkassky and Y. Ma, “Selecting of the loss function for robust linear regression,” Neural computation, 2002.
8. Scholkopf, B., Burges, C., and Smola, “Advances in kernel methods: support vector machine,” MIT Press, 1999.
9. A.K. Jain, “Statistical pattern recognition: a review,” IEEE Trans. On PAMI, 2, 4-37, 2001.