Highlight Ranking for Sports Video Browsing

Xiaofeng Tong, Qingshan Liu, Yifan Zhang, Hanqing Lu
National Lab of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences
P.O.Box 2728, Beijing, China, 100080
{xftong, qsliu, yfzhang, luhq}@nlpr.ia.ac.cn

ABSTRACT
Sports video has been extensively studied for its wide viewer-ship and tremendous commercial potentials. Many studies focused on highlight extraction for summarizing a lengthy video. In this paper, we present an advanced highlight analysis system for sports video browsing, in which highlight evaluation and ranking are concerned besides highlight detection. First, we use replay detection to efficiently localize the highlights. Then incorporating with domain-specific knowledge, we adopt several significant cues to evaluate the importance degree of the highlights with support vector regression. Finally, the highlights are ranked with descending sort according to their importance value. The ranking results can provide a hierarchical video browsing and customized content delivery scheme. Initial experimental results on soccer videos show an encouraging performance comparing with human subjective evaluation.

Categories and Subject Descriptors
H.3.1 [**Content Analysis and Indexing**]: *abstract methods, indexing methods*.

General Terms
Management, Measurement, Design, Experimentation.

Keywords
Replay detection, Highlight ranking, Video browsing.

1. INTRODUCTION
Sports video has been widely studied for its large number of viewer-ship and tremendous commercial potentials. For a lengthy sports game, only a few parts are attractive to audience. Therefore, highlights based sports video summarization attracted much attention in the recent years [1-7]. For example, L. Duan *et al* [5] proposed a mid-level representation framework for event analysis. An audio based highlights detection scheme was designed in [2]; A. Ekin *et al* [6] presented a heuristic based analysis method. However, these works almost focused on highlights extraction and video summarization. Litter work has been developed to evaluate the extracted highlights.

In this paper, we present an advanced highlight analysis scheme, in which the highlights are ranked according to their important values besides the highlights extraction. Because all the special events occurred in the sports videos can be called highlights, it is obvious that different highlights have different importance to audience. For instance, most of us more like to browse the goal and shoot events than red-yellow card events in soccer video. Moreover, highlights ranking is also benefit for hierarchical video browsing and customized content delivery against limited network/device capacity. In the proposed system, we first simply introduce a robust and effective method to extract highlights. Then, we use the Support Vector Regression (*SVR*) to evaluate the interesting degree (importance value or confidence) for each highlight with the help of domain-specific knowledge. The evaluation result can be taken as evidence for ranking.

The contributions of this paper can be concluded as follows:

1) We present an advanced highlight analysis system, which can give an automatic evaluation and ranking for highlights besides highlights localization. The highlights ranking can provide a hierarchical video browsing and customized content delivery.

2) We propose a novel highlight evaluation and ranking algorithm. With the help of domain knowledge, *SVR* is performed to evaluate the highlights based on several significant cues. The output of *SVR* is regarded as the confidence of the highlights.

The rest of this paper is organized as follows: the overview of this system is introduced in Section 2. The highlights detection method is simply presented in Section 3. Section 4 gives the details of highlight ranking scheme. Experiments are reported in Section 5, and followed by conclusions in Section 6.

2. SYSTEM OVERVIEW
The proposed system is composed of three components, i.e., highlight detection, evaluation, and ranking for browsing. Highlights are usually replayed with slow-motion patterns to show the details in broadcast video, so we can localize the highlights according to replay detection. A complete highlight contains a few shots ahead the replay, because they are the original source of the replay scene. In this paper, we put our emphasis on highlight evaluation and ranking. We adopt *SVR* to evaluate each highlight with the help of domain-specific knowledge. The output of *SVR* is regarded as the importance value of the highlights for hierarchical browsing and customized content delivery. The flowchart is illustrated in Figure 1.

MM'05, November 6–11, 2005, Singapore.

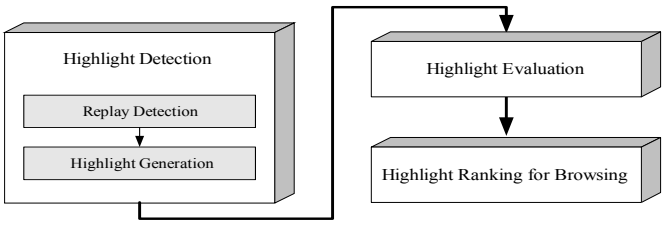

Figure 1. Flowchart of the proposed system

3. Highlight Detection

For broadcast sports videos, interesting or important segments, i.e., highlights, are generally played-back with a slow-motion pattern to show its details for audience. Thus, a replay scene can be taken as a significant indication of a highlight [9]. We use the replay detection to localize the highlights in this work.

3.1 Replay Scene Detection

The previous work of replay detection can be categorized into two classes: logo [5] and context [15] based methods. But both of them are not robust due to inevitable mistake in logo detection and difficulties in modeling replay pattern.

We consider both the logo and the intra- and inter-shot context for replay detection [10]. Our method comprises three steps: First, we extract some logo-transitions through frame-to-frame difference characterization, and then we get the logo-template through clustering based on these transitions. Second, we use the extracted logo-template to detect the replay logos in the whole video by template matching. Finally, due to false alarms in logo detection or video editing, we adopt intra- and inter-shot context aided by the SVM learner to identify replay scenes located by pairs of logos. Experiments demonstrated the effectiveness of the method [10]. An example of a replay scene is shown in Figure 2.

Figure 2. A replay scene

3.2 Highlight Localization

A highlight can be taken as a special play-unit [8]. A play-unit is defined as an individual segment that usually contains some semantic cues, and it should: 1) reflect the appropriate inherent structure of video data; 2) be consistent with human understanding for sports; 3) facilitate video indexing and retrieval. We localize the highlight play unit with replay extension.

A replay scene is just an indicator of a highlight, but it is inadequate to represent a whole highlight segment. Commonly, a highlight scene is composed of: interesting live play shots, replay scenes, and live break shots, such as close-ups and audience views, etc.

For sports video analysis, often we first parse the video into shots and classify them into several pre-defined categories that correspond to certain clear meaning [11]. For instance, generally long field-view and medium field-view shots correspond to play process and the others (such as close-up, replay, audience, etc.) indicate break state in soccer videos. A play-unit starts with a long field-view shot and ends with a break shot. Successive play shots are merged into one, but this strategy does not applied for consecutive break shots. The procedure of the highlight play-unit detection contains the following steps: 1) find the start point through forward search from the beginning of the replay scene. A highlight play-unit should begin with a long field-view shot meeting certain motion and duration constrains ahead of the replay. The segment from the long field-view shot to the replay scene contains the interesting event that is played back by the replay scene. 2) Find the end point through backward search from the end of the replay scene. A play-unit should end either with the replay scene or a break shot that is relevant with the live play scene and closely follows the replay scene. The segment between the start point and end point is regarded as a highlight play unit.

4. Highlight Evaluation and Ranking

4.1 Highlight Evaluation

Highlight detection cannot indicate how much interesting a highlight is, i.e., the confidence or important value. In this subsection, we present a scheme to evaluate the confidence of highlights.

It is obvious that we have to rely on domain knowledge to give the evaluations of the highlights, for the highlights in different sports game represent different meaning. In soccer domain, the highlights usually occur at the following cases: a goal, a shoot, an interesting attack, severe foul (such as card), offside, and others [12]. In this work, taking the soccer video as an example, we evaluate a highlight with the following cues (see Figure 3):

1) Length of a replay scene (*RP*). Usually, the longer a replay scene is, the more attractive the segment is [13].

2) Duration of goalmouth views in a long field-view shot before the replay scene (*GM*). The goalmouth views are often shown before the replay scene in the cases of interesting goal, shoot and attack.

3) Audience views after the replay scene (*AU*). An excited audience shot will be displayed after an interesting event according to general video editing rule.

4) Goal-net within the replay (*GN*). According to observation, the goal-net views appearing in the replay often indicates a highly interesting segment.

5) Scoreboard superimposed on long field-views (*SB*). These views are shown after a successful goal (a score). In these views, the caption containing score information is usually superimposed onto long field-views.

The replay scene can be detected by combining replay-logo and context information as mentioned above. A goalmouth view is determined by approximate estimation of slant angle of the field-view. An audience view is characterized by its lower field-ratio and high complexity texture feature. A goal-net view takes the field as background, and has certain range of contrast, entropy, energy of its Gray-Level Co-occurrence Matrix. A scoreboard is a manual-label caption that can be considered as a special texture aligned by vertical strokes. It can be detected by local-accumulated gradient [14], which consists of gradient computation, run-length smoothing, morphological open operation, region segmentation and region verification. The details of extracting these five features can be seen in [8].

Figure 3. Five cues for highlight evaluation

Actually, the five cues are selected empirically according to domain-specific knowledge and broadcasting rules. It is hard to automatically mine and determine the valid cues for highlight ranking.

We take the duration lengths of the above five cues as the observations for highlight evaluation. Thus we can get a feature vector comprised of (*GM*, *RP*, *GN*, *AU*, *SB*) to represent a highlight. We use a Support Vector Regression to estimate the importance of a highlight. In the experiments, the *RBF* kernel is used. The output of *SVR* is used to rank the highlights.

4.2 Highlight Ranking for Browsing

For most users, they often have the following experiences or criterions in browsing a video: 1) they prefer the highlights rather than the whole program; 2) they incline to first browse the more interesting scenes; 3) sometimes, they only concern the top interesting highlights due to device capacity and transmission time.

For browsing, we represent a highlight with a three-element vector: $h_i =< t_i, d_i, c_i >$, where t_i is the start time of the highlight, and d_i is its duration (length), and c_i corresponds the confidence. In previous work, the detected highlights are just sorted by their start time:

$$H_t = \{h_i\}, \exists h_i.t > h_j.t, \forall i > j$$

But according to the confidence of highlights, the ranking result is:

$$H_c = \{h_i\}, \exists h_i.c > h_j.c, \forall i > j$$

The set H_t and H_c are corresponding to the highlights ranking results according to time and highlight confidence respectively. In this paper, we focus on the latter.

Due to limited time, a user only wants to browse some more interesting highlights. This can be easily achieved by ranking with confidences of the highlights. For examples, if the total delivery highlights are limited within the time of Th_t, the returned result set should be:

$$R_t = \{h_i, i = 1, 2, \cdots, c \,|\, h_i \in H_c, \sum_i h_i.d <= Th_t\}$$

Sometimes if the browsing task is limited in number of returned items, for an example not exceeds *n*, the highlight set for delivery is:

$$R_c = \{h_i, i = 1, 2, \cdots, n \,|\, h_i \in H_c\}$$

5. EXPERIMENTS

We conduct the experiments on four complete soccer games: Portugal vs. England and Portugal vs. Holland in Europe Cup 2004, Brazil vs. England and Germany vs. USA in FIFA 2002. They are all captured from TV recorder.

Firstly, we detect all replay scenes, and extend them to segment the highlight scenes. To subjectively evaluate the highlights and get the ground truth data, we design a program for manual highlights confidence labeling. The scores of highlight confidence labeling are limited within an interval of 0 to 10. The more interesting a highlight is, the higher score will be assigned. We invited four individual persons to independently label the confidence score according to their own understanding of sports highlight. They are all sports fans and have rich experience of sports video enjoyment. All the subjective labeling results are equally treated and directly put together. The ground truth of evaluation of a highlight is the average of all subjective scores without any predilection.

With the subjective evaluation score, we can construct an evaluation tool that takes several key attributes mentioned above as observations for *SVR*. Then the output of *SVR* is used for automatic highlights evaluation and ranking. To measure the performance of the automatic highlights evaluation tool, we compare the results generated by the algorithm with those of manual labeling. In our initial experiments, we use three videos to train *SVR*, and the rest one for testing.

In the experiment, we use the game "video 1: Portugal vs. England in Euro2004" (eleven highlights contained), and the game "video 2: Brazil vs. England in FIFA2002" (ten highlights included) to test the performance of automatic evaluation and ranking method. The comparison between manual subjective and automatic evaluations is shown in Figure 4 and 6. The evaluation scores are restricted within the interval [0, 10]. We use the absolute difference to measure the evaluation performance.

$$d_i = \left| s_i^a - s_i^m \right|$$

where s_i^a and s_i^m represent the i^{th} automatic evaluation and manual labeling score respectively. We calculate the mean and variance of the difference serial $\{d_i\}$. For video 1, the *mean* = 1.21 and, *var* = 0.79; for video 2, *mean* = 1.60, *var* = 1.65. The difference will decrease if normalization of the scores is performed.

It can be found from the comparison results that the importance value output by *SVR* basically consists with subjective evaluation. Some errors and differences mostly result from two factors: (1) Selected features cannot definitely reflect the content of the highlights; and (2) Different subjects have different understanding on highlights, which bring with some excursions in the ground truth. In the future work, we will mine more valuable features and incorporate other sophisticated methods for evaluating the highlights.

After evaluation, the ranking can be carried out according to scores. The comparison between manual and automatic ranking results is drawn in Figure 5 and 7. The overall highlights are ranked with a descend order. If more than one segment is denoted the same score, they are assigned the same rank. From the results, we can see that the trend of the manual labeling and automatic ranking results are approximately similar, and the proposed ranking method is encouraging.

After highlight ranking, we can browse the scenes according to different requirements, such as hierarchical highlight browsing, content delivery according to network/device capacity, budget or cost of time of economy for data transmission. The browsing priority can be freely chosen by the users according to their requirements.

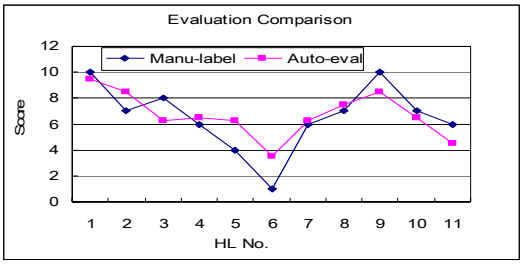

Figure 4. Evaluation comparisons for video 1

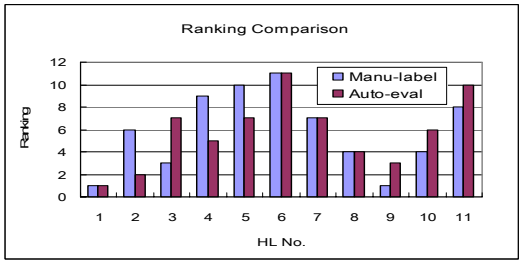

Figure 5. Ranking comparison for video 1

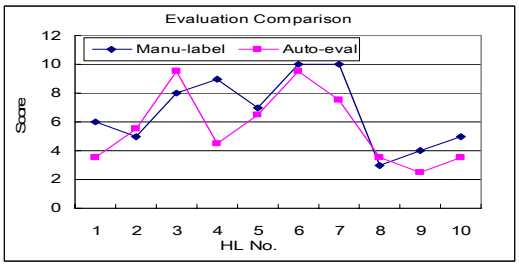

Figure 6. Evaluation comparison for video 2

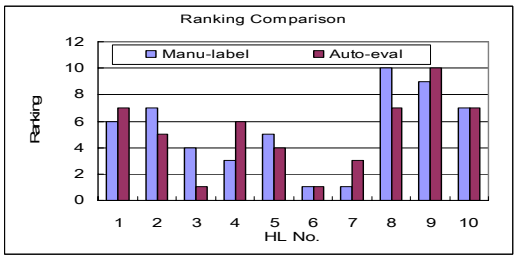

Figure 7. Ranking comparison for video 2

6. CONCLUSIONS

This paper proposed a system for highlight evaluation and ranking for hierarchical browsing and customized content delivery. Different from previous work, this system focused on automatic highlight evaluation and ranking rather than typical highlight detection. We ranked the highlights by five cues with domain knowledge and an SVR. The results of ranking can be used to facilitate browsing required in different situations. Initial experiments show the promising performance of the proposed method.

In the future, we will use more data to test our evaluation method, and use the data mining technique to perform automatic feature selection and highlight confidence measurement. We will also extend this work to other sports genres.

7. ACKNOWLEDGEMENT

This work is supported by National Natural Science Foundation of China (Grant No. 60475010 and 60121302), and an international joint project between China and Singapore.

8. REFERENCES

[1] D. Yow, B. Yeo, M. Yeung, and B. Liu, "Analysis and presentation of soccer highlights from digital video", *Proc. of ACCV' 95*, 1995.

[2] Y. Rui, A. Gupta, A. Acero, "Automatically extracting highlights for TV baseball programs", *Proc. ACM Multimedia*, pp.105-115, 2000.

[3] S. Nepal, U. Srinivasan, G. Reynolds, "Automatic detection of 'goal' segments in basketball videos", *Proc. of ACM Multimedia*, Ottawa, pp.261-269, 2001.

[4] Y. Gong, M. Han, W. Hua, W. Xu, "Maximum entropy model-based baseball highlight detection and classification", *Computer Vision and Image Understanding*, 96, pp.181-199, 2004.

[5] L. Duan, M. Xu, T. Chua, Q. Tian, C. Xu, "A mid-level representation framework for semantic sports video analysis", *Proc. of ACM Multimedia*, pp. 33-44, 2003.

[6] A. Ekin, A. Tekalp, "Automatic soccer video analysis and summarization", *IEEE Trans. Image Processing*, 12(7), pp. 796-807, 2003.

[7] K. Wan, C. Xu, "Robust soccer highlight generation with a novel domain-specific feature extractor", *Proc. of ICPR*, 2004.

[8] X. Tong, Q. Liu, L. Duan, H. Lu, C. Xu, Q. Tian, "A unified framework for semantic shot representation of sports video", *ACM Workshop MIR 2005*.

[9] B. Li, J. Errico, H.Pan, and I. Sezan, "Bridging the semantic gap in sports", *SPIE Storage and Retrieval for Media Databases 2003*. vol. 5021, pp.314-326, 2003.

[10] X. Tong, H. Lu, Q. Liu, H. Jin, "Replay detection in broadcasting sports videos", *Proc. Conf. Image and Graphics*. pp. 337-340, 2004.

[11] X. Tong, Q. Liu, H. Lu, and H.Jin, "shot classification in sports video", *Proc. of Int'l Conf. Signal Processing*, 2004.

[12] J. Dai, L. Duan, X. Tong, C. Xu, Q.Tian, H. Lu, and J. Jin, "Replay scene classification in soccer video using web broadcast text", *Proc. of ICME*, Netherlands, July, 2005.

[13] Y. Takahashi, N. Nitta, N. Babaguchi, "Automatic video summarization of sports videos using metadata", *Prof. PCM*, Tokyo, Japan, Nov, 2004, pp. 272-280.

[14] C.Wolf, J.Jolin, and F.Chassaing, "Text Localization, Enhancement and Binarization in Multimedia Document", *Proc. Int'l Conf. Pattern Recognition*, pp. 1037-1040, 2002.

[15] J. Wang, E. Chng, C. Xu, "Soccer replay detection using scene transition structure analysis", *Proc. of ICASSP 2005.*