Providing On-Demand Sports Video to Mobile Devices

Qingshan Liu Zhigang Hua Cunxun Zang Xiaofeng Tong Hanqing Lu
National Laboratory of Pattern Recognition, Institute of Automation,Chinese Academy of Sciences
No. 95 East Road, Zhong Guan-Cun, Beijing, 100080, China
{qsliu, zghua, cxzang, xftong, luhq}@nlpr.ia.ac.cn

ABSTRACT
This paper introduces a system for providing on-demand sports video to mobile devices, which has two main contributions. First, we construct an infrastructure for extracting and delivering the highlights instead of the whole sport videos to mobile clients, which can significantly reduce the bandwidth consumption. Second, we design an advanced UI for the mobile clients to effectively browse and interact with the video highlights. To validate the practicality and effectiveness of this system, we conduct the experiments on several real soccer videos. The results demonstrated that more than 65% of bandwidth consumption could be reduced. Moreover, the initial user study results show that the mobile users could interact effectively with the interface to seek or navigate sports videos.

Categories and Subject Descriptors
H.5.2 [**User-Centered Design**]: *Video content analysis, UI design*.

General Terms
Algorithm, Performance, Design, Experimentation.

Keywords
Multimedia content adaptation, sports video, highlight extraction

1. INTRODUCTION
Currently, the handheld Internet-capable devices have been undergoing a booming prosperity in people's daily lives. With these portable devices, it is convenient for the mobile users to access the Internet anywhere and anytime. As the multimedia contents are becoming widely available on the Internet, it is necessary to facilitate the mobile users to access them. To let the mobile users really enjoy the ease of video resources, there still exist some hurdles that need to be crossed [3]. Among them, the limited accessible bandwidth is one key bottleneck.

To reduce the wireless network load of mobile devices, many efforts have been put on multimedia adaptation. For instance, the ROI coding scheme and Spatial/SNR scalability in the JPEG2000 standard [2] provides a functionality of progressive encoding and display. It is useful for fast database access as well as for delivering different resolutions to terminals with different capabilities. The works [1] propose an attention model based image adaptation approach. Instead of treating an image as a whole, it manipulates each region-of-interest in the image separately, which allows delivery of the most important regions when the bandwidth is limited or the screen size is small.

Due to the increasing availability and use of digital video data on the Internet, video compression for delivery will be an important performance factor on the Internet. Schojer et al [8] proposes an architecture called QBIX that supports real-time adaptation in the compressed and decompressed domain driven by the clients' terminal capabilities. There are some other criteria proposed to adapt the quality of the videos. Examples are periodic caching of layered coded videos (LVC), quality adjusted caching of GoPs (group of pictures), adaptive caching of LVC in combination with congestion control [6,7]. Most them only focus on compressing and caching contents in order to reduce the data transmission for fast delivery over the limited bandwidth. Therefore, the results are often not consistent with human perception on small displays because of excessive resolution reduction or quality loss.

This leads us to design a novel system for providing video services to mobile clients based on video content analysis. As is clear that, various video parts attract different attentions from users. Thus, this system focuses on reducing bandwidth consumption with transmitting important video segments to the users rather than compressing video in the quality domain. Our first prototype is designed to the sports video because it is widely available on the Internet and attracts numerous audiences. Generally, the highlights in the broadcast sports video are indicated by the replay patterns [5][11]. Based on the replay detection, an efficient highlight extraction algorithm is developed in the system. Furthermore, an advanced UI is designed to effectively browse these highlights on the mobile clients. The experimental results demonstrate that this system can reduce about two folds of the bandwidth consumption. Moreover, the initial user study demonstrates the effectiveness of the UI design.

2. SYSTEM DESIGN
The system architecture is presented in Figure 1, which comprises two core components, i.e. the highlight extraction component at the video server side, and the client UI component at the mobile client side. The system workflow can be formulated as:

1) The mobile client begins a request to an online sports video.
2) The front end of sports video server gets the client request and forwards the request to the video fetching component.
3) The video fetching component examines the client request and determines whether to call the highlight extraction component.
4) Video service component transfers the original or highlight video to the fro nd end.
5) The client interacts with the video via the local browsing UI.

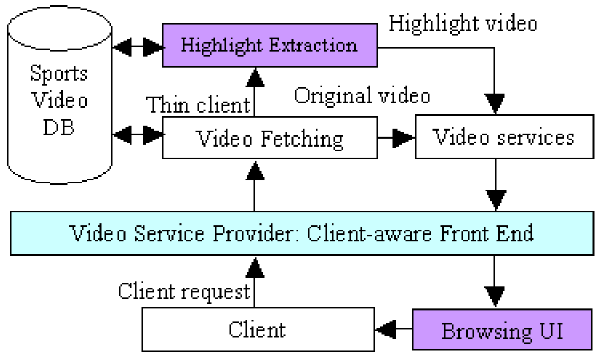

Figure 1. System architecture of our approach.

2.1 Highlight Extraction Algorithm

Highlights represent very interesting parts of a game [5][11][12]. In broadcast sports video, the highlights are often replayed with slow-motion patterns, so the highlights can be achieved by replay detection. The previous approaches of replay detection can be categorized into two classes: logo based [5] and context based [11]. But both of them are not robust due to inevitable mistake in logo detection and difficulties in modeling replay pattern.

We develop a robust replay detection algorithm to segment highlights, in which both the logo and context are considered. At the beginning and end of the replay, often there appears a logo-transition shown in Figure 2. Based on this observation, we first use the logo-transition detection to get the logo template, and then we can get the logo-pair segments with logo template matching. Finally, we adopt intra- and inter-shot context aided by the SVM learner to identify replay scenes located by a pair of logos. We describe the three-step algorithm in more details as follows.

1) Logo-transition Detection

According to extensive observations, a typical logo-transition usually lasts for 0.5-0.8 seconds or 15-24 frames. Especially, the intensity difference between two adjacent frames measured by mean square difference (MSD) is usually like a plateau shape, as shown in Figure 3. The formula of MSD between two adjacent frames f_k and $f_{k\text{-}1}$ is:

$$MSD(k) = \frac{1}{M \cdot N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} [f_k(i,j) - f_{k-1}(i,j)]^2$$

where M, N are height and width of a frame image. The logo-transition detection can be performed through plateau-like shape detection over the sequence of frame-to-frame difference [10].

2) Logo Detection

Based on the properties that the logo is superimposed and near vertical symmetry, we extract some logo candidates from the logo transitions. With these logo candidates, we first regard their center as the preliminary logo template. In order to eliminate the effect of background, we then compute the average image of the qualified candidates that are near to the center, and take it as the final logo-template.

Since the replay-logo is usually invariable all through the video, we use the extracted logo-template to detect all the replay-logos in the same video with template matching. Color (normalized color correlation coefficient) and shape (invariable shape moments) features are employed.

Figure 2. An example for the Logo-transition.

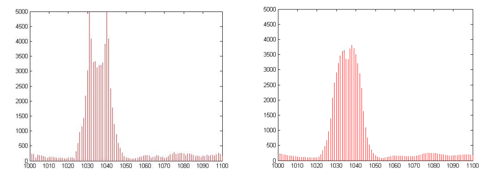

Figure 3. Plateau shape like MSD in the logo-transition.

3) Replay identification

Now, a video can be divided into segments with taking logos as boundaries. Due to false alarms in logo detection or video editing, we further analyze the following features on a segment to robust discriminate replay scenes: 1) Length of the segment (LS). 2) Counts of four shot types: long view shot (CLS), medium view shot (CMS), close-up (CCS) and out-field shot (COS). 3) Average length of shots in the segment (ALS). 4) Motion information: mean and variance of motion activity (MMA and VMA) between two successive frames. Motion activity is defined as the mean of magnitude of motion vectors in a motion map.

Thus, a segment is represented by a feature vector with eight elements, i.e. <SL, CLS, CMS, CCS, COS, ALS, MMA, VMA>. Based on these features, we employ a Support Vector Classifier to perform segment classification, in which the RBF kernel is used. Five-fold cross validation and grid-search criterion are applied to select the optimal kernel parameters.

Now, with the detected replay, we can easily localize the highlights in the sports videos. In addition, a certain number of shots (usually no more than 5 in our practice) ahead a replay are also contained in the highlights, because these shots are the very sources of the replay.

2.2 Sports Video Highlight Transmission

After the highlights are extracted, the video server transfers these video segments to the mobile clients instead of the whole video contents. In our system, in addition to the highlights, an XML file of the highlight structure of a soccer video is also transmitted. An example for this XML file is shown in Figure 4.

```
<Video Name="EURO2004_Portu_Eng" Length=197856>
    <Highlight No=1>
        <Position StartFrame=2505 EndFrame=3223/>
        <Overview Frame=2505/>
        <Content File="seg1.mpg"/>
    </Highlight>
    <Highlight No=2>
        <Position StartFrame=6706 EndFrame=7223/>
        <Overview Frame=6706/>
        <Content File="seg2.mpg"/>
    </Highlight>
    …
</Video>
```

Figure 4. An example for the highlight structure file.

Figure 5. An example for the highlight browsing user interface on smart phone.

For each highlight, the serials of its starting and ending frames are indicated in this file. In our current prototype, each start frame is by default assigned as the overview image for a highlight segment. In future work, we will employ sophisticated algorithms to acquire the most representative frame as the highlight overview. Furthermore, the highlight content segments are also presented in this file. We can see that the bytes for the sports video delivery are greatly reduced, since only highlight segments are transmitted in our case.

2.3 Browsing UI Design

After the video server responds the highlights to the mobile clients, it is necessary to design a client-side browsing UI to effectively view these video segments. We design a browsing interface on a Dopod 575 smart phone with a 176x224 pixel resolution and Microsoft Smartphone 2003 as the operation system. As shown in Figure 5, the highlight viewer comprises three main components, namely the play panel on the top row, the progress bar on the middle row and the highlight overview panel on the bottom row.

A notable feature of this interface is that all the three components are responsive to each other. For example, when a user operates in the progress bar, the play panel is navigated to the corresponding video part, and the highlight overview panel will be also automatically switched to the corresponding highlight.

- The highlight overview panel presents overviews for all of the highlights in a video according to the temporal sequence. Due to small display size on the smart phone, this panel can only display three highlight overviews concurrently. Users can use the left and right key button to freely switch the previous or posterior highlights. When the user clicks on any highlight overview, that highlight will be placed to play on the play panel, and the progress bar will also be scrolled to the corresponding position.
- The video play panel is designed with a larger display area in order to let users clearly view the video content. This panel is responsive to both the highlight overview panel and the progress bar: a drag on the progress bar or a click of the highlight panel will generate an automatic trigger to the play panel that will play the corresponding video content.
- The progress bar is designed to show playing status of a video. As shown in the figure, three different colors are identified in various parts of the progress bar. Detailedly, the blue color refers to a video highlight segment, the white color indicates the interim video content between two highlights, and the red color represents a highlight segment that is currently played.

We believe that such a well-designed UI will facilitate the users to browse through a lengthy video easily. The initial user study results demonstrate the effectiveness of this browsing UI, which will be reported later in the experiments.

3. EXPERIMENTS

We carried out experiments on several recent real soccer videos selected from the World Cup 2002, which are listed in Table 1. All the videos are the 90-minute soccer games with a high-quality resolution and a large size.

Table 1. The experimental dataset: four soccer videos.

ID	1	2	3	4
Match	Cameron vs. Germany	China vs. Turkey	England vs. Sweden	USA vs. Portugal

The Dopod 575 smart phone is set as the mobile client, and a powerful PC machine is selected as the video server that provides the video services the mobile clients. The communication between the two machines is connected by the wireless Bluetooth. Such a test bed setup can reflect the performance of our approach by serving sports video highlights to mobile clients.

3.1 Highlight Extraction Performance

We first look into the performance of our highlight extraction algorithm. As has been described above, the highlights in a sports video are always indicated by the replay segments. Thus, the highlight extraction performance is directly related to the replay detection. Table 2 reports the results of the replay detection on the soccer videos. It can be seen from the table that, the precision can maximally reach 97.7%, and the recall ranges from 69% to about 96%. The quantitative results showed that our system could effectively identify the.

The missing replay detections are mostly caused by the logo missing in the period of physical recording. For instance, in the match of China vs. Turkey, it happens that an on-going replay will be frequently interrupted. Thus, the ending logo is lost, which is the due reason responsible for this low precision of detection. The false detection is minor, mainly due to the wrong identification of logos. In future work, we would strengthen the replay detection by studying and incorporating more sophisticated techniques.

Table 2. The results of replay detection.

ID	Detect	False	Missing	Precision	Recall
1	38	2	4	95.0%	90.5%
2	38	6	17	86.4%	69.1%
3	43	1	2	97.7%	95.6%
4	58	4	9	93.5%	86.6%

3.2 Bandwidth Optimization

Due to the highlight transmission rather than the whole video delivery, our approach can manage to save the bandwidth consumption of the mobile clients. We carried out a quantitative analysis of the real bandwidth optimization result. Figure 6 presents the transmission bytes by the four original videos versus their highlights. The rightmost column is the average statistical result of the five sports videos. On average, more than 65% of the bandwidth consumption could be reduced with our highlight transmission approach. The result demonstrated that our approach could significantly save the bandwidth usage.

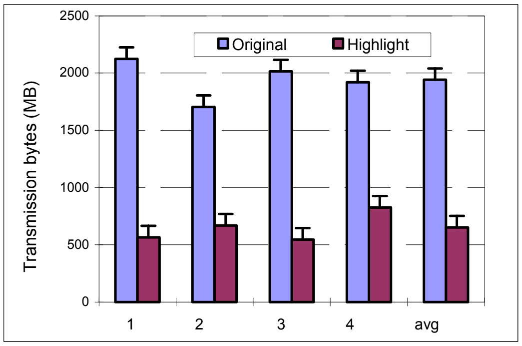

Figure 6. The transmission bytes of the original vs. highlight.

3.3 User Study

We first conducted a user study to validate the highlight representation of lengthy sports video to the mobile clients. Our study involved 4 computer science graduate students, who are familiar with the operations on mobile devices. They were told to use the client browsing UI to view the video highlights on the phone, which would help them to get familiar with this system.

Initially, we fielded a questionnaire concerning the usability of the system among the users to get their feedbacks: 1) the highlight representation is able to keep the main information of a sports video, not causing severe information loss; and 2) The highlight video helps to view a sports video more easily on mobile devices than a whole video. Both of the two questionnaire items were answered on a five-scale Likert scale where 1=strongly disagree and 5=strongly disagree. The result for the first problem is 4.0 on average for various users, indicating that the highlight could keep the information of video content. The result score for the second problem was about 4.3, further demonstrating the effectiveness of our UI design on helping users to browse the video.

We then asked the four users into two groups, and the two groups were respectively assigned the tasks of finding out the goals in the original soccer videos and in the highlights. The time for each task was recorded, and the average time cost (measured by seconds) for various users over all the four videos are shown in Table 3. The rightmost column in the table represents the percentage of the browsing time reduction. As can be seen from the table, our approach can achieve about 2-fold reduction of the finding time. This result verifies that the users could easily interact with our system to seek or navigate a video.

Table 3. The average time cost of the goal-finding tasks.

Approach	Highlight	Original	Save Ratio
Time cost	160	475	66.3%

4. CONCLUSIONS

This paper presented a system for providing on-demand sports video highlights to mobile devices. The system uses an effective highlight extraction algorithm to transfer the highlight segments instead of the whole videos to the mobile clients. An effective browsing UI was also designed for the mobile clients to interact the highlight video. As a result, the system is capable of reducing more than 65% of wireless bandwidth usage. Moreover, the preliminary user study demonstrated that mobile users preferred such a highlight representation. We will apply our system to more sports videos for further investigations of its real effects.

5. ACKNOWLEDGEMENT

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

6. REFERENCES

[1] L.Q. Chen, X. Xie, X. Fan, W.Y. Ma, H.J. Zhang, and H.Q. Zhou, A visual attention model for adapting images on small displays, ACM Multimedia Systems Journal, 2003.

[2] C. Christopoulos, A. Skodras, and T. Ebrahimi, The JPEG 2000 still image coding system: an overview, IEEE Trans. on Consumer Electronics, 46(4): 1103-1127, 2000.

[3] W.Y. Ma, I. Bedner, G. Chang, A. Kuchinsky, and H.J. Zhang, A framework for adaptive content delivery in heterogeneous network environments, Proc. of Multimedia Computing and Networking 2000.

[4] R. Mohan, J.R. Smith, and C.S. Li, Adapting multimedia Internet content for Universal Access, IEEE Trans. on Multimedia, 1(1):104-114, 1999.

[5] H. Pan, B. Li, and M. Sezan, Automatic detection of replay segments in broadcast sports programs by detection of logos in scene transitions, Proc. of Int'l Conf. on Acoustics, Speech, and Signal Processing, 2002.

[6] R. Rejaie and J. Kangasharju, Mocha: A Quality Adaptive Multimedia Proxy Cache for Internet Streaming. Proc. of Int'l Workshop on Network and Operating Systems Support for Digital Audio and Video, 2001.

[7] M. Sasabe, N. Wakamiya, M. Murata, and H. Miyahara. Proxy Caching Mechanisms With Video Quality Adjustment. Proc. of Internet Multimedia Management Systems, 2001.

[8] P. Schojer, L. Böszörményi, H. Hellwagner, B. Penz, and S. Podlipnig, Architecture of a quality based intelligent proxy (QBIX) for MPEG-4 videos. Proc. of Int'l Conference on World Wide Web. Budapest, Hungary, May 2003.

[9] J.R. Smith, R. Mohan, and C.S. Li, Content-based transcoding of images in the Internet, Proc. of Int'l Conf. on Image Processing 1998, Chicago, USA, Oct. 1998.

[10] M. Wu, W. Wolf, and B. Liu, An algorithm for wipe detection, Proc. of Int'l Conf. on Image Processing 1998, Chicago, USA, Oct. 1998.

[11] J, Wang, E. Chng, and C. Xu, Soccer replay detection using scene transition structure analysis. Proc. of Int'l Conf on Acoustics, Speech, and Signal Processing, 2005.

[12] N. Babaguchi, Y. Kawai, T. Ogura, T. Kitahashi, Personalized Abstraction of Broadcasted American Football Video by Highlight Selection, IEEE Trans Multimedia, 6:575-586,2004.