A Framework for Providing Adaptive Sports Video to Mobile Devices

Cunxun Zang Qingshan Liu Xiaofeng Tong Hanqing Lu
National Laboratory of Pattern Recognition, Institiue of Automation, Chinese Academy of Sciences
P.O.Box 2728, Beijing, China
{cxzang, qsliu, xftong, luhq}@nlpr.ia.ac.cn

ABSTRACT
In this paper, we present a novel framework that serves adaptive sports video to mobile users. Our framework combines content-based sports highlights extraction and quality-domain video compression technologies in video server, capable of reducing wireless bandwidth consumption more effectively. We develop a robust replay-based highlights extraction method, and propose a content-based video streaming coding scheme to handle the problems of bandwidth and capacity of computation. To validate the practicality and effectiveness of our system, we conduct the experiments on several real soccer videos. The experimental results demonstrate the robustness of our highlights extraction method and more than 77.5% of the bandwidth consumption could be reduced.

Categories and Subject Descriptors
H.5.1 [**Multimedia Information Systems**]: *Multimedia streaming, Video content analysis.*

General Terms
Algorithm, Performance, Design, Experimentation.

Keywords
Multimedia content adaptation, sports video, highlight extraction.

1. INTRODUCTION
Sports video has a wide viewership in the world. However, most audiences cannot sit at TV or PC to see their interested games due to many causes. Currently hand-handled devices have been becoming an alternative choice for watching sports video through wireless network. **S**mart mobile devices are becoming popular hand-handled devices. It is reported that global shipments of smart mobile devices is 13,004,680, up 75% year-on-year in Q3 2005, and the continuing shift from standalone handhelds to converged smart mobile devices was confirmed again today [4].

However, it is known that mobile devices have three limitations with present technology, i.e., the bandwidth, computation capacity, and the size of display. And among them, the limited accessible bandwidth is one key bottleneck. To solve this problem, many efforts have been put on multimedia adaptation. [1][2][3] focus on compressing and caching contents in order to reduce the data transmission for fast delivery over the limited bandwidth, and [8] focuses on reducing bandwidth consumption with transmitting important contents to the users. In this paper, we combine quality-domain video compression with content-based sports highlights extraction to reduce the bandwidth consumption more effectively, and construct a new system infrastructure, which has three core components including sports video analysis, wireless video transmission, and UI design.

WiCon'06, August 2–5, 2006, Boston, MA, United States.

For mobile clients, our system provides two kinds of sports video data, i.e., the original sports video and highlights video. –The latter is appropriate for mobile clients with limited bandwidth resource. Presently several schemes have been proposed to extract the highlights from broadcasting sports video[5][6][9], but most them have a complex computation. In this paper, a robust replay-based highlight extraction method is adopted [11], because highlights are generally replayed by slow motion patterns. We combine the logo matching with context analysis for replay identification in the system.

Our system uses the video streaming for wireless network transmission. Due to the limitation of computation capacity, it is impossible to directly use some general coding schemes, such as, Mpeg-2,4 H.264 and AVS[7]. In this paper, we customized normal general video coding schemes for mobile devices from two aspects. The one is to select the Intra-frame adaptively, and the other is to pay more attention on interesting regions coding. Taking an example of soccer video, the shot detection and field segmentation are considered for video streaming coding.

2. SYSTEM DESIGN
The system architecture is presented in Figure 1, which comprises three core components, i.e., the highlight extraction module, the video streaming encoder, and the mobile client UI. The former two are both at the video server side and the latter is at the mobile client side.

For mobile clients, our system provides two kinds of sports video data, the original sports video and highlights video, the latter is appropriate for mobile clients with limited bandwidth resource.

The system workflow can be formulated as:

1) The mobile client sends a request to the server console component on the server through wireless network. Two kinds of request are defined: one is for the video outline, details of which are shown in Figure 7; the other is for the sports video data, where the original and highlights videos are both provided.

2) The server console component examines the mobile client request: for the video outline, to deliver the result from the Outline Fetching component directly to the mobile client through wireless network; and for sports video data, to determine whether to fetch original video or to extract highlight video data; then to deliver the raw video data to the video streaming encoder.

3) The video streaming encoder encodes the received raw video data from the server console based on the content, and then transmits the encoded video data streaming to the mobile client.

4) The mobile client interacts with the video streaming via the browsing UI.

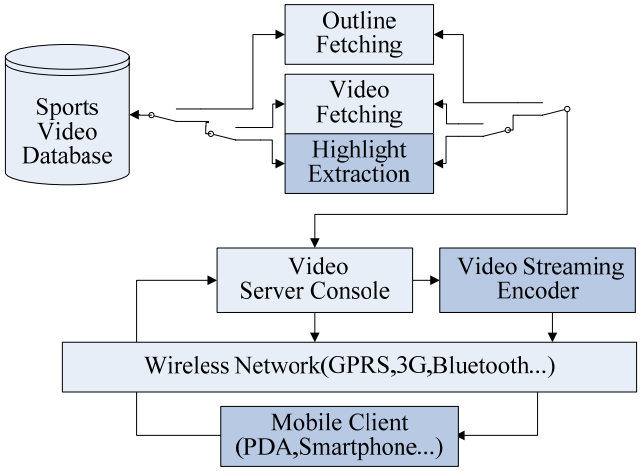

Figure 1. System architecture of our approch.

2.1 Sports Video Analysis

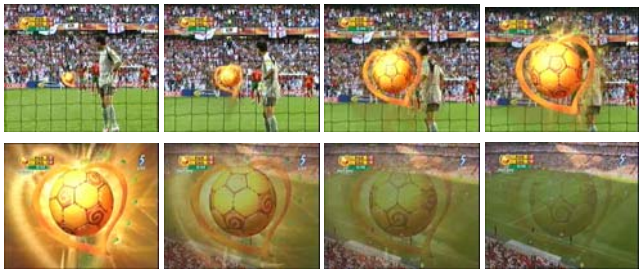

Figure 2. An example of logo-transition (8 frames are displayed).

Highlights represent very interesting parts of a game [5] [6][9]. In broadcast sports video, highlights are often replayed with slow-motion patterns, so the highlights can be obtained by replay detection. It is well known that in most cases there exists a special transition at both start and end of a replay, in which a highlighted logo comes in and out gradually. We call this transition as "logo-transition", an example of which is shown in Figure 2. The previous approaches of replay detection can be categorized into two classes: logo based [5] and context based [6]. But both of them are not robust due to inevitable mistake in logo detection and difficulties in modeling replay pattern.

In this paper, we adopt a robust replay detection algorithm to segment highlights, in which both the logo and context are considered. The framework is shown in Figure 3. 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. The details of the algorithm are discussed in our previous work [8].

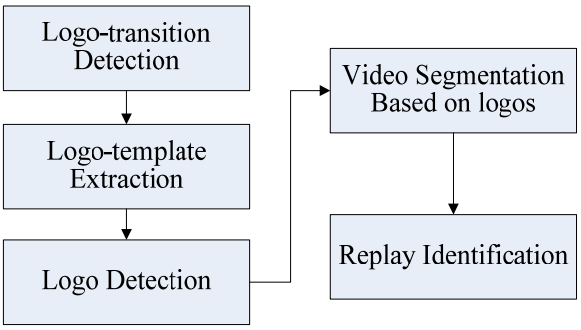

Figure 3. Flow of Replay Detection.

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 Video Streaming Customization

Due to the limitation of bandwidth and computation capacity, a coding scheme of low-rate and low computation cost in decoding is desired. In our system, a content-based video streaming coding scheme is proposed for mobile devices to further improve its performance.

1) Variable-Period Intra-Frame Selection Based on Shot Detection

In order to reduce the complexity of computation, most of current general coders adopt simple prediction models, some of which only have Intra and Predictive frames, but this simplification also leads to the decrease of the prediction performance. So how to select the appropriate Intra frames turns to be important.

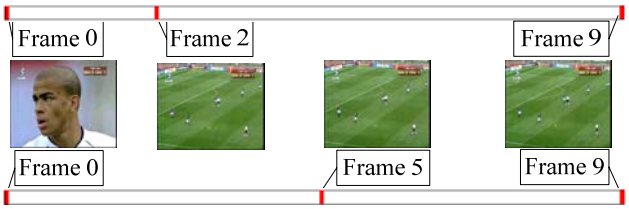

Figure 4. mismatch of GOPs and shot periods.

In most of current general coders , a video sequence is commonly divided into groups of pictures (GOP) in fixed-length time order. In fact, video sequences are made up of shots, and each of them contains successive frames with similar content so the first frame of each shot is fit for an Intra frame. However, GOPs with fix length cannot always keep matching shots completely. A simple example of such mismatch is demonstrated in Figure 4. It is obvious that Frame 2 is more suitable for the intra frame of the

second period. Frames 1, 2, 3 wouldn't predict precisely if Frame 5 is chosen as the intra frame.

To overcome this problem, we construct a scheme of adaptively selecting Intra frames based on shot detection. Since we need to detect shots in highlight extraction, no extra computation is demanded. We adopt a classical histogram method for shot detection [10][13], for there are almost shot cuts in sports video. According to the results of shot detection, we modify the fixed length GOPs to a variable length GOPs.

To avoid the length of GOPs too long or short, two thresholds (θ_1, θ_2) are adopted to limit the rang of the length. θ_1 is set 5 by default as the lower boundary; while θ_2 is twice of the original fixed length of one GOP as the upper boundary. Besides θ_2, a too-long GOP will be divided into two parts in average.

2) Attention Region Coding

In addition, the audiences often focus on the active region. For example, they seldom care what happens out of soccer field. So decreasing the visual quality of such non-attention regions a little cannot affect the visual effect while it can reduce the size of the encoded video data. Here, we also propose an attention region-coding scheme base on field and object segmentation for soccer games.

Field segmentation is also done in highlight extraction. Detailed algorithm was discussed in our previous work[12]. Figure 5 shows a segmentation result.

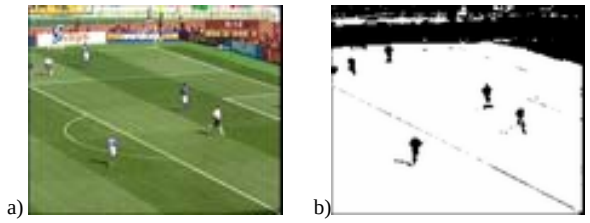

Figure 5. Field segmentation. (a) Original and (b) Segmented

Now we begin to generate the attention regions. It is easy to get the active objects in Long shots, for the soccer field is the main region of a frame. But some cases should be considered in medium shots, such as the case of Figure 6.

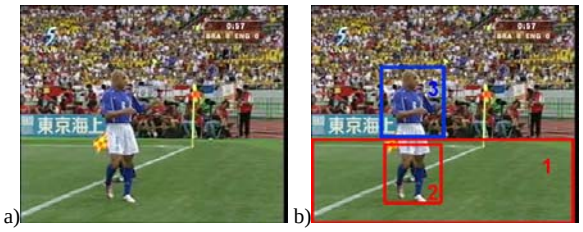

Figure 6. Field segmentation. (a) Original and (b) Segmented

In such cases, we first detect the region of the player inside the soccer field marked by Number 2 in Figure 6(b), according to the method in [11], and then we predict the region outside marked by Number 3, whose position and size are calculated approximately by the size and view direction of the soccer field.

Finally, we construct a scheme to adjust the QP value of every MB (microblock) by different degree of attention. Generally, the values of QP of non-attention regions are increased and attention regions are kept with the original QP scheme, which are shown as in

$$QP = \begin{cases} QP + N & (\textit{for non-attention region}) \\ QP & (\textit{for attention region}) \end{cases} \quad (1)$$
$$N \in [1,8]$$

Where N is the increased quantization step value for non-attention regions. A reference range of N is from 1 to 8 as experience and that of N is different in different video coders.

To evaluate the visual quality of different regions, a region based PSNR is defined as in

$$PSNR = 10 \times \log_{10} \frac{255^2}{MSE} dB \quad (2)$$
$$MSE = \frac{1}{|R|} \sum_{(x,y) \in R} (I(x,y) - I'(x,y))^2$$

Where R is a polygonal region.

2.3 Sports Video Highlight Transmission

After connected with the mobile clients, the video server first transfers the outline information of sports video, which contains highlights structure and description of sports videos, to the mobile clients. In our system, the outline information is recorded in an XML file and is transmitted in form of a compressed ZIP file. An example for the outline XML file is shown in Figure 7.

```
<Video Name="FIFA2002_ Brazil_England"
        Length=67040  Num_of_Highlights=13>
    <Highlight Index=0>
        <Position StartFrame=8142 EndFrame=8321/>
        <Overview Frame=8142/>
        <Description Title="section1_Bra_Eng"
        Content="This is the content of highlight 0 in Brazil
        vs England FIFA2002"/>
        <Video File="sec1_FIFA02_BRA_ENG"/>
    </Highlight>
    …
</Video>
```

Figure 7. An example for the highlight structure file.

For each highlight, the serials of frame information are indicated in the outline XML file. In our current system, the information for every highlight segment contains the starting and ending frames, the description of title and content and the outline overview. And each start frame is by default assigned as the overview image for a highlight. In the future work, we will employ sophisticated algorithms to obtain the more representative frame as the overview. Furthermore, the highlight content segments are also presented.

3. EXPERIMENTS

We carried out experiments on four real soccer videos selected from the World Cup 2002 and the European Cup 2004, which are listed in Table I. The frame rate of video sources is downsampled from 30 fps to 10 fps, and the frame resolution is from CIF-format (352x288) to QCIF (176x144).

Table 1. The experimental dataset.

ID	1	2	3	4
Soccer Videos	World Cup 2002		European Cup 2004	
	Brazil vs England	Germany vs USA	Portugal vs Holland	Portugal vs England

Based on our previous works, the video coding scheme WM3 was chosen as our basic scheme, which is designed for mobile devices by AVS[7]. In fact, our content-based video streaming coding scheme for mobile devices does not rely on any certain video coding scheme and we will work on applying our scheme based on MPEG and H.264.

The mobile client is the Dopod 696 PDA, which is running on Microsoft WindowsCE.Net PocketPC 2003 as the operation system and a Pentium4 3.0G Hz computer is selected as the sports video server. Both of them communicate over the wireless Bluetooth network.

3.1 Highlight Extraction Performance

As described above, the highlight extraction performance is directly related to the replay detection. Figure 8 reports the results of the replay detection on the selected soccer videos listed in Table I. The maximum precision can reach 96.23%, and the recall ranges from 72.41% to about 91.07%. The quantitative results showed that our system could effectively identify the sports highlights. The missing detections of replay are mostly caused by the logo loss in the period of physical recording. For instance, 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.

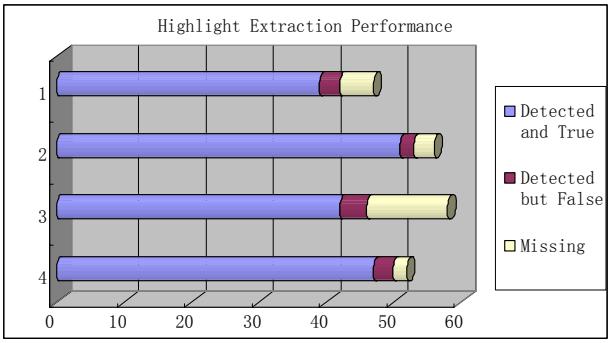

Figure 8. Highlight Extraction Performance

In future work, we would strengthen the replay detection by studying and incorporating more sophisticated techniques.

3.2 Bandwidth Optimization

Due to our highlight extracting method and content-based video streaming coding scheme, the bandwidth consumption can be reduced significantly and results are shown above.

1) Optimized by highlight extraction.

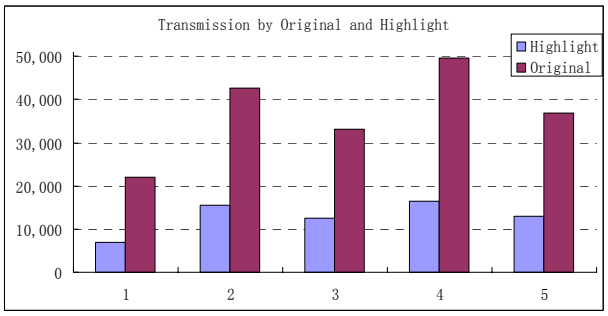

Figure 9. Transmission by Original and Highlight

We carried out a quantitative analysis of the real bandwidth optimization results, which are demonstrated that our approach of highlight extracting could significantly save the bandwidth consumption. Figure 9 gives the comparison of data bytes transmitted by original and highlight sports video. 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.

2) Optimized by VPIS (Variable-Period Intra-frame Selection)

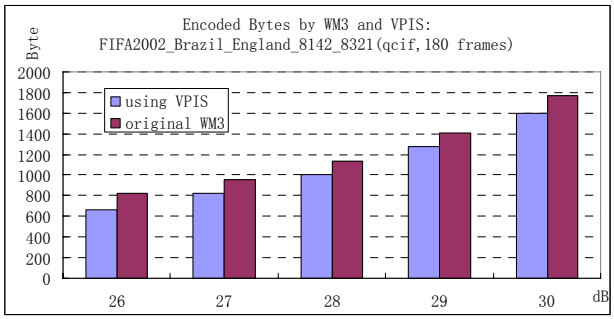

Figure 10. Encoded Bytes by WM3 and VPIS

Figure 10 shows the encoded data of 180 frames of a highlight(frame 8142~8321) in Video ID1 respectively by WM3 and VPIS in different values of PSNR. On average, about 13% of the bandwidth consumption could be reduced.

3) Optimized by ARC(Attention Region Coding)

Figure 11 reports the encoded data bytes of highlights of the four original videos respectively by WM3 and ARC. About 35.8% of the bandwidth consumption can be reduced.

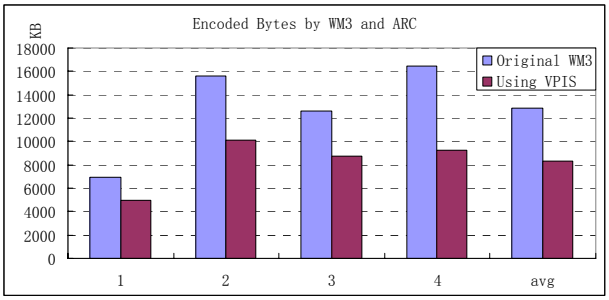

Figure 11. Encoded Bytes by WM3 and ARC

We also compare the PSNR of attention regions and non-attention regions and the results are shown in Figure 12. The PSNR values of attention regions are almost same in the two methods, and those of non-attention regions are decreased by several dBs with ARC. Moreover, the initial user study results demonstrated that nearly no one perceives the difference in the visual quality, because the PSNR-decreased regions don't attract the attention. The user study experiments are described in the following.

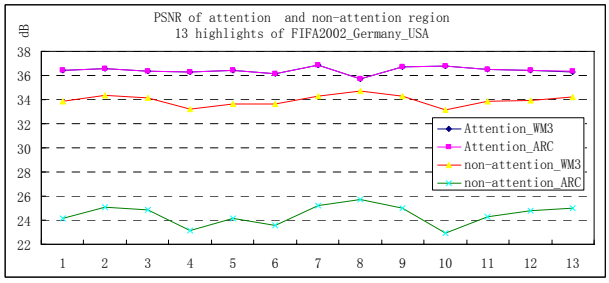

Figure 12. PSNR of attention and non-attention region.

3.3 User Study

We conducted a user study to evaluate the visual quality of video encoded by attention-region coding, which involved 8 computer-science students, who are familiar with the operations on mobile devices.

Firstly, they were taught to use the client browsing UI to view the sports video. Then we asked the eight users to view several pairs of sports video and each pair was encoded respectively by WM3 and ARC. For each pair, we fielded a questionnaire concerning the visual quality of each pair among the users to get their feedbacks: The two video sequences are the same in the visual quality, which should be answered on a five-scale likert scale where 1=strongly disagree and 5=strongly disagree. The result in Table II shows that our approach can keep the visual quality of video basically and can not affect users broadcasting sports video.

Table 2. The results of similarity in visual quality.

ID	HL1	HL2	HL3	HL4	HL5	HL6	AVG
VQ	3.63	4.13	4	3.88	4.5	4.13	4.04

VQ: visual quality; AVG: average of VQ in 6 HL(highlights)

4. CONCLUSIONS

This paper presents a novel framework that serves adaptive sports video to mobile users. We combine content-based sports highlights extraction and quality-domain video compression technologies in video server, capable of reducing wireless bandwidth consumption more effectively. In our system, a robust replay-based highlights extraction method is developed and a content-based video streaming coding scheme is proposed to handle the problems of bandwidth and capacity of computation. In our coding scheme, a variable-period Intra-frame selection based on shot detection and an attention region coding based on field and object segmentation are proposed. The experimental results demonstrate the robustness of our highlights extraction method and more than 77.5% of the bandwidth consumption could be reduced. We will apply our system to more sports videos for further investigations of its real effects.

5. ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China (Grant No. 60475010 and 60121302) and the National Key Basic Research and Development Program (973)(Grant No. 2004CB318107).

6. REFERENCES

[1] 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.

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

[3] 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.

[4] http://www.canalys.com/pr/2005/r2005102.htm

[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 IEEE Int'l Conf. on Acoustics, Speech, and Signal Processing, Orlando, Florida, May 2002

[6] J, Wang, E. Chng, and C. Xu, "Soccer replay detection using scene transition structure analysis", IEEE Intl Conference on Acoustics, Speech, and Signal Processing, Mar 2005.

[7] Liang Fan, Siwei Ma, Feng Wu, "Overview of AVS video standard", IEEE International Conference on Multimedia and Expo (ICME), vol. 1, pp 423-426, 2004.

[8] Liu, Q., Hua, Z., Zang, C., Tong, X., and Lu, H. 2005. "Providing on-demand sports video to mobile devices". In Proceedings of the 13th Annual ACM international Conference on Multime.

[9] 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.

[10] Rainer Lienhart, “Comparison of Automatic Shot Boundary Detection Algorithms”, In Image and Video Processing VII 1999, Proc. SPIE 3656-29, Jan. 1999.

[11] Tong Xiaofeng,Liu Qingshan, Lu Hanqing, and Jin Hongliang, “shot Classification in Sports Video”, Seventh Int¡¯l Conf. on Signal Processing (ICSP 04), Beijing, China, 2004.

[12] Xiao-Feng Tong, Han-Qing Lu, Qing-Shan Liu, “A Three-Layer Event Detection Framework and Its Application in Soccer Video”, accepted by Proc. of IEEE Int’l Conf. On Multimedia & Expo 2004, Taiwan, June 27-30, 2004.

[13] Zhang, H., Kankanhalli, A., and Smoliar, S.W. 1993. “Automatic partitioning of full-motion video”. Multimedia Syst. 1, 1 (Jan. 1993), 10-28.