Sports Video Retargeting

Liang Shi[1,2],Jinqiao Wang[1],Ling-yu Duan[3], Hanqing Lu[1]
[1]National Lab of Pattern Recognition,Institute of Automation, Chinese Academy of Sciences, Beijing 100190,China
[2]Beijing University of Posts and Telecommunications, Beijing 100876, China
[3]Institute of Digital Media, School of EE&CS, Peking University, Beijing 100871, China
freybupt@gmail.com,{jqwang, luhq}@nlpr.ia.ac.cn,lingyu@pku.edu.cn

ABSTRACT

With the proliferation of diverse multimedia terminals, the request for elegantly retargeting videos to different display devices is evident, especially in sports. This demonstration presents a Sports Video Retargeting(SVR) technique, that utilized domain based structure parsing to build a semantic importance map for video retargeting. The system enables flexible and coherent aspect-ratio change of the output sports videos with a spatial-temporal 3D rectilinear grid framework, which are free from significant loss of information or distortion on salient and important regions. Results in various sports type have shown that SVR is promising for content adaptation on mobile media.

Categories and Subject Descriptors

I.2.10 [**Vision and Scene Understanding**]: Video analysis

General Terms

Algorithms, Experimentation

Keywords

domain-based analysis, Video Retargeting

1. INTRODUCTION

Watching videos in diversified terminals such as televisions, notebooks, cellular phones, PDAs, PSPs and so on is becoming a popular way for communicating and sharing media information. Toward adapting source videos to these devices with proper resolutions and aspect ratios, video retargeting has attracted research efforts in recent years. As a sort of popular entertainment video, sports video has great commercial values. To satisfy various terminal users, video retargeting is dedicated to elegantly adapting videos to different display devices with clear and smooth imaging quality.

Two major problems are involved in video retargeting: (1) Aspect ratio change, e.g. mapping 4:3 videos to 16:9 wide screen of a HDTV set or iPhone; (2) Scale change, e.g. playing higher resolution videos on a smaller screen like PDA or mobile terminals. Aspect ratio change generates prominent deformation on human, face, and other perceptible objects, while scale change makes meaningful objects (e.g. ball in soccer) too small to be identified. Besides industrial techniques of squeezing, stretching, center cropping and black padding, region of interest based cropping [1] and content-aware resizing [4][3] have been adopted to solve these two problems. However, the performance is nevertheless far from users' requirement, especially in sports videos.

Figure 1: An example of sports video retargeting in different terminals, including TV, cell-phone, HDTV, PSP, iphone, and PDA.

This demonstration presents a technical implementation of the Sports Video Retargeting. Since structure information in most sports videos is much easier to obtain, we first parse the sports video hierarchically, then formulate a more accurate descriptor called semantic importance map(SIM). We argue that the objects' proportion and photographic invariance is perceptibly critical, and the visual continuity and smoothness is also essential for viewers' satisfaction. Our system is adaptable to both scale and aspect-ratio change.

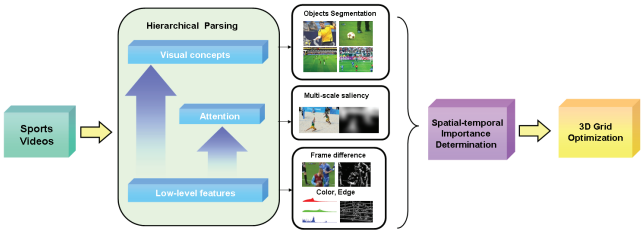

Figure 2: System overview.

MM'09, October 19–24, 2009, Beijing, China.
ACM 978-1-60558-608-3/09/10.

2. SYSTEM OVERVIEW

As shown in Fig.2, our system is composed of hierarchical semantic parsing and grid-based optimization. The former measures from low-level features to the domain-based objects in order to form a spatial-temporal importance map, and the latter adopt a time-consistent grid scaleplate to resize the frames non-homogeneously, which aims at maintaining the proportion of focused regions and ensure changeless regions to be still in a sequence of frames.

3. IMPLEMENTATION DETAILS

3.1 Structure Parsing of Sports Video

Structure analysis is an elementary step for semantic mining in video data. Although semantic structure parsing for general videos is difficult, it becomes much easier in sports videos with well-defined video structure, specialized sports syntax and fixed number of canonical view types.

In sports video, four types of shot size are usually used: long, medium, close-up and out-field. Since different shot types focus on different objects, domain knowledge is necessary for higher accuracy. In this paper, the shot type, playfield and in-field objects are chosen as typical features. Playfield usually contains one or two dominant colors, which are extracted in a long shot frame by taking the mean value of each color component around their respective histogram peaks. The statistics is computed at start-up then automatically applied it to the latter period of the video. Then, based on the field contour, we use color different with binary segmentation to extract in-field objects, as shown in Fig.3.

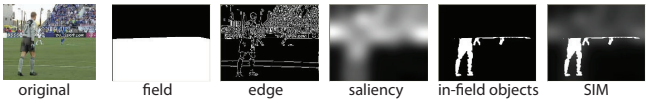

Figure 3: Semantic sports video parsing results.

3.2 Semantic Importance Map

Focus extraction and selection is critical for video retargeting, especially in sports videos. Low-level features including visual attention, gradient and motion cannot directly represent the semantic importance, thus are debatable in retargeting. In view of adaptive strategy for each shot type, we combine structure parsing results with edge detection and visual attention[2] under a naive Bayesian framework, and build the semantic importance map considering both domain knowledge and visual attractiveness, as shown in Fig.3.

3.3 Spatial-temporal 3D rectilinear grid

In view of semantic importance map, we build a 3D rectilinear grid as a scaleplate for adaptive video retargeting. We put more efforts to maintain proportions of important regions and ensure visual coherency. Both spatial and temporal distortions are considered: In the space domain, we redistribute distortions through out the frame; In the time domain, the grid lines are restrained by those of the former frame, so that the retargeted video will be visually smooth and continuous. Benefited from the small amount of grid marks (compared to pixel-mapping in [4]), the retargeted streaming can be optimized by gradient descent in real-time, as shown in Fig.4.

Figure 4: Grid optimization.

4. EXPERIMENT

In the test case, sports videos cover 15 types of sports including soccer, American football, basketball, volleyball, badminton, Ping-Pong ball, tennis, baseball, gymnastics, athletics, cycling, racing car, boxing, surfing and swimming, covering both indoor and outdoor, individual race and team event, typical sports types with varied proportion of sports field (grass playground, bi-color plastic surface, racetrack, swimming pool, indoor wrestling field, surface of ice, etc.) We collect 237 sports video clips, 30 seconds long in average, and 1780 key frames for test. With an AMD 1.6GHz dual core laptop with 1GB memory, typical videos by 448×336 could be processed in real-time.

Results indicate that our method are adaptable to diversified sport and shot types with varied scene layout, which supports downsizing, upsizing and change of aspect ratio. We also compare the results with cropping [1] and resizing [4]. Extensive evaluations by user study show the superiority of our SVR system.

5. CONCLUSIONS

We have proposed a unified system for sports video retargeting with invariant objects proportion and coherent visual effect. We integrated hierarchical semantic parsing and visual attention to build up a domain-based Semantic Importance Map as a more accurate descriptor. With a $3D$ time-consistent grid framework, video frames are optimized incorporating both spatial deformation and temporal discontinuity.

6. ACKNOWLEDGEMENTS

This work is supported by National Natural Science Foundation of China No. 60833006, Natural Science Foundation of Beijing No. 4072025, and 973 Program (Project No. 2010CB327900, 2010CB327905). Also this work was supported by National Basic Research Program of China under contract No. 2009CB320902, the research fund from National Laboratory of Pattern Recognition, Institute of Automation, CAS, and in part by the research award of Microsoft Research Asia Internet Services Theme.

7. REFERENCES

[1] F. Liu and M. Gleicher. Video retargeting automating pan and scan. In *Acm Multimedia*, 2006.

[2] Y. Ma and H. Zhang. Contrast-based image attention analysis by using fuzzy growing. In *ACM Multimedia*, 2003.

[3] Y.-S. Wang, C.-L. Tai, O. Sorkine, and T.-Y. Lee. Optimized scale-and-stretch for image resizing. In *ACM Trans. Graph*, 2008.

[4] L. Wolf, M. Guttmann, and D. Cohen-Or. Non-homogeneous content-driven video retargeting. In *ICCV*, 2007.