

Title	Automatic Summarization of Broadcasted Soccer Videos with Adaptive Fast-forwarding
Author(s)	Chen, Fan; De Vleeschouwer, Christophe
Citation	2011 IEEE International Conference on Multimedia and Expo (ICME): 1-6
Issue Date	2011-07
Type	Conference Paper
Text version	author
URL	<a href="http://hdl.handle.net/10119/10284">http://hdl.handle.net/10119/10284</a>
Rights	Copyright (C) 2011 IEEE. Reprinted from 2011 IEEE International Conference on Multimedia and Expo (ICME), 2011, 1-6. This material is posted here with permission of the IEEE. Such permission of the IEEE does not in any way imply IEEE endorsement of any of JAIST's products or services. Internal or personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution must be obtained from the IEEE by writing to <a href="mailto:pubs-permissions@ieee.org">pubs-permissions@ieee.org</a> . By choosing to view this document, you agree to all provisions of the copyright laws protecting it.
Description	

# AUTOMATIC SUMMARIZATION OF BROADCASTED SOCCER VIDEOS WITH ADAPTIVE FAST-FORWARDING

Fan Chen\*

Japan Adv. Insti. of Sci. and Tech.  
School of Information Science  
Nomi, Ishikawa, Japan  
Email: chen-fan@jaist.ac.jp

Christophe De Vleeschouwer†

Université catholique de Louvain  
ICTEAM  
Louvain-la-Neuve, Belgium  
Email: christophe.devleeschouwer@uclouvain.be

## ABSTRACT

When summarizing a video consisting of temporally continuous actions, e.g., surveillance videos or team-sport videos, fast-forwarding usually provides a better solution than content truncation to organize a semantically more complete and thus easily understandable story. Based on the resource allocation framework initially introduced in [1], we propose a fully automatic summarization system for broadcasted soccer videos, which supports adaptive fast-forwarding as well as content truncation. The proposed system involves two major contributions (i) an improved algorithm for refining the boundaries of far-view clips based on the extracted camera movement; (ii) a process to organize a summary by determining both the selected clips and their optimal fast-forwarding strategies, considering the scene-changing tempos of those clips and the given duration constraint. Experiments on real-life broadcasted soccer videos demonstrate the relevance and the efficiency of our proposed method.

**Index Terms**— Video Summarization, Resource Allocation, Adaptive Fast-Forwarding, Soccer Video Analysis

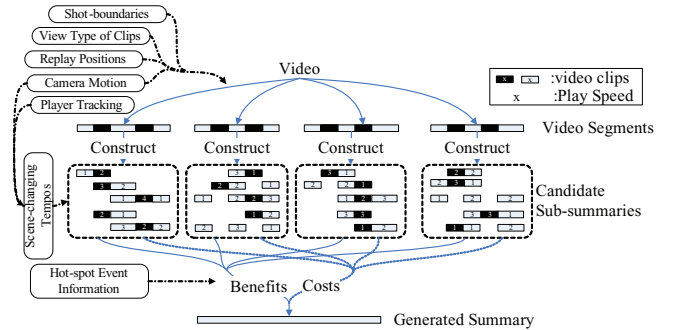
## 1. INTRODUCTION

Soccer video summarization has been widely studied in the past decades. Different from many previous works providing a solution for fast-browsing of highlighted moments[2][3], we intend to provide a concise video with well organized story-telling, from which the audience could enjoy the highlight moments of all events that best satisfy their interest. To achieve this purpose, two issues need to be considered: 1) Flexible organization of personalized summaries is required to handle various user preferences, both in terms of preferred action, and in terms of the desired length and narrative style of the summary. However, by investigating the literature, we found that personalization of story organization in most methods for sport video summarization was mainly focused on assignment of event significance [4] and extraction of specified view types [5], according to user preferences. 2) Story-telling artifacts caused by abrupt switching of stories should also be avoided to help the audience to get a comfortable viewing experience. Continuity of clips in generated summary was considered for better story-telling in [6]. [7] organizes stories by considering a graph model for managing semantic relations among concept entities. By controlling both local and global story-telling, [8] presented a resource allocation framework

for summarizing soccer videos with various personalization flexibilities of story-telling (including continuity, redundancy, and prevention of story-telling artifacts, etc.) as well as personalized retrieval of events-of-interest, which provides us a base for integrating new personalization functionalities.

Compared to the method explained in [8], our summarization system proposes two major improvements: (1) Besides clip selection, we also enable adaptive fast-forwarding of video contents, by analyzing the motion complexity of the current scene, so as to provide the audience a more complete story with suppressed visual discomfort caused by fast-forwarding; (2) Accordingly, we implement a method for refining the cutting of long far-view shots, so that the fast-forwarding is applied to a group of frames sharing similar motion complexities. Note that this refinement of far-view clip cutting also offers more natural and semantically meaningful truncation of shots than the one used in [8].

We will introduce the overall framework of the proposed summarization system in Section 2, along with explaining the required meta-data in Section 3 and the major processing steps in Section 4. Special emphasis is given to the explanation of the fast-forwarded display of long actions. We give experimental results in Section 5 and conclude it in Section 6.



**Fig. 1.** System overview. The whole process is envisioned in a divide and conquer paradigm.

## 2. OVERALL FRAMEWORK

We start with a simple explanation of the resource allocation based summarization, which was proposed in [8] for providing flexible personalized summarization of sport videos. This framework interprets the summarization problem as finding the optimal allocation of duration resources into video segments, according to the needs and

\*JAIST's contribution has been partly funded by the Japanese Grants-in-Aid for Scientific Research (Start-up Support).

†UCL's contribution has been partly funded by the European FP7 APIDIS project, the Walloon Region WALCOMO project, and the Belgian NSF.

preferences of the user. For an efficient processing, the whole process is envisioned in the divide and conquer paradigm, as depicted in Fig.1. The whole video is first cut into short clips by using a shot-boundary detector. These short clips are then organized into video segments. A sub-summary or local story is defined as one way to select clips within a segment. Several contending sub-summaries are then considered to present the segment, so that not only the content, but also the narrative style of the summary can be adapted to user requirements. By tuning the benefit and the cost of the sub-summaries, it is able to balance -in a natural and personal way- the semantic (what is included in the summary) and narrative (how it is presented to the user) aspects of the summary. The final summary is regarded as a collection of non-overlapping sub-summaries to maximize the overall benefit, under the given user-preferences and duration constraint.

A sport action usually consists of a sequence of several temporally closely related action units, e.g., several passings before the final shooting. When the action lasts long (which is often the case in the soccer video), it has to be cut into shorter clips for a finer summarization, where each clip only covers a part of the whole action. Missing of any intermediate clip of this action could result in the difficulty of game understanding, because the audience is then forced to reestablish himself in the game by inferring the semantic connection between the video contents before and after this clip. Besides clip selection, fast-forwarding provides another option to compress the time needed to render the action. When dealing with a less important clip, we prefer to include its fast-forwarded version than simply exclude it from the final summary, so as to reach a better balance between satisfaction of the overall temporal duration and generation of a smoother and more complete story. All those arguments motivate us to consider a summarization method with adaptive fast-forwarding.

Our system is different to the intelligent fast-forwarding system proposed in [9], in the following aspects: 1) We perform summarization as well as fast-forwarding. Temporal compression via fast-forwarding is upper bounded due to the existence of maximum tolerated playback speed, where summarization via content truncation is required to provide even shorter summaries; 2) Rather than directly assigning a unique fast-forward speed to each video clip, we allow personalized determination of optimal fast-forward speeds, through a global optimization of the overall benefit with respect to the specified narrative and semantic constraints, which hence provides more flexibility to satisfy various user preferences.

Our system builds on the same major processing steps as [8], i.e., it first performs video segmentation and then prepares candidate sub-summaries for each segment. The global story of the final summary is then organized by selecting a subset of non-overlapping sub-summaries. The major difference locates in that different fast-forwarded versions of clips will also be considered during the preparation of candidate sub-summaries. Accordingly, boundaries of far-view clips need to be further refined by grouping frames of similar motion complexity, so as to avoid mixing contents that should be presented at different fast-forwarding speeds. Furthermore, the benefit function should be revised to reflect this change, and additional meta-data (namely, camera motion and scene-changing tempo) need to be extracted to support this process.

### 3. META-DATA COLLECTION

As depicted in Fig.1, various meta-data are required at different stages of the proposed system. They include both production actions (position of replays, shot-boundaries, view types) and event infor-

mation. Production actions are exploited both to cut the video into semantically meaningful segments and to tune the benefits assigned to the various local stories of a segment. In contrast, event information acts at a global level, and directly impacts the (non)inclusion of segments in the summary. As done in [8], we also detect replays from producer-specific logos[10], extract shot-boundaries with a detector that has been developed in [11] to better deal with smooth transitions, recognize the view-type by using the method in [12], and automatically locate hot-spots by analyzing audio signals [13]. Detailed explanation of those methods are omitted due to page limitation. In this section, we focus on explaining the acquisition of two kinds of additional meta-data required by adaptive fast-forwarding, i.e., camera motion for refining the clip boundaries of action units, and scene-changing tempo for evaluating the scene motion complexity and thus determining the suitable fast-forwarding speed.

As a pure production action, camera motion itself carries no direct information about game events. However, since the general principles of sport video production prefer that the camera view always covers the ball and the dominant players, camera actions complying with those principles indirectly reflect the game situation, e.g. it is possible to infer the time of exchanging offensive/defensive roles by analyzing the direction and duration of camera motions. We extract the camera motion based on optical flow analysis, as depicted in Fig.2, where the camera motion in each frame is defined as the average of optical motion vectors from all grassland pixels.

Scene-changing tempo serves as a clue for guiding adaptive fast-forwarding. Rather than defining a scene-changing tempo to reflect the intensity of the real game (which hence makes this tempo different from the hot-spot), we are more interested in searching for a metric to evaluate the motion complexity of the scene, e.g., the fluctuation of the camera view or the diversified movement of multiple players. The contents with a higher scene-changing tempo prefer a lower play speed so that the audience can easily follow the game in a visually comfortable way, while the contents with a lower tempo prefer a higher play speed so as to save some temporal resources for rendering other important actions. Given a clip, the fluctuation of its camera view is evaluated by the standard deviation of the camera motion vectors, while the complexity of diversified player movements is defined as the average of standard deviation of players' moving speeds in each frame<sup>1</sup>. The scene-changing tempo is then defined as a weighted sum of the above two terms, as shown in Fig.3.

### 4. RESOURCE CONSTRAINED SUMMARIZATION

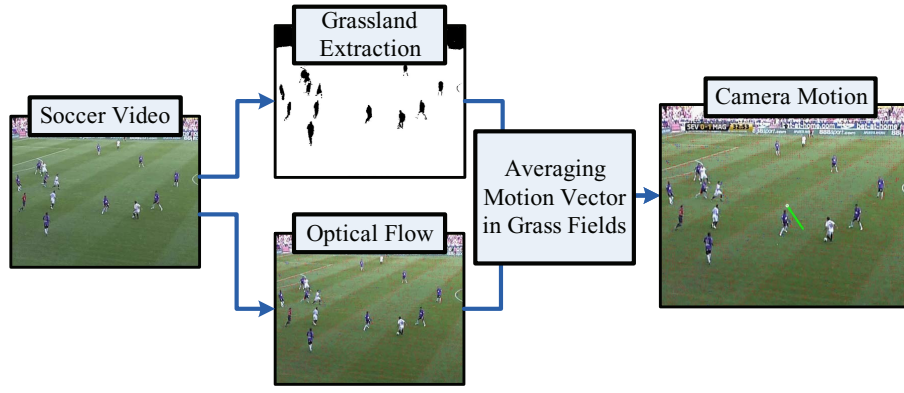
As explained in Section 2, we form the final summary by collecting at most one sub-summary for each segment, from a set of candidate sub-summaries. Let  $\mathbf{a}_{mk}$  denote the  $k^{th}$  sub-summary of the  $m$ -th segment. The overall benefit of the whole summary is defined as accumulated benefits of all selected sub-summaries, i.e.,

$$\mathcal{B}(\{\mathbf{a}_{mk}\}) = \sum_m \mathcal{B}(\mathbf{a}_{mk}) \quad (1)$$

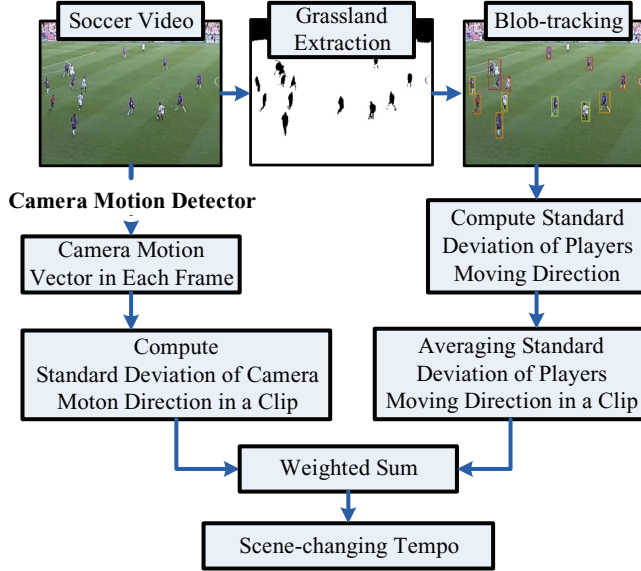
with  $\mathcal{B}(\mathbf{a}_{mk})$  being defined as a function of the user preferences, and of the highlighted moments. Our major task is to search for the set of sub-summaries index  $\{k^*\}$  that maximizes the total payoff  $\mathcal{B}(\{\mathbf{a}_{mk}\})$  under the length constraint  $\sum_m |\mathbf{a}_{mk}| \leq u^{LEN}$ , with  $u^{LEN}$  being the user-preferred length of the summary.

Under strict constraints, the problem is hard and relies on heuristic methods or dynamic programming approaches to be solved. In

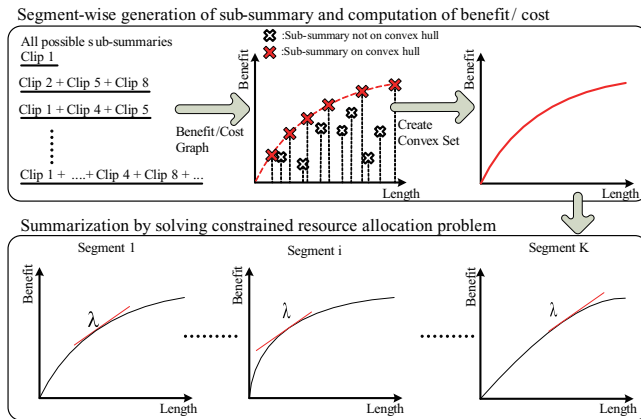
<sup>1</sup>Another advantage of using the standard deviation is that it avoids accurate compensation of player speed with respect to camera motions.



**Fig. 2.** Extraction of camera motion vectors by averaging motion vectors of grassland pixels computed from the optical flow analysis. The camera direction is represented by the green bold line starting from the center of the frame.



**Fig. 3.** Scene-changing tempo of a given far-view clip is computed from both camera motion and player motion, which reflects the variation speed of the scene in the clip.

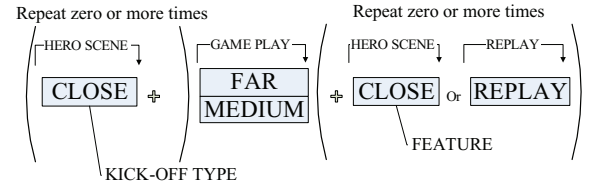


**Fig. 4.** Working flow in our summarization framework.

contrast, when some relaxation of the constraint is allowed, Lagrangian optimization and convex-hull approximation can be considered to split the global optimization problem in a set of simple block-based local decision problems [14, 15]. The convex-hull approximation consists in restricting the eligible summarization options for each sub-summary to the (benefit, cost) points sustaining the upper convex hull of the available (benefit, cost) pairs of the segment. Global optimization at the video level is then obtained by allocating the available duration among the individual segment convex-hulls, in decreasing order of benefit increase per unit of length [16]. This results in a computationally efficient solution that can still consider a set of candidate sub-summaries with various descriptive levels for each segment. In Fig.4, a diagram summarizes the working process of summary organization by solving a resource allocation problem.

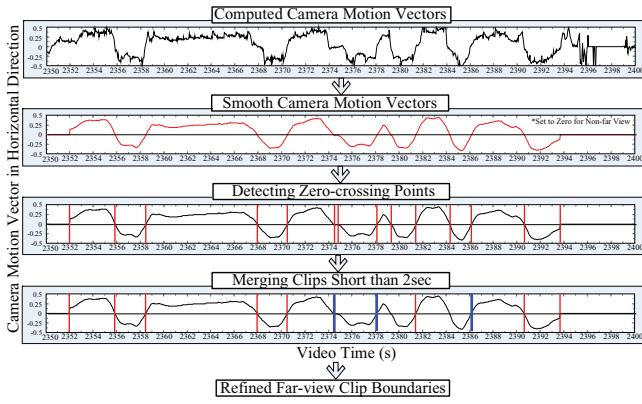
#### 4.1. Video segmentation

A video segment covers several successive clips closely related in terms of story semantic. By considering construction of sub-summaries in each segment independently, we trade-off summarization between efficiency of computation and controllability of story organization.



**Fig. 5.** Soccer game video segmentation is based on view-type sub-sequence matching.

Transitions in the state of the game motivate camera or view type switching, and are thus reflected in the production actions[17]. In fact view-types of clips has been used in many previous works, e.g., in [5] to help detect exciting events(i.e., game parts with both close-ups and replays). In [8], they defined a rule to segment the video based on the monitoring of the production actions, instead of (complex) semantic scene analysis tools. With the detected view-type (as Far-view/Closeup/replays) for all clips, they group successive clips following the view-structure in Fig.5.



**Fig. 6.** For far view, we refine the boundary of each clip, by analyzing the camera shifting direction of current scene.

A generic shot-boundary detector usually cannot deal with clip separation in a long far-view scene, which is required by a finer summarization. Here, we propose a method for refining the boundaries of far-view clips for two major purposes: 1) Although missing of intermediate clips is unavoidable, we hope that selected clips could be kept as locally intact as possible; 2) We intend to group scenes of similar motions together into a clip, so that adaptive fast-forwarding speed could be applied to different clips.

Along with the plot in Fig.6, we explain our method for clip boundary refinement. Only the horizontal component of camera motion vectors is used. After smoothing the curve, all zero-crossing points are detected, which stand for the changing times of attacking direction. If a period between two zero-crossing points is shorter than 2sec, we merge it into a neighbour clip of a shorter duration, because it could be a noise or a local direction change caused by dribbles.

#### 4.2. Local story organization with adaptive fast-forwarding

One major advantage of using the resource allocation framework is that it allows highly personalized nonlinear story organization via flexible definition of benefits. In the present paper, the benefit is defined as

$$\mathcal{B}(\mathbf{a}_{mk}) = \sum_{j \in \mathbf{a}_{mk}} \mathcal{I}_{mkj} \mathcal{G}(\mathbf{a}_{mk}, u^P, u^T) \mathcal{P}_{mk}^{CR}(u^C, u^R) \mathcal{P}_{mk}^F, \quad (2)$$

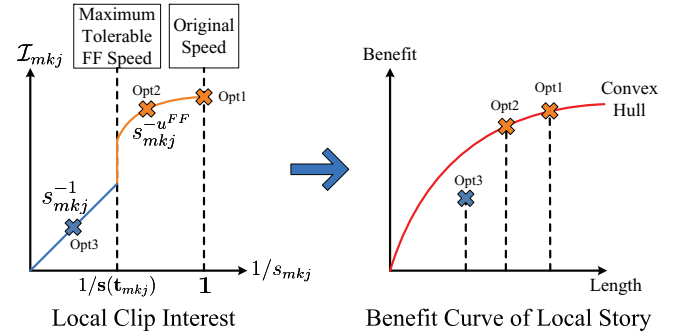
which includes accumulated semantical importance of selected clips  $\sum_{j \in \mathbf{a}_{mk}} \mathcal{I}_{mkj}$  and extra gain  $\mathcal{G}(\mathbf{a}_{mk}, u^P, u^T)$  from user favorite player  $u^P$  and team  $u^T$ , and also evaluates narrative preferences on story-telling (e.g., penalty  $\mathcal{P}_{mk}^{CR}$  on user specified story continuity  $u^C$ , and story redundancy  $u^R$ ). Satisfaction of general production principles is also evaluated through the penalty for forbidden cases  $\mathcal{P}_{mk}^F$ , to avoid frustrating visual/story-telling artifacts (e.g., over-short/incomplete local stories).

The appearance of a story-telling pattern in our final summary is firstly determined by its appearance in the convex hull of the benefit/cost curve of each segment. Hence, we achieve adaptive fast-forwarding by controlling the occurrence chance of various fast-forwarded versions in these convex hulls, which is implemented by tuning the convexity of local clip interest under various playback speed, as depicted in Fig.7.

Formally, we inject the adaptive fast-forwarding support into interest  $\mathcal{I}_{mkj}$ , which depends on the interest at normal speed  $\mathcal{I}_{mj}$ ,

$$\mathcal{I}_{mkj} = \begin{cases} \mathcal{I}_{mj}(1/s_{mkj})^{u^{FF}}, & s_{mkj} \leq s(t_{mkj}) \\ \mathcal{I}_{mj}/s_{mkj}, & s_{mkj} > s(t_{mkj}) \end{cases}, \quad (3)$$

where  $s_{mkj}$  stands for the selected fast-forwarding speed of clip  $j$  in sub-summary  $\mathbf{a}_{mk}$ .  $s(t_{mkj})$  defines the highest tolerable fast-forwarding speed under scene-changing tempo  $t_{mkj}$ . Four different speeds are enabled in current implementation, i.e., 1.0x, 1.5x, 2.0x and 2.5x. We normalize the  $t_{mkj}$  to range  $[0, 1]$  over all far-view clips, and then determine  $s(t_{mkj})$  by evenly distributing clips into four speed slots.  $u^{FF}$  is a positive parameter to control the user-preference on fast-forwarding, which needs to be set to  $u^{FF} \leq 1.0$  to assure convexity, so as to allow a fast-forwarded version to emerge as candidate sub-summaries in the convex hulls. When  $u^{FF} = 1.0$ , only normal speed play will be adopted. The lower the  $u^{FF}$  goes, the higher the fast-forwarding speeds are favored.



**Fig. 7.** The fast-forwarding support is implemented by revising the definition of local clip interest, where the new version encourages fast-forwarded clips under a given playback speed to emerge in the convex hull of local story organization, but suppresses those over the maximum tolerated speed in the produced story.

Finally, for close-up views, we set  $t_{mkj} = 0.5$ . For a replay clip, we only allow normal speed, since if we have time resources to render a replay, we expect it to present the action in the clearest way.

#### 4.3. Global story organization

The global duration resource is allocated among the available local sub-summaries to maximize the aggregated benefit. We solve this resource allocation problem by using the Lagrangian relaxation [16], whose main theorem reads that if  $\lambda$  is a non-negative Lagrangian multiplier and  $\{k^*\}$  is the optimal set that maximizes

$$\mathcal{L}(\{k\}) = \sum_m \mathcal{B}(\mathbf{a}_{mk}) - \lambda \sum_m |\mathbf{a}_{mk}| \quad (4)$$

over all possible  $\{k\}$ , then  $\{\mathbf{a}_{mk^*}\}$  maximizes  $\sum_m \mathcal{B}(\mathbf{a}_{mk})$  over all  $\{\mathbf{a}_{mk}\}$  such that  $\sum_m |\mathbf{a}_{mk}| \leq \sum_m |\mathbf{a}_{mk^*}|$ . Hence, if  $\{k^*\}$  solves the unconstrained problem in Eq.(4), then it also provides the optimal solution to the constrained problem in Eq.(1), with  $u^{LEN} = \sum_m |\mathbf{a}_{mk^*}|$ . Since the contributions to the benefit and cost of all segments are independent and additive, we can write

$$\sum_m \mathcal{B}(\mathbf{a}_{mk}) - \lambda \sum_m |\mathbf{a}_{mk}| = \sum_m (\mathcal{B}(\mathbf{a}_{mk}) - \lambda |\mathbf{a}_{mk}|). \quad (5)$$



From the curves of  $\mathcal{B}(\mathbf{a}_{mk})$  with respect to their corresponding summary length  $|\mathbf{a}_{mk}|$ , the collection of points maximizing  $\mathcal{B}(\mathbf{a}_{mk}) - \lambda|\mathbf{a}_{mk}|$  with a same slope  $\lambda$  produces one unconstrained optimum. Different choices of  $\lambda$  lead to different summary lengths. If we construct a set of convex hulls from the curves of  $\mathcal{B}(\mathbf{a}_{mk})$  with respect to  $|\mathbf{a}_{mk}|$ , we can use a greedy algorithm to search for the optimum under a given constraint  $u^{\text{LEN}}$ . The approach is depicted in Fig.4 and explained in details in [15]. In short, for each point in each convex hull, we first compute the forward (incremental) differences in both benefits and summary-lengths. We then sort the points of all convex-hulls in decreasing order of  $\lambda$ , i.e., of the increment of benefit per unit of length. Given a length constraint  $u^{\text{LEN}}$ , ordered points are accumulated until the summary length gets larger or equal to  $u^{\text{LEN}}$ . Selected points on each convex-hull define the sub-summaries for each segment.

## 5. EXPERIMENTAL RESULTS

We use a soccer video along with a list of 50 automatically extracted audio hot-spots in the following experiments. We only provide some representative results here. Their corresponding videos and more experimental results are available in the supplemental material associated to this paper [18].

We first present some results of meta-data extraction. We then investigate the behavior of the proposed system, and assess the performance by analyzing its output results and compare them to those generated by the previous method in [8].

### 5.1. Results on meta-data extraction

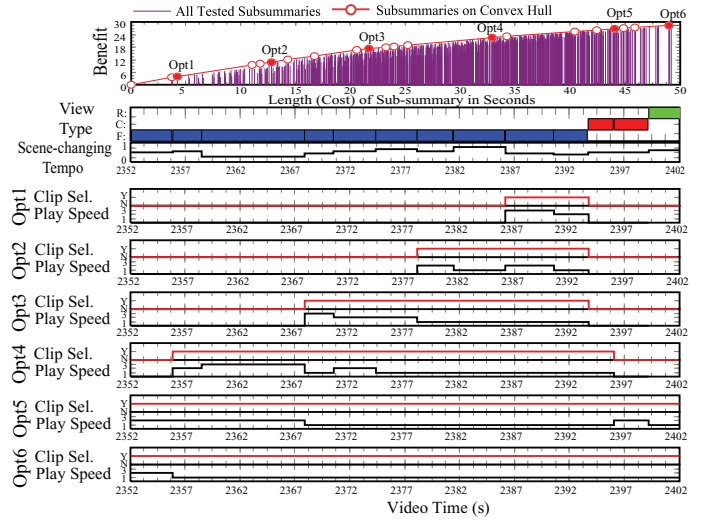
In the supplemental material, we provide video results of optical flow, camera motion detection, and player tracking. Scene-changing tempos are plotted along with the summarization results later.

In Fig.6, we plotted the estimated camera motion and refined far-view clip boundaries. Due to the intrinsic slow pace of soccer (comparing to other games, e.g., basketball) and the general principle of production to suppress unnecessary camera motion, camera shifting usually sticks to one moving direction for a certain period. As we can see from Fig.6, the far-view shot is cut into a limited and thus processable number of clips. When two consecutive long clips are different in camera shifting directions, it usually means that two teams have exchanged their offensive/defensive roles in the game. Hence, by checking the corresponding videos, we found that it also provides a good cut of an action in a semantic way.

Due to the lacking of ground-truth data for camera motion or far-view boundary refinement (which are not easy to get, because it not only needs tedious manual annotation, but also requires subjective evaluation to validate the manually annotated data), no quantitative results are provided here. The reviewers are invited to access the videos in the supplemental material[18] to make their own opinions.

### 5.2. Behavior of the summarization system

We investigate the behavior of the proposed system by inspecting the way the optimal sub-summary grows with the increasing duration constraint. In Fig.8, we plot benefit-cost convex-hulls computed for a sample segment. Below the convex-hull, a graph shows the boundaries of clips in this segment along with their corresponding view-types and scene-changing tempos. Six different duration constraints lead to six different optimal sub-summaries, plotted in the six subgraphs. In each sub-graph, the upper part shows the selection status of a clip, and the lower part shows its fast-forwarding speed. We have three major observations from Fig.8:



**Fig. 8.** Benefit-cost convex-hulls computed for a sample segment. Below the convex-hull, a graph shows the boundaries of clips in this segment along with their corresponding view-types and scene-changing tempos. Six different duration constraints lead to six different optimal sub-summaries, plotted in the six subgraphs. In each sub-graph, the upper part shows the selection status of a clip, and the lower part shows its fast-forwarding speed. (Play Speed: 1=1.0x, 2=1.5x, 3=2.0x)

a.) A global tendency is that the overall play speed decreases when the duration constraint increases, so as to satisfy both a busy audience to fast-browse the game and an audience with more leisure to enjoy the video.

b.) The second observation is that a lower scene-changing tempo favors more on a higher play speed and a higher one takes in a lower play speed, as reflected, e.g., by 2378s-2387s in Opt2, 2367s-2378s in Opt3, and 2355s-2367s in Opt4.

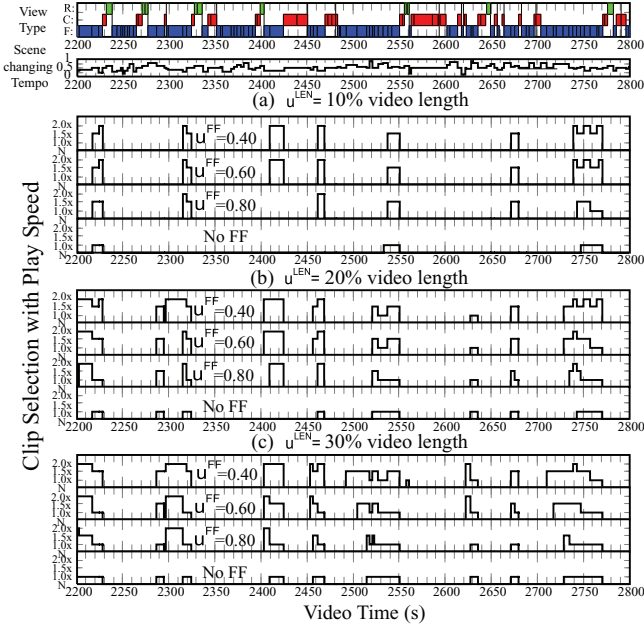
c.) However, play speed does not always coincide with the scene-changing tempo, e.g., the result from 2387s-2392s in Opt1. The reason is that the play speed is not only affected by the scene-changing tempo, but is also determined by the base interest of a clip. The base interest of a clip is computed from its temporal distance to the end of the action. Hence, although the clip between 2391s-2393s has a lower scene-changing tempo than its preceding clip, any speed up of this clip will result in a higher benefit lost, which prevents it from taking a high fast-forwarding speed. We regard this consequence as a reasonable and positive effect, because a clip having a higher interest deserves more chances to be presented in a normal speed.

### 5.3. Performance of the proposed system

Finally, we inspect the performance of the proposed system in generating a global story with adaptive fast-forwarding. In Fig.9, we plot the summaries resulting from our fast-forwarding enabled method. The first subgraph presents the view-structure of segments, while the second gives the computed scene-changing tempos. Four cases are compared, i.e., our method with  $u^{\text{FF}} = 0.40, 0.60$  and  $0.80$  to the method in [8] (NO FF), under three duration constraints, i.e.,  $u^{\text{LEN}} = 10\%, 20\%$  and  $30\%$  of the video length.

It is easy to find that both a smaller  $u^{\text{FF}}$  and a smaller  $u^{\text{LEN}}$  intend to increase the overall play speed. Hence, it provides us a way to personalize the fast-forwarding functionality by tuning  $u^{\text{FF}}$ . Fur-

thermore, by saving the duration resources through fast-forwarding, it is able to either include more actions, such as three actions within 2300s-2500s for  $u^{LEN} = 10\%$ , or complete an action, e.g., 2450s-2480s in the result of  $u^{LEN} = 30\%$ . All these demonstrate the advantage of including adaptive fast-forwarding into summarization.



**Fig. 9.** We plot the summaries resulting from our fast-forwarding enabled method. The first subgraph presents the view-structure of segments, while the second gives the computed scene-changing tempos. Four cases are compared, i.e., our method with  $u^{FF} = 0.40$ , 0.60 and 0.80 to the method in [8] (NO FF), under three duration constraints, i.e.,  $u^{LEN} = 10\%$ , 20% and 30% of the video length.

We have built up a base for personalized video summarization with adaptive fast-forwarding, and validated its efficiency in reaching our design goals, which is the major contribution of the present work. Further subjective evaluations are still required to explore the optimal fast-forwarding speed in the sense of user perception, and to refine the relationship between the fast-forwarding speed and the scene-changing tempo. They are left as our future work.

## 6. CONCLUSIONS

We proposed a framework for producing personalized summaries of broad-casted soccer videos, including both clip selection and adaptive fast-forwarding. By analyzing the soccer video, we defined the metrics for evaluating camera motion and scene-changing tempos, so as to determine the optimal cutting points within a long far-view shot, and to provide referential clues on optimal play speed in the generated summary. Instead of a rigid determination of fast-forwarding speed, we efficiently select the optimal combination from candidate summaries by solving a resource-allocation problem. Experiments demonstrate the efficiency of the proposed system. All these exploits a way to summarize a video consisting of actions with strong temporal continuity by combining both video truncation and fast-forwarding, which is not only applicable to sport video game, but also meaningful in many other application scenarios such as surveillance.

## 7. REFERENCES

- [1] F. Chen and C. De Vleeschouwer, "A resource allocation framework for summarizing team sport videos," in *ICIP'09*, 2009, pp. 4349–4352.
- [2] B.L. Tseng and J.R. Smith, "Hierarchical video summarization based on context clustering," in *Internet Multimedia Management Systems IV: Proceedings of SPIE*, Nov. 2003, vol. 5242, pp. 14–25.
- [3] Zhu Li, Guido M. Schuster, and Aggelos K. Katsaggelos, "Minmax optimal video summarization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, pp. 1245–1256, 2005.
- [4] N. Babaguchi, Y. Kawai, T. Ogura, and T. Kitahashi, "Personalized abstraction of broadcasted american football video by highlight selection," *Multimedia, IEEE Transactions on*, vol. 6, no. 4, pp. 575 – 586, 2004.
- [5] B. Li, H. Pan, and I. Sezan, "A general framework for sports video summarization with its application to soccer," in *ICASSP*, Hong-Kong, April 2003, pp. 169–172.
- [6] M. Albanese, M. Fayzullin, A. Picariello, and V.S. Subrahmanian, "The priority curve algorithm for video summarization," *Information Systems*, vol. 31, no. 7, pp. 679 – 695, 2006.
- [7] B.W. Chen, J.C. Wang, , and J.F. Wang, "A novel video summarization based on mining the story-structure and semantic relations among concept entities," *IEEE Trans. on Multimedia*, vol. 11, no. 2, pp. 295–312, Feb. 2009.
- [8] F. Chen, C. De Vleeschouwer, H.D. Barrobes, J.G. Escalada, and D. Conejero, "Automatic and personalized summarization of audio-visual soccer feeds," in *ICME'10*, 2010, pp. 837–842.
- [9] S.B. Jun, K.R. Yoon, J.E. Lee, and S.Y. Lee, "Intelligent video system," in *United States Patent Application No. 20020039481*, 2002.
- [10] H. Pan, P. van Beek, and M.I. Sezan, "Detection of slow-motion replay segments in sports video for highlights generation," in *ICASSP'01*, 2001.
- [11] I.A. Fernandez, F. Chen, F. Lavigne, X. Desurmont, and C. De Vleeschouwer, "Browsing sport content through an interactive h.264 streaming session," in *MMEDIA 2010*, 2010, pp. 155 –161.
- [12] A. Ekin, A.M. Tekalp, and R. Mehrotra, "Automatic soccer video analysis and summarization," *IEEE Trans. on Image Process.*, vol. 12, no. 7, pp. 796–807, July 2003.
- [13] H. Duxans, X. Anguera, and D. Conejero, "Audio based soccer game summarization," in *BMSB'09*, 2009, pp. 1–6.
- [14] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 36, no. 9, pp. 1445 –1453, Sept. 1988.
- [15] A. Ortega, "Optimal bit allocation under multiple rate constraints," in *DCC'96*, 1996, pp. 349 –358.
- [16] H. Everett, "Generalized lagrange multiplier method for solving problems of optimum allocation of resources," *Operations Research*, vol. 11, no. 3, pp. 399–417, 1963.
- [17] J. Owens, "Tv sports production," *Focal Press*, 2007.
- [18] Additional results on web, "http://www.jaist.ac.jp/~chen-fan/apidis/www/icme2011/main.htm," 2011.