# Personalized Summarization of Broadcasted Soccer Videos with Adaptive Fast-forwarding

Fan Chen[1] and Christophe De Vleeschouwer[2]

[1] Japan Advanced Institute of Science and Technology, Nomi 923-1211, Japan,
chen-fan@jaist.ac.jp
[2] Université catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium,
christophe.devleeschouwer@uclouvain.be

**Abstract.** We propose a hybrid personalized summarization framework that combines adaptive fast-forwarding and content truncation to generate comfortable and compact video summaries. We formulate video summarization as a discrete optimization problem, where the optimal summary is determined by adopting Lagrangian relaxation and convex-hull approximation to solve a resource allocation problem. Subjective experiments are performed to demonstrate the relevance and efficiency of the proposed method.

**Key words:** Personalized Video Summarization, Adaptive Fast-forwarding, Soccer Video Analysis

## 1 Introduction

Video summarization techniques address different purposes, including fast browsing [6], retrieval [14], behaviour analysis [15], and entertainment. We intend to generate from the source video(s) a concise version with well organized story-telling, from which the audience can enjoy the contents that best satisfy their interest. Two kinds of information are essential for producing semantically relevant and enjoyable summaries: *Semantic information* of the scene directly evaluates the importance of frames for producing semantically relevant summaries; *Scene activity* is associated to the changes of the scene presented to the audience. Conventional content-truncation-based methods mainly maximize the semantic information associated to the content played during the constraint browsing period, e.g. using fast-browsing of highlighted moments [10]. However, semantic information extracted from individual images/segments fails to model a complicated story-telling with strong dependency in its contents. In contrast, conventional fast-forwarding-based methods mainly sample the video frames at a rate that increases with the measured scene activity, defined via optical flow [13] or the histogram of pixel differences [8]. By only evaluating changes in the scene, it is difficult to assure the semantic relevance of the summary. The application of pure fast-forwarding based methods is also constrained by the fact the highest tolerable playback speed is bounded due to the limitation of visual perception [9].

We thus propose an approach that truncates contents with intolerable playback speeds and saves time resources for better rendering the remaining contents. We design a hybrid summarization method with both content truncation and

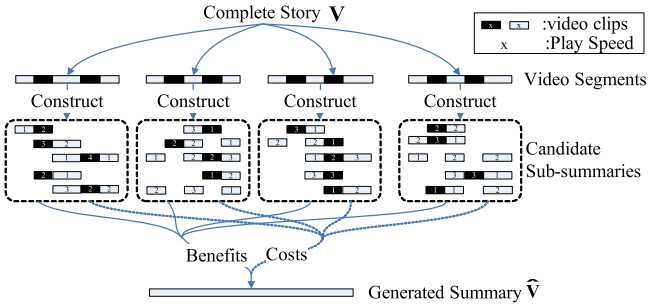



**Fig. 1.** Conceptual diagram of the overall proposed summarization process envisioned in a divide and conqure paradigm.

adaptive fast-forwarding to provide continuous as well as semantically relevant summaries with improved visual comfort. We select playback speeds from a set of discrete options, and introduce a hierarchical summarization framework to find the optimal allocation of time resources into the summary, which enables various story-telling patterns for flexible personalized video summarization. Our resource allocation summarization in [2] only considers content truncation. In [1], we only considered the semantic information of pre-defined video segments in evaluating the benefit of the summary, and scene activity was only adopted for heuristically determining the maximum tolerable speed, in an independent process from the resource allocation optimization. In the present paper, we determine both the truncation actions and the playback speeds in a soft way through a unified resource allocation process that considers both semantic information and scene activity. Furthermore, subjective tests are now performed to validate the adaptive fast-forwarding principle.

The paper is organized as follows. In Section 2 we introduce the proposed summarization framework. In Section 3, we present experimental results. Finally, we conclude the paper in Section 4.

## 2 Resource Allocation Framework

Our resource-allocation-based framework interprets the summarization problem as finding the optimal allocation of duration resources $u^{\mathrm{L}}$ into video segments, according to various user preferences. We design the whole process using the divide and conquer paradigm (Fig.1(a)). 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 defines one way to select clips within a segment. Several sub-summaries can be generated from a segment: 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 sub-summaries, we balance -in a natural and personal way- the semantics (what is included in the summary) and the narrative (how it is presented to the user) of the summary. The final summary is formed by collecting non-overlapping sub-summaries to maximize the overall benefit, under the user-preferences and duration constraint.

Let the video be cut into $N^{\mathrm{C}}$ clips, with the $i^{th}$ clip $\mathcal{C}_i$ being $\mathcal{C}_i = \{t|t = t_i^S, \cdots, t_i^E\}$. $t_i^S$ and $t_i^E$ are the index of its starting and ending frames. These video clips are grouped into $M$ segments. A set of candidate sub-summaries is considered for each segment, from which at most one sub-summary can be selected. We denote the $k^{th}$ sub-summary of the $m$-th segment $\mathcal{S}_m$ as $\mathbf{a}_{mk}$, which is a set of playback speeds for all its clips, i.e., $\mathbf{a}_{mk} = \{v_{ki}|i \in \mathcal{S}_m\}$. $v_{ki}$ is the playback speed assigned to the $i^{th}$ clip if the $k^{th}$ sub-summary $\mathbf{a}_{mk}$ is adopted.

Let $\mathbf{b}_m = \{b_i|i \in \mathcal{S}_m\}$ be the list of base benefits for all clips in $\mathcal{S}_m$. Our major task is to find the set of sub-summaries that maximizes the total payoff

$$\hat{\mathbf{V}}^* = \arg\max_{\hat{\mathbf{V}}} \mathcal{B}(\{\mathbf{a}_{mk}\}|\{\mathbf{b}_m\}), \tag{1}$$

subject to $\sum_{m=1}^{M} |\mathbf{a}_{mk}| \leq u^{\mathrm{L}}$. We define $|\mathbf{a}_{mk}|$ as the length of summary $\mathbf{a}_{mk}$,

$$|\mathbf{a}_{mk}| = \sum_{i \in \mathcal{S}_m} \frac{t_i^E - t_i^S}{v_{ki}}. \tag{2}$$

The overall benefit of the whole summary is defined as accumulated benefits of all selected sub-summaries:

$$\mathcal{B}(\{\mathbf{a}_{mk}\}|\{\mathbf{b}_m\}) = \sum_{m=1}^{M} \mathcal{B}_m(\mathbf{a}_{mk}) \tag{3}$$

with $\mathcal{B}_m(\mathbf{a}_{mk})$ being defined as a function of the user preferences, of the highlighted moments, and of the playback speeds as described in the following.

### 2.1 Video Segmentation

We divide the soccer video into clips, according to the detected production actions, such as position of replays, shot-boundaries and view types. We detect replays from producer-specific logos [12], extract shot-boundaries with a detector proposed in [7] to better deal with smooth transitions, and recognize the view-type by using the method in [4]. We segment the video based on the monitoring of production actions by analysing the view-structure [2] instead of using (complex) semantic scene analysis tools.

### 2.2 Local Story Organization

One major advantage of the resource allocation framework is that it allows highly personalized story organization, which is achieved via flexible definition of benefits. We define the benefit of a sub-summary as

$$\mathcal{B}_m(\mathbf{a}_{mk}) = \sum_{i \in \mathcal{S}_m} \mathcal{B}_i(v_{ki}) \mathcal{B}_{mi}^{\mathrm{P}}(\mathbf{a}_{mk}), \tag{4}$$

which includes accumulated benefits of selected clips. $\mathcal{B}_i(v_{ki})$ computes the base benefit of clip $i$ at playback speed $v_{ki}$,

$$\mathcal{B}_i(v_{ki}) = b_i(1/v_{ki})^{\beta}. \tag{5}$$

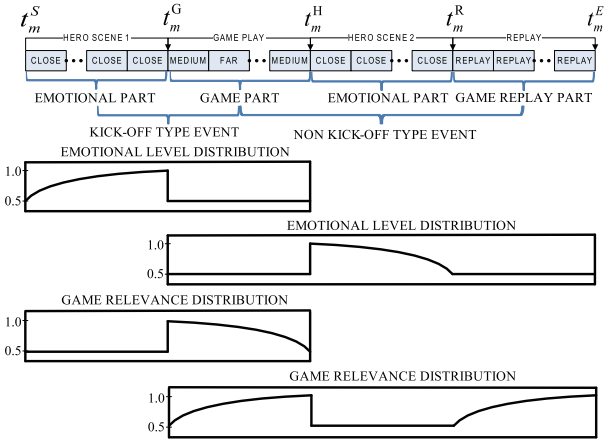



**Fig. 2.** The base benefit of a clip is evaluated from the game relevance and emotional level, defined as functions of clip view-types. The decaying process is modelled by hyperbolic tangent function. $t_m^G$, $t_m^H$, $t_m^R$ are starting times of game play, hero scene, and replay in the $m$-th segment, respectively.

$b_i$ is the clip benefit, defined as

$$b_i = |t_i^E - t_i^S| \left(\overline{f_t}\,\overline{a_t}\right)^{\alpha}, \tag{6}$$

which consider the average semantic information $\overline{f_t}$ and scene activity $\overline{a_t}$. As in [1], we automatically locate hot-spots by analyzing audio signals [3], whose (change of) intensity is correlated to the semantic importance of each video segment. The benefit of each frame $t$ within each segment is further evaluated from its relevance to the game $f_t^G$ and its level of emotional involvement $f_t^E$. The frame information $f_t$ is computed as

$$f_t = 0.25 f_t^E + 0.75 f_t^G. \tag{7}$$

$f_t^G$ mainly evaluates the semantic relevance of a clip in presenting the game progress, while $f_t^E$ evaluates the importance of a clip in revoking the emotional involvement of the audience, e.g. via closeup view of a player. Hence, the above fixed weight favours game related contents in the summary. We define $f_t^G$ and $f_t^E$ by propagating the significance of the detected hot-spot event according to the view type structure of the segment, as depicted in Fig.2. The decaying process was modelled by using the hyperbolic tangent function, because it is bounded and is integrable thus simplifying the computation of $\overline{f_t}$.

Scene activity $a_t$ is defined on the fluctuation of the camera view or the diversified movement of multiple players. Given a clip, the fluctuation of its camera view $\overline{\tau^{\mathrm{M}}}$ is evaluated by the average standard deviation of the motion vectors in the clip, while the complexity of diversified player movements $\overline{\tau^{\mathrm{P}}}$ is defined as the average standard deviation of players' moving speeds in the clip. As shown in Fig.3, the average information $\overline{a_t}$ is then defined as a weighted sum of the above two terms,

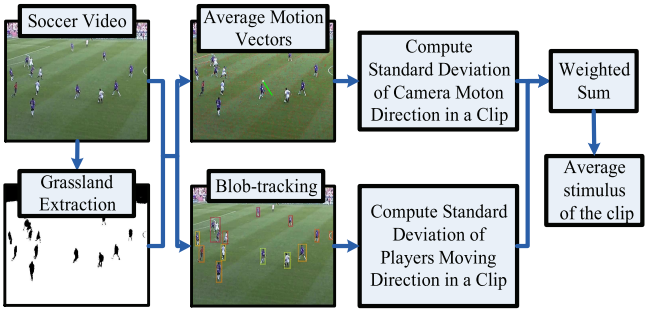



**Fig. 3.** We evaluate the average stimulus in a far-view clip by estimating information associated to scene activity from camera motion and player motion, which are computed on average motion vector in the grassland region and tracked player positions.

$$\overline{a_t} \propto \begin{cases} \overline{\tau^{\mathrm{M}}} + \overline{\tau^{\mathrm{P}}}, & \text{far view} \\ \overline{\tau^{\mathrm{M}}}, & \text{otherwise} \end{cases} \tag{8}$$

which is normalized to $[0 \quad 1]$ for far-view and non-far-view clips independently. Using the standard deviation avoids the need of accurate compensation of player speed with respect to camera motions. $\mathcal{B}^{\mathrm{P}}_{mi}(\mathbf{a}_{mk})$ evaluates the extra benefits by satisfying specific preferences:

$$\mathcal{B}^{\mathrm{P}}_{mi}(\mathbf{a}_{mk}) = \mathcal{P}^{\mathrm{O}}(v_{ki}, u^{\mathrm{O}})\mathcal{P}^{C}_{mki}(u^{C})\mathcal{P}^{F}_{mk}. \tag{9}$$

$\mathcal{P}^{\mathrm{O}}(v_{ki}, u^{\mathrm{O}})$ is the extra gain obtained by including user's favorite object $u^{\mathrm{O}}$ specified through an interactive interface,

$$\mathcal{P}^{\mathrm{O}}(v_{ki}, u^{\mathrm{O}}) = \begin{cases} 1.5, v_{ki} < \infty, \exists t \in \mathcal{C}_i, u^{\mathrm{O}} \text{exists in} I_t, \\ 1.0, \text{otherwise.} \end{cases} \tag{10}$$

We favour a continuous story-telling by defining $\mathcal{P}^{C}_{mki}(u^{C})$

$$\mathcal{P}^{C}_{mki}(u^{C}) = 1 + u^{C}(2 - \delta_{\frac{1}{v_{ki}v_{k(i+1)}},0} - \delta_{\frac{1}{v_{ki}v_{k(i-1)}},0}), \tag{11}$$

where $\delta_{a,b}$ is the Kronecker delta function, and $u^{C}$ is fixed to 0.1 in our experiments. Satisfaction of general production principles is also evaluated through $\mathcal{P}^{F}_{mk}$, which takes 1 for normal case and 0.001 for forbidden cases (or a value that is small enough to suppress this case from being selected), to avoid unpleasant visual/story-telling artifacts (e.g. too-short/incomplete local stories). We only allow normal speed for a replay clip in local story organization. If time resources to render a replay are available, we present the action in the clearest way.

### 2.3 Global Story Organization

The global-duration resource is allocated among the available sub-summaries to maximize the aggregated benefit (Eq.1). When relaxation of constraints are allowed, Lagrangian optimization and convex-hull approximation can be considered to split the global optimization problem in a set of simple block-based

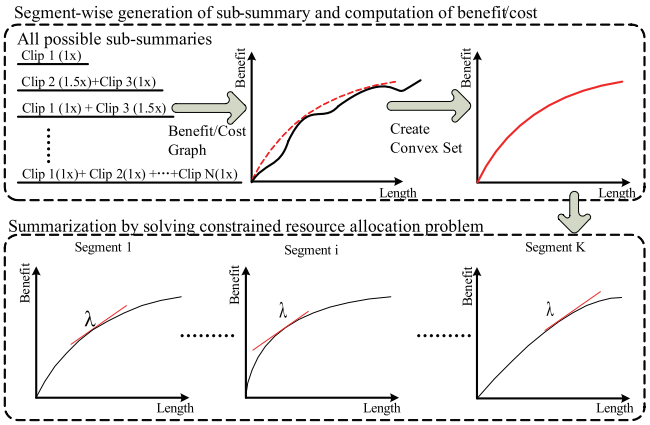



**Fig. 4.** Lagrangian relaxation and convex-hull approximation are adopted to solve the resource allocation problem, which restrict the eligible summarization options to the convex hulls of benefit-to-cost curves of the segments, where the collection of points from all convex-hulls with a same slope $\lambda$ produces one optimal solution under the corresponding summary length.

decision problems [11]. The convex-hull approximation restricts 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 is obtained by allocating the available duration among the individual segment convex-hulls [5], which results in a computationally efficient solution. Fig.4 summarizes the summarization process.

We solve this resource allocation problem by using the Lagrangian relaxation [5]: if $\lambda$ is a non-negative Lagrangian multiplier and $\{k^*\}$ is the optimal set that maximizes

$$\mathcal{L}(\{k\}) = \sum_{m=1}^{M} \mathcal{B}_m(\mathbf{a}_{mk}) - \lambda \sum_{m=1}^{M} |\mathbf{a}_{mk}| \tag{12}$$

over all possible $\{k\}$, then $\{\mathbf{a}_{mk^*}\}$ maximizes $\sum_{m=1}^{M} \mathcal{B}_m(\mathbf{a}_{mk})$ over all $\{\mathbf{a}_{mk}\}$ such that $\sum_{m=1}^{M} |\mathbf{a}_{mk}| \leqslant \sum_{m=1}^{M} |\mathbf{a}_{mk^*}|$. Hence, if $\{k^*\}$ solves the unconstrained problem in Eq.12, then it also provides the optimal solution to the constrained problem in Eq.1, with $u^{\mathrm{L}} = \sum_{m=1}^{M} |\mathbf{a}^*_{mk}|$. Since the contributions to the benefit and cost of all segments are independent and additive, we can write

$$\sum_{m=1}^{M} \mathcal{B}_m(\mathbf{a}_{mk}) - \lambda \sum_{m=1}^{M} |\mathbf{a}_{mk}| = \sum_{m=1}^{M} (\mathcal{B}_m(\mathbf{a}_{mk}) - \lambda |\mathbf{a}_{mk}|). \tag{13}$$

From the curves of $\mathcal{B}_m(\mathbf{a}_{mk})$ with respect to their corresponding summary length $|\mathbf{a}_{mk}|$, the collection of points maximizing $\mathcal{B}_m(\mathbf{a}_{mk}) - \lambda |\mathbf{a}_{mk}|$ 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 $\mathcal{B}_m(\mathbf{a}_{mk})$ with respect to $|\mathbf{a}_{mk}|$, we can use a greedy algorithm to search for the

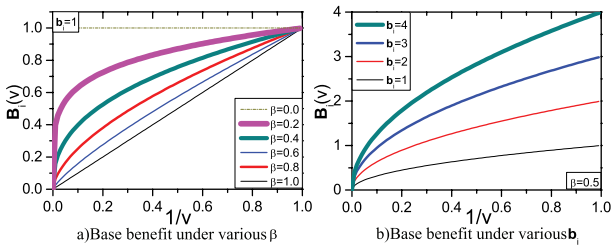



**Fig. 5.** Clip benefit complies with convex-hull approximation and the greedy algorithm adopted for solving the resource allocation problem.

optimum under a given constraint $u^{\mathrm{L}}$. The approach is depicted in Fig.4 and explained in details in [11]. 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. Ordered points are accumulated until the summary length gets larger or equal to $u^{\mathrm{L}}$.

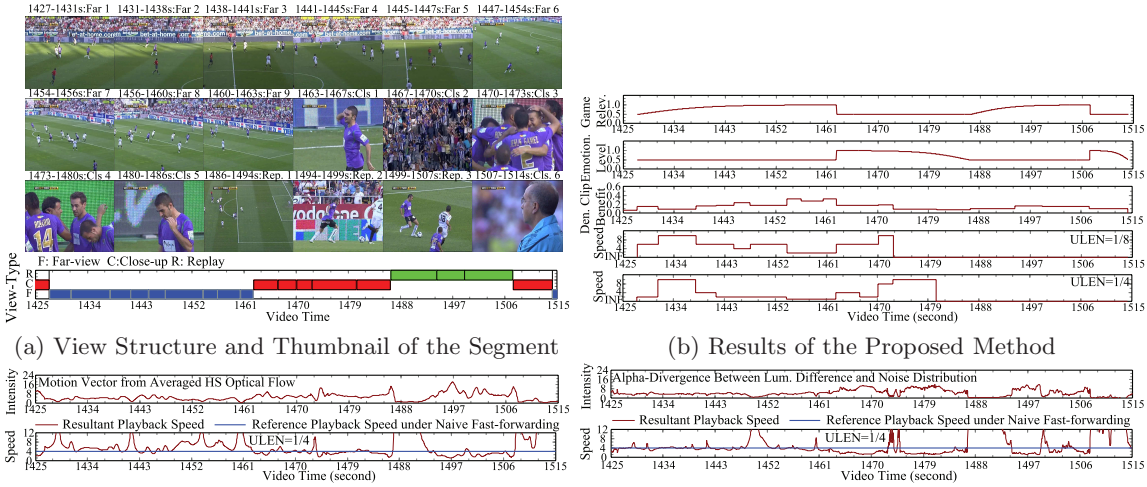
Fig.5 shows the clip benefit $\mathcal{B}_i(v)$ w.r.t. $1/v$ under various $\beta$ and $b_i$ values, so as to analyse the behaviour the clip interest defined in Eq.5 in the above optimization process. Fig.5(a) reveals that the whole curve is convex when $0 < \beta < 1$, which thus enables various options of playback speeds to appear in the benefit/cost convex hulls. In Fig.5(b), we found that the clip with a higher base interest $b_i$ has the same slope value at a slower playback speed. Accordingly, in the above greedy algorithm, slower playback speed will be first assigned to semantically more important clips in the sense of high information.

## 3 Experimental Results

The proposed framework aims at focusing on summarization with adaptive fast-forwarding and semantically relevant and personalized story telling. Those properties are explored through a comparative analysis with state of the art methods. The soccer video used for performance evaluation is 3 hours long with a list of 50 automatically extracted audio hot-spots. Seven different speed options, i.e., 1x, 2x, 4x, 6x, 8x, 10x, and $+\infty$(for content truncation), are enabled in the current implementation, so as to provide comparative flexibility in fast-forwarding control to those methods with continuous playback speeds. Here, $a$x stands for the $a$ times of the normal playback speed. We compared the behavior of our proposed method to the following two methods:

- Peker et al. [13] achieve the adaptive fast-forwarding via constant activity sub-sampling.
- Höferlin et al. [8] determine the activity level by computing the alpha-divergence between the luminance difference of two consecutive frames and the estimated noise model. The adjusted sampling interval is then set to be linearly proportional to the activity level.

The results of the proposed and comparison methods are shown in Figs.6. We use $\alpha = \beta = 0.5$ and plot the results from different methods. We made the




(a) View Structure and Thumbnail of the Segment

(b) Results of the Proposed Method

(c) Results of [13]

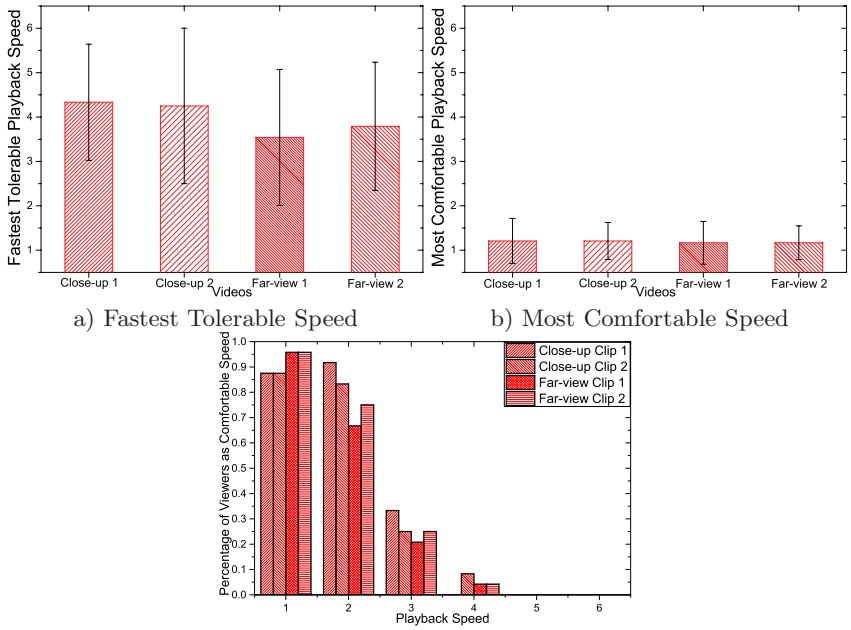
(d) Results of [8]

**Fig. 6.** Summaries produced for the broadcasted soccer video. The first subgraph presents the view-structure of segments and clip thumbnails. Resultant playback speeds from three methods are plotted with the corresponding clip benefit (ULEN for $u^{\mathrm{L}}$).

following major observations: both the optical flow and the alpha divergence failed to correctly measure the intensity of scene activities or the importance of the events; compared to the linear playback speed control in [13] and [8], our framework allows flexible personalization of story organization. We can suppress redundant contents in the replays for higher compaction, consider story continuity, and remove very short clips to avoid flickering; playback speeds of different clips in [13] and [8] maintain the same ratio, when the length of target summary changes, while our method performs non-linear time allocation under different target summary lengths, owing to the flexible definition of clip benefit.

We first subjectively evaluate the suitable playback speeds (Fig.7). 25 participants (including 11 females and 14 male, age from 20-40) were asked to specify their highest tolerable playback speed, comfortable playback speeds and the most comfortable playback speed when presented four groups of video samples with various playback speeds. The highest tolerable speed for far views is lower than that of the close-up views. We consider this as a result that understanding far-view need attentional perception to follow the players. Audiences still feel comfortable in faster playback speeds, which is the base of adaptive fast-forwarding. The most comfortable speed is selected to be the original speed that was produced by experts.

We then collect the global impression of the audiences in comparatively evaluating the generated summaries. We asked 23 participants (including 10 females and 13 males, age from 20-40) to give their opinions on the most preferred result when presented a group of three summaries generated by the above different methods (in the random order), from their *completeness*, *comfort*, and *effectiveness* of time allocation. We plot the results of evaluating summaries under two




a) Fastest Tolerable Speed b) Most Comfortable Speed

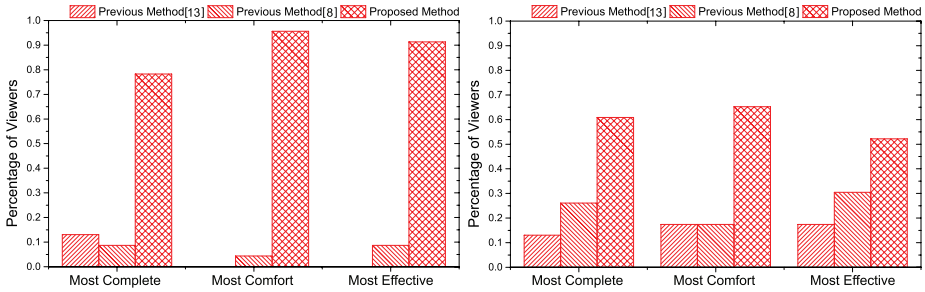
c) Comfortable Speeds

**Fig. 7.** Results of the first subjective evaluation from 25 participants on their feedback under various fast-forwarding speeds when browsing the broadcasted soccer videos.

different compaction ratios (i.e. 1/8 and 1/4) in Fig.8. We make the following observations: our method outperforms the other two methods in generating more complete summaries for highly compact summarization (1/8), which supports our idea of introducing content truncation to save time resources for presenting key events in a clearer way; our method produces more comfortable summaries from the broadcasted soccer video, where both 1/8 and 1/4 are too high for an adaptive fast-forwarding method to produce a comfortable video without truncating some contents. In order to slow down a key event, we have to raise the playback speed of other contents to a much higher level in exchange for the equivalent time resource, which results in flickering and lowers the visual comfort of the summary; our method is evaluated to be the most effective in allocating playback speeds for presenting the actions of interest, especially under a high compaction ratio.

## 4 Conclusions

We proposed a framework for producing personalized summaries that enables both content truncation and adaptive fast-forwarding. Instead of a rigid determination of the fast-forwarding speed, we efficiently select the optimal combination from candidate summaries, which is solved efficiently as a resource-allocation problem. Subjective experiments demonstrate the proposed system by evaluating summaries from broadcasted soccer videos. We will further extend our hybrid method of content truncation and adaptive fast-forwarding. Both semantic in-




a) High Compaction Ratio (1/8) b) Low Compaction Ratio (1/4)

**Fig. 8.** Results of the second subjective evaluation test from 23 viewers, by collecting their global impression on the summaries, in the sense of completeness, comfort and the efficiency of time allocation.

formation and scene activity are important in producing a semantically relevant and visually comfort summary. We will thus consider both types of information in our future work.

## References

1. Chen, F., De Vleeschouwer, C.: Automatic summarization of broadcasted soccer videos with adaptive fast-forwarding. In: ICME'11. pp. 1–6 (July 2011)
2. Chen, F., De Vleeschouwer, C.: Formulating team-sport video summarization as a resource allocation problem. TCSVT 21(2), 193–205 (Feb 2011)
3. Duxans, H., Anguera, X., Conejero, D.: Audio based soccer game summarization. In: BMSB'09. pp. 1–6 (May 2009)
4. Ekin, A., Tekalp, A., Mehrotra, R.: Automatic soccer video analysis and summarization. TIP 12(7), 796–807 (2003)
5. Everett, H.: Generalized lagrange multiplier method for solving problems of optimum allocation of resources. Operations Research 11(3), 399–417 (1963)
6. Ferman, A., Tekalp, A.: Two-stage hierarchical video summary extraction to match low-level user browsing preferences. TMM 5(2), 244–256 (2003)
7. Fernandez, I., Chen, F., Lavigne, F., Desurmont, X., De Vleeschouwer, C.: Browsing sport content through an interactive h.264 streaming session. In: MMEDIA'10. vol. 1, pp. 155–161 (Jun 2010)
8. Höferlin, B., Höferlin, M., Weiskopf, D., Heidemann, G.: Information-based adaptive fast-forward for visual surveillance. Multimedia Tools Appl. 55, 127–150 (2011)
9. Holcombe, A.O.: Seeing slow and seeing fast: two limits on perception. Trends in Cognitive Sciences 13(5), 216 – 221 (2009)
10. Li, Z., Schuster, G.M., Katsaggelos, A.K.: Minmax optimal video summarization. TCSVT 15, 1245–1256 (2005)
11. Ortega, A.: Optimal bit allocation under multiple rate constraints. In: DCC'96. pp. 349–358 (1996)
12. Pan, H., van Beek, P., Sezan, M.I.: Detection of slow-motion replay segments in sports video for highlights generation. In: ICASSP'01. vol. 3, pp. 1649–1652 (2001)
13. Peker, K.A., Divakaran, A., Sun, H.: Constant pace skimming and temporal subsampling of video using motion activity. In: ICIP'01. vol. 3, pp. 414–417 (2001)
14. de Silva, G.C., Yamasaki, T., Aizawa, K.: Evaluation of video summarization for a large number of cameras in ubiquitous home. In: ACM MM'05. pp. 820–828 (2005)
15. Zhu, G., Huang, Q., Xu, C., Rui, Y., Jiang, S., Gao, W., Yao, H.: Trajectory based event tactics analysis in broadcast sports video. In: ACM MM'07. pp. 58–67 (2007)