Effect of Personalization on Retrieval and Summarization of Sports Video

Noboru Babaguchi   Kouzou Ohara   Takehiro Ogura
2-1 Yamadaoka Suita, 565-0871 Japan
Osaka University

Abstract

Personalization is one of the most important mechanisms to make multimedia systems easy to use. In video applications, its embodiment is to tailor video contents for a particular viewer. For this purpose, we are now developing a system of retrieving and browsing video segments, called video portal with personalization (VIPP). VIPP is characterized by 1) supporting the viewer’s access to video contents and making a summarized video clip by taking his/her preference into account, and 2) acquiring the viewer’s profile from his/her operations automatically. In this paper, we discuss the effect of personalization on retrieval and summarization of sports videos on VIPP.

1. Introduction

Personalization is a mechanism of making an information system adaptable to its user. The system can be made easier to use as we continue to use it. Therefore, personalization can be viewed as one of the most important functions in designing information systems. In video applications, personalization is sometimes considered in the context of personal TV systems[1, 2, 3]. We think that an alternative embodiment of personalization is to tailor the video contents for a particular viewer[4]. In this paper, we aim at personalized tailoring in retrieval and summarization of sports videos, taking account of the viewer’s preference. For example, we Japanese may be more interested in the batting result of Matsui, who is a player of Yankees, than in the game result. Thus we may prefer to see a digest including all scenes of Matsui’s at-bats rather than see a digest including highlight scenes of the game.

Although a lot of methods for personalization in WWW applications have been reported[5, 6], they cannot be directly applied to multimedia applications. Also, Smyth et al. [2] have proposed a method of automatically acquiring a user’s preference for video genres by investigating whether he/she is interested in the TV program or not. However, it might be difficult for their method to identify objects or events of interest in videos. Since a video has the temporal axis, we have to examine the viewer’s preference to each video segment which is a temporal segment of a video stream.

We are now developing a video portal system which provides the functions such as efficient retrieval and browsing of sports videos. It helps the viewer access the target video segment, by taking full advantage of its metadata[7]. In addition, monitoring the operations by the viewer enables the system to automatically acquire his/her preference as a profile. We call this system Video Portal with Personalization, abbreviated as VIPP. VIPP associates the viewer’s operations such as retrieving and browsing with keywords in the metadata, updating a preference degree of each keyword. The resulting profiles are utilized both for sorting the video segments in retrieval and for generating a summarized video. In this paper, we present VIPP with emphasis on the effect of personalization on retrieval and summarization of sports video.

2. Video Portal with Personalization

Our video portal system, VIPP, deals with sports videos accompanied by their metadata which describes their semantic structure and contents[7]. The semantic structure of a video recording a broadcast sports game can be represented as a tree, called a game tree, which is hierarchic. For example, the game tree of baseball is composed of the nodes of the 1st through 9th innings, and moreover the inning node is composed of the at-bat nodes, and so on. Semantic contents such as objects and events occurring in the video are also described. This metadata is described with MPEG-7[8]. Ideally metadata should be produced automatically. To this end, we have to achieve automated analysis of the semantic contents of videos. However, since highly reliable analysis of such contents is very difficult at present, we assume the metadata is given manually in this case.

Using the metadata, VIPP provides the following useful functions:

- Retrieval by text: VIPP can retrieve video segments by query keywords such as ‘Homerun’ and ‘Matsui’ that the viewer specifies by detecting them in the metadata. When browsing the retrieved video segments, playback control can be performed.
**Presentation of highlight scenes**: VIPP can retrieve highlight scenes by searching descriptions in the metadata related to the highlights, e.g. scoring innings of Yankees.

**Presentation of tree structures**: VIPP can display logical compositions of videos based on the game tree on its interface for quick access to the video segment corresponding to each node in the tree.

**Video summarization**: To help the viewer understand the whole contents of the video in short, VIPP can generate a summarized video clip[9], i.e. a video digest whose length is specified by the viewer.

**Personalization**: VIPP can reflect on the viewer’s preference in retrieval and summarization of videos.

**Learning to personalize**: VIPP can automatically acquire the viewer’s preference while he/she operates it.

Figure 1 shows a conceptual diagram of VIPP. In what follows, we proceed to describe VIPP’s mechanism related to personalization.

### 3. Personalized Retrieval and Summarization

This section describes a viewer’s profile and how personalized retrieval and summarization of videos can be realized.

#### 3.1 Profile

A profile is a formal description of a viewer’s preference. A fundamental description $< k, P_k >$ is a pair of a keyword $k$ and its weight $P_k$, called a preference degree. The weight $P_k$ expresses how large the viewer’s preference for the keyword $k$ is. A set of fundamental descriptions forms the profile.

It is noted that VIPP has two kinds of profiles: the profile for people and that for events, which represent the people and events in which the viewer is interested, respectively. The keywords in the profile for people are the player’s names, while those in the profile for events are the play names such as ‘SingleHit’ and ‘Homerun.’ An example of the profile is shown in Figure 2.

#### 3.2 Personalization

These two profiles are utilized both for sorting retrieved video segments and for generating video digests. VIPP tries to determine keywords of interest from the profiles in the following way: VIPP first normalizes the weights of all the keywords so that the maximum weight in both profiles can become 1. For each profile, the average of all the normalized weights $\tau$ is calculated. VIPP can determine the keywords whose weights exceed $\tau$ as the keywords of interest.

Let us describe procedures of personalized retrieval and summarization as follows:

**<Personalized retrieval of video segments>**

VIPP evaluates a fitness value of each video segment to be retrieved. The fitness value of the video segment is defined as the sum of the weights of keywords of interest in the profile. Then, each keyframe of the video segments which are sorted in order of the magnitude of their fitness values is browsed. The keyframe is the first image frame of each video segment. The top-ranked video segment should be what the viewer wants to see most. Each video segment can be displayed as a video clip under playback control.

**<Personalized video summarization>**

1) VIPP accepts the desired length of the resulting digest given by the viewer.

2) VIPP calculates a significance degree of each video segment, taking into account the following factors[9]:

   A) Temporal order of the video segments: A scene appears at the latter or last stage of the sports game is likely to be significant.

   B) Attributes of the video segments: A scene changing the game states, e.g. turning the game, is likely to be significant. A live scene is more significant than a replay scene.

   C) Fitness value of the video segments: A scene that matches well with the profile is significant.
These factors can be individually arranged in making the video digest.

3) VIPP repetitively selects video segments in order of the significance degree until the sum of each segment length exceeds the predefined digest length.

4) VIPP finally produces a video clip by connecting the selected video segments in original temporal order. The highlight scenes in the video digest flows chronologically.

4. Learning to Personalize

This section discusses learning to personalize and formulates a procedure of profile acquisition. Generally, personalization is realized by monitoring the explicit or implicit feedbacks given by the user. In this paper, we focus only on the implicit feedbacks because we think personalization should adapt to the user unconsciously. As such feedbacks from a viewer, VIPP monitors his/her operations of retrieval and browsing. In retrieval of video segments, we consider 1) retrieval by text, and 2) retrieval for presentation of highlight scenes. In browsing video segments retrieved, we consider 1) selection of video segments from the ordered list resulting from retrieval, and 2) playback, stop, fast-forward, and rewind for browsing.

The profile acquisition procedure is repeated while the viewer is using VIPP. Each time he/she takes one of the above operations, the weights of keywords in the profiles are updated. The operations can be identified by monitoring which of the playback buttons is touched on its interface. Initially both profiles are empty and the weights of keywords are set to 0. Specifically, the new weight $P'_k$ of a keyword $k$ whose current weight is $P_k$ is given by the following equations.

In case of retrieval,

\[ P'_k = \alpha \times \eta_k + P_k \times \delta^{1-\eta_k}, \tag{1} \]

where $\eta_k = 1$ if $k$ was a keyword given for retrieval; $\eta_k = 0$ otherwise.

In case of browsing, we consider the tree structure of video segments. Namely,

\[ P'_k = \beta \times \rho_k + P_k \times \gamma^{\sigma_k} \times \delta^{1-\rho_k}, \tag{2} \]

where $\rho_k = 1$ if $k$ appears in the video segment selected for browsing or if $k$ appears in a descendent of the selected segment in the game tree; $\rho_k = 0$ otherwise. $\sigma_k = 1$ if $k$ appears only in the video segment ranked above the selected one, but not selected; $\sigma_k = 1$ otherwise.

Equation (1) increases the weight of the keyword given for retrieval by $\alpha$. Similarly, Equation (2) increases the weights of the keywords appearing in the video segment the viewer selected for browsing by $\beta$. In case of browsing, if the selected segment has its descendents in the game tree, the weights of the keywords appearing in them are increased. Instead, the weights of the other keywords are decreased by $\delta$. Especially in case of browsing, the weights of the keywords appearing only in the video segments ranked above the selected one in the result of the retrieval but not selected are further decreased by $\gamma$. Decreasing the weights of such keywords allows VIPP to adapt to the change of the viewer’s interests by forgetting past information gradually. The parameters $\alpha$, $\gamma$, and $\delta$ are constants, whereas $\beta$ is given based on the length of the segment which the viewer actually browses.

5. Experimental Results

To verify the effectiveness of profile acquisition, we conducted the following experiments. As subjects for the experiments, we had eight viewers. Each of them decided a different set of five keywords beforehand. We then let them each use VIPP for one or two hours, evoking the procedure of profile acquisition. After they performed more than a hundred of operations, we sorted the acquired keywords in order of their preference degree, and checked where the pre-decided five keywords were ranked in all the acquired keywords of the profile.

In this experiment, we used nine broadcast videos of Japanese professional baseball games. The average length of the videos is about 210 minutes. The metadata of these videos totally contains 95 and 25 keywords about players and events, respectively.

The experimental results indicate that 73% of the pre-decided keywords for people and 55% of those for events were placed in the top-5 rank. Further, about 90% of these keywords for both people and events were placed in the top-10 rank. These rates are the average values of all the eight subjects. Unsuccessful cases were due to low frequency of some event keywords. We confirm that this procedure of profile acquisition works well if VIPP accepts more than a hundred of the viewer’s operations. In some cases, the profile turned characteristic at its early stage. The details about this issue are shown in [10].

Figure 3 shows an instance of the transition of preference degrees of the pre-decided keywords about people. As can be seen, the preference degree of the keywords which were not used
through the viewer’s operations was gradually decreased. In contrast, if the keywords were actually used, their degree got large. For example, when the keyword ‘Igawa’ was selected at the 35th and 95th operations, its degree was abruptly changed.

Let us consider the change of the profile descriptions as operations by a particular viewer are going on. As mentioned previously, VIPP has two kinds of profiles: the profile for people and that for events. In this case, we let two subjects use VIPP and make operations for about two hours. One subject had only the profile for people and made 164 operations. The other subject had only the profile for events and made 182 operations.

Figures 4 and 5 show the change of the profile for people and that for events, respectively. In both figures, each description has only the profile for events and made 182 operations.

We examined how the retrieval result of video segments about players was changed when a query keyword was given as ‘Hit,’ an event name, and the acquired profile for people (Fig.4) was used. Figure 6 shows the change of retrieval results for players. In Fig.6, the video segments marked by ‘×’ disappear at the next stage. The scenes of ‘Hamanaka,’ the player the viewer most preferred, were located in the top-three rank. In addition, we investigated how the retrieval result of video segments about events was changed when a query keyword was given as ‘Nioka,’ a player name, and the acquired profile for events (Fig.5) was used. Figure 7 shows the change of retrieval results for events. The scenes about the events located in the higher rank in the profile for events were actually retrieved.

Finally two video digests were produced at each stage based on each profile. The given digest length was two minutes. Figures 8 and 9 show the change of digest results based on the profile for people and that for events, respectively. We can observe that the video digests are also personalized according to the profiles. In particular, the video digest based on the profile for people (Fig.8) at the final stage included only the scenes of the viewer’s favorite player, ‘Hamanaka.’

Let us here discuss how the results of video digests should be evaluated. He et al.[11] claimed that ideal video digests should have the four desirable properties: conciseness, coverage, context and coherence. However, no evaluation scheme has been unfortunately realized in a concrete form. Furthermore, personalized video digests we deal with would be far more difficult to evaluate because of the subjectivity. Making a good scheme of the evaluation is an open problem.
In this experiment, the duration of using VIPP, which is equal to that of profile acquisition, was only a couple of hours. If the duration got much longer, the profile would fit the viewer’s preference more. Further analysis of long duration of learning should be explored.

6. Conclusion

In this paper, we proposed a personalized video portal system, called VIPP. Associating the viewer’s operations with the keywords in the metadata allows VIPP to automatically acquire his/her preference for people and events in videos. VIPP offers a variety of useful functions for quick access to the video contents: keyword based retrieval, highlight based summarization, and tree structure based browsing. We verified that the profiles can act as a bias in retrieval and summarization of the videos, and that hundreds of operations by the viewer can generate appropriate profiles. Our future work is VIPP’s evaluation for a considerable period of time.

References


