Personalized Abstraction of Broadcasted American Football Video by Highlight Selection

Noboru Babaguchi, *Member, IEEE*, Yoshihiko Kawai, Takehiro Ogura, and Tadahiro Kitahashi, *Member, IEEE*

***Abstract*—Video abstraction is defined as creating shorter video clips or video posters from an original video stream. In this paper, we propose a method of generating a personalized abstract of broadcasted American football video. We first detect significant events in the video stream by matching textual overlays appearing in an image frame with the descriptions of gamestats in which highlights of the game are described. Then, we select highlight shots which should be included in the video abstract from those detected events reflecting on their significance degree and personal preferences, and generate a video clip by connecting the shots augmented with related audio and text. An hour-length video can be compressed into a minute-length personalized abstract. We experimentally verified the effectiveness of this method by comparing man-made video abstracts.**

***Index Terms*—Broadcasted sports video, event detection, highlight, personalization, video abstraction.**

I. Introduction

VIDEO abstraction is defined as creating shorter video clips or video posters from an original video stream [1]. In recent years, it has become one of the most demanding video applications. The scheme of video abstraction is divided into two classes. The first is to create a concise video clip by temporally compressing the amount of the video data. It is sometimes referred to as video skimming and its actual examples are movie trails and sports digests. Recent researches by Smith *et al.* [2], Lienhart *et al.* [3], He *et al.* [4], Oh *et al.* [5], and Babaguchi [6] are classified in this class. The second is to provide image keyframe layouts representing the whole video contents on a computer display like a storyboard. It is suitable for at-a-glance presentation by means of spatial visualization. Such systems were developed by Yeung and Yeo [7], Uchihashi *et al.* [8], Chang *et al.* [9], or Toklu *et al.* [10].

In this paper, we present a new method of abstracting *sports video*, specifically broadcasted TV programs of American football games, taking *personalization* into consideration. This method belongs to the first class as above, and can be viewed as video abstraction based on *highlights* that are closely related to semantical video contents. Generating a highlight based abstract of sports video requires detecting *significant events* like score events, which are candidates of the highlights, in the video stream. We think that the existing methods such as [2]–[4] take little account of detecting significant scenes based on its semantical contents, because they are based mainly on surface features of the video such as shot boundaries, camera works, shot framing and sound magnitude.

Let us consider how the events should be detected. For this purpose, automated analysis of video contents is ideal. However, at present, highly reliable detection by means of image analysis is very difficult. Our solution is to make use of external metadata, called *gamestats*, which are available via websites or newspapers. Since the progress and result of the game are described in the gamestats, we can easily know when significant events occurred. Thus, the event detection problem can be reduced into linking segments of the video stream with the descriptions of the gamestats. To achieve this linking, we focus on *textual overlays* that may appear at times in an image frame. They are sure to be present in the broadcasted sports video and have rich information about its contents. In a sense, they are special visual objects describing the video contents. We try to recognize the text shown in the overlay and to identify the time point when the event took place in the video stream.

As described before, we concentrate on personalization [11], [12] in making video abstracts. It has been extensively attempted in a variety of application fields such as web mining and user interfaces. Also in the field related to video processing, the idea of personalization has been recently introduced to personal TV applications [13]–[15]. On the other hand, our aim is to apply it to video abstraction. We emphasize that because the significance of scenes could vary according to personal *preferences* and *interests*, the resultant video abstracts should be individual. For example, one who favors the San Diego Chargers wants to see more scenes about them than scenes about their opponent. Namely, the personalization can be thought of as tailoring the video contents for a particular viewer. To this end, a *profile* is provided to collect personal preferences such as his/her favorite team, player and play. It is expected to act as a bias in making video abstracts.

The rest of this paper is organized as follows. Section II surveys related work on video abstraction. In Section III, we describe a method of detecting significant events in the video stream by means of recognition of the textual overlays. In Section IV, we describe a method of generating video abstracts based on highlight shots. In Section V, we experimentally evaluate the performance of this method, and compare the

Manuscript received March 27, 2001; revised November 22, 2002. This work was supported in part by a Grant-in-Aid for scientific research from the Japan Society for the Promotion of Science and by Telecommunications Advancement Organization of Japan, and was performed while the authors were with the ISIR, Osaka University. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. HongJiang Zhang.

N. Babaguchi is with the Department of Communication Engineering, Osaka University, Osaka 565-0871, Japan (e-mail: babaguchi@comm.eng.osaka-u.ac.jp).

Y. Kawai and T. Ogura are with NHK, Tokyo 150-8001, Japan.

T. Kitahashi is with the Department of Informatics, Kwansei Gakuin University, Hyogo 669-1337, Japan.

Digital Object Identifier 10.1109/TMM.2004.830811

generated abstracts with man-made abstracts. Section VI gives concluding remarks.

II. Related Work

As stated in Section I, the scheme of video abstraction is divided into two classes: *time compression* and *spatial expansion*. We first describe existing methods of the time compression type. Smith *et al.* [2] proposed a method of great interest, what is called video skimming. They extracted significant information from video such as audio keywords, specific objects, camera motions and scene breaks with integrating language, audio, and image analyzes. They reported the compression ratio was about 1/20 although the essential content was kept. Lienhart *et al.* [3] tried to assemble and edit scenes of significant events in action movies, focusing the on actor/actress's closeup, text, and sound of gunfire and explosion. It can be pointed out that these two methods are based on surface features of the video rather than on its semantical contents. Oh *et al.* [5] developed a method of abstracting video using its interesting scenes. Their method is able to automatically uncover the remaining interesting scenes in the video by choosing some interesting scenes. They reported successful results for news video. Babaguchi [6] discussed video abstraction based on its semantical content in the sports domain. To select highlights of a game, an impact factor for a significant event in two-team sports was proposed. He *et al.* [4] proposed a method to create summaries for online audio-video presentations. They used pitch and pause in audio signals, slide transition points in the presentation, and users' access patterns. Their compression rate ranged from 1/5 to 1/4. What is more important in their work is to show the desirable properties for an ideal abstract, the four C's: conciseness, coverage, context and coherence. They evaluated the generated abstracts from these four C's viewpoints.

Let us next mention existing methods of the spatial expansion type. Their goal is to visualize the whole contents of the video. Because of the keyframe layouts, they are suitable for the at-a-glance browsing. Yeung and Yeo [7] proposed a method to automatically create a set of video posters (keyframe layouts) by the dominance value for each shot. Also, Uchihashi *et al.* [8] presented a similar method of making video posters whose size can be changed according to the importance measure. Chang *et al.* [9] made shot-level summaries of time ordered shot sequence or hierarchical keyframe clusters, as well as program-level summaries.

The method proposed in this paper belongs to the first class, the time compression type. The live sports broadcasts of our concern are usually redundant. In fact, scenes in play are part of the whole video, and highlights are also part of the scenes. Therefore, time compression can be strongly required in some situations. When busy, we might be happy if we could view a two minutes video abstract for a two-hour game.

III. Detection of Significant Events

This method consists of two major steps. The first step is to detect significant events in the original video stream according to the descriptions in the gamestats. The second step is to make a video abstract by connecting highlight shots and reflecting on personal preferences.

```
Louisiana Superdome
New Orleans, Louisiana
January 26, 1997
Attendance: 72,301
MVP: Desmond Howard, KR-WR, Green Bay

                          SCORING
New England           14  0  7  0 -- 21
Green Bay             10 17  8  0 -- 35

1stQ  3:32  GB-  Rison 54 pass from Favre (Jacke kick)
1stQ  6:18  GB-  FG Jacke 37
1stQ  8:25  NE-  Byars 1 pass from Bledsoe (Vinatieri kick)
1stQ 12:27  NE-  Coates 4 pass from Bledsoe (Vinatieri kick)
2ndQ  0:56  GB-  Freeman 81 pass from Favre (Jacke kick)
2ndQ  6:45  GB-  FG Jacke 31
3rdQ 13:49  GB-  Favre 2 run  (Jacke kick)
3rdQ 11:33  NE-  Martin 18 run  (Vinatieri kick)
3rdQ 11:50  GB-  Howard 99 kick return  (2-pt Chmura from
                 Favre, T-Formation)

TEAM STATISTICS                  N.E.          G.B.
Total First Downs                 16            16
  Rushing                          3             8
    :                              :             :
```

Fig. 1. Gamestats descriptions.

We now proceed to describe how significant events are detected. Since event detection is a typical problem about contents analysis of video, a large number of methods for it have been reported so far. As far as the event detection in broadcasted sports video is concerned, various methods, e.g., [16], [17], attempted achieving it mainly through image analysis, and other methods [18]–[21] tried integrating video, audio and text analysis. However, highly reliable detection by such analysis has not been attained yet. As an alternative approach, we make use of the gamestats and link the video segments with their descriptions. Because we deal with American football TV programs that are live on-air, e.g., NFL professional football and NCAA college football, almost all of the gamestats of our concern can be obtained at the websites of the team, the football organization, or the sports news.

Fig. 1 shows an example of descriptions in gamestats. Here, the time in the game is simply called the *game time*. The gamestats contain the following data: a) game time when an event took place (e.g., 1st Qtr. 3 min, 32 s), called the *event time*, b) sorts of events (e.g., touchdown), c) names of players and teams related to the events, and so on. These data can become good indexes to video contents.

All we have to do is to detect, in the video stream, each event that occurred at the time specified in the gamestats. Note that the time in the gamestats means not the media time but the game time. In this case, the media time is, for example, the time from the beginning of the TV program or the recording tape. Apparently there is no direct correspondence between the media time and the game time. In what follows, we will mention detecting events through analysis of textual overlays.

A. Identification of Event Frames

We attempt to recognize text expressing the game time in the overlay, and then to identify an *event frame*. We think an event occurs in the shot including the event frame. There are several kinds of overlays appearing in an actual sports TV program. One of these indicates the information about the current status of the game such as quarter, time and score, as shown in Fig. 2.

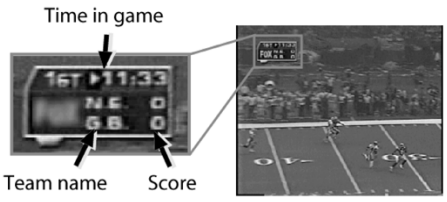


Fig. 2. Example of an overlay.

As far as the same TV program is concerned, both the location where the overlay appears and the layout of textual data are kept unchanged.

To identify the event frame, an overlay model is employed. In the model, we describe the location where the overlay should appear in the image frame as a rectangular region. We further describe layout structure of the overlay to represent where its textual data is located. For identification of event frames, the rectangular region for time display is of our particular interest. Original image patterns in these regions are exploited as templates in pattern matching.

First of all, we perform presence analysis of the overlays. Pattern matching between the overlay model and the image frame informs us whether the target overlay is present in the frame. Next, we realize a digit recognizer to identify the event frame. In the overlay, the game time is expressed as four digits: the first two digits represent minutes, while the last two ones do seconds. Employing a template corresponding to each digit representing the event time to be searched, we find a good matching position in the region for time display. Matching is made by sliding the template horizontally and vertically for each digit. If the event time described in the gamestats is found with the recognizer, we determine that the image frame with the matched overlay is the event frame. In this way, detection of the significant events is realized by searching the text indicating the event time.

B. Detection of Event Shots

Based on the identified event frame, we try to detect *event shots*. Note that a shot is defined as consecutive image frames at a single camera view. We here classify the event shots into the following four types.

i) *Live-play (Live) shot*: a shot where the game is actually in play. This is defined as a shot including the event frame.
ii) *Replay shot*: a shot from different perspectives or with slow motion. Live and replay shots essentially express the same event. A replay shot can be automatically detected by identifying a pair of Digital Video Effects (DVEs), DVE-IN, and DVE-OUT [22], because in sports TV programs it is sandwiched in between DVE-IN and DVE-OUT as shown in Fig. 3. DVE is either a gradual and spatial shot change operation which is similar to a wipe or a special CG shot.
iii) *Pre-play shot*: a shot just before the live shot.
iv) *Post-play shot*: a shot just after the live shot.

The shot boundary between the live and pre-play or post-play shots is detected by color histogram difference between blocks of the neighboring frames. The relationship among these elements is graphically illustrated in Fig. 3.

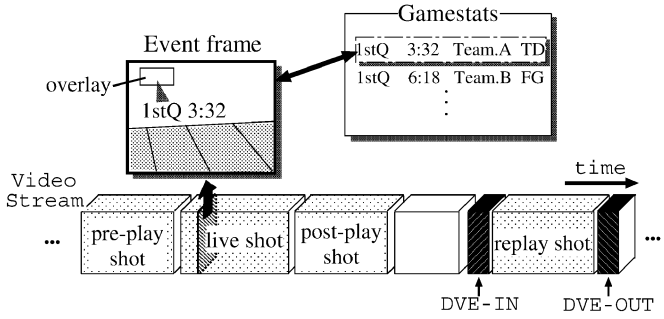


Fig. 3. Relationship among event shots, an event frame and gamestats.

IV. Generation of Personalized Video Abstract

This section presents the process of generating video abstracts from the detected significant events. Once such events are determined, we select *highlights* of the game from all the events, considering *profile descriptions*. Then we connect the highlight shots in original temporal order.

A. Generating Rules

The generating rules for the video abstract [6] are as follows.

ABSTRACT ::= **INTRO HIGHLIGHT ENDING**
INTRO ::= **INTRO-VC INTRO-ST**
ENDING ::= **ENDING-ST ENDING-VC**
HIGHLIGHT ::= [**PRE-VC EVENT-ST POST-VC**]$^{+}$
EVENT-ST ::= **PRE-PLAY-ST*** **LIVE-ST**$^{+}$
REPLAY-ST* **POST-PLAY-ST***

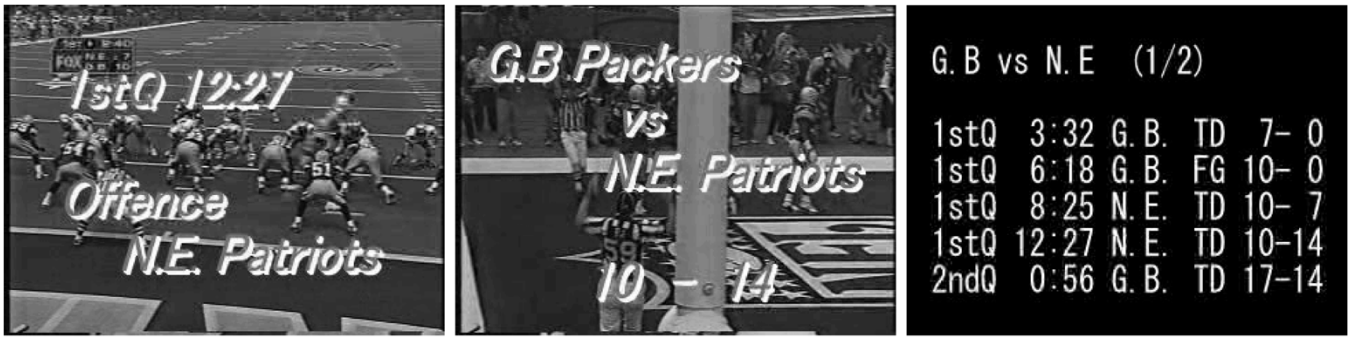
Here, **ST** and **VC** stand for a shot and a video caption, respectively, and the notation $[]^{+(*)}$ means one (zero) and more repetitions. These rules are derived from actual composition of sports digests in news TV programs. The generated **ABSTRACT** consists of not only actual highlight shots but also video captions explaining them. The captions include the following data.

INTRO-VC: date, location, team-name,….
ENDING-VC: result, gamestats,….
PRE-VC: quarter, time, offense-team,….
POST-VC: score, team, player,….

In addition to these, synchronized audio is associated with the shot. Examples of superimposed video captions of **PRE-VC**, **POST-VC**, and **ENDING-VC** are shown in Fig. 4(a)–(c), respectively. **PRE-VC** and **POST-VC** explain what the highlight is about. **ENDING-VC** summarizes the game.

B. Profile

We strongly claim that a video abstract has to be personalized because significance of events could change individually. For this purpose, we provide a profile to collect personal preferences and interests. Its items are as follows: a) *favorite teams*, b) *favorite players*, and c) *events to want to see*. Specifications to make the abstract are included as well. Specifically, they are d) *range of the video stream* to be abstracted, and e) *length of the abstract*. This method aims at personalization by referring to these profile descriptions.



(a) (b) (c)

Fig. 4. Examples of superimposed video captions. (a) PRE-VC. (b) POST-VC. (c) ENDING-VC.

C. Significance Degree of Events

The highlights of the game depend on the significance of each event. We think that its significance can be estimated in terms of event rank, event occurrence time, and the profile. Three kinds of degrees are defined as follows.

1) Event Rank: In this paper, we assume that a game between two teams, team A and team B, is played, and that the team's goal is to get more scores than its opponent. Under this assumption, there are three states of the game situation: "the two teams tie," "team A leads," and "team B leads." If a score event can change the current state into a different state, we call it a *state change event* (SCE) [6]. It is evident that SCE's are candidates of the highlights.

The rank of various events is defined as follows.

Rank 1: SCE's. They are defined by following the transition of the game score.
Rank 2: not SCE's but exceptional score events, e.g., kickoff return touchdown.
Rank 3: events closely related to score events.
Rank 4: all other events that are not Rank 1 to 3.

Now, *rank based significance degree* of an event I_r, $(0 \leq I_r \leq 1)$ is defined as

$$I_r(E_i) = 1 - \frac{r_i - 1}{3} \cdot \alpha$$

where r_i, $(1 \leq r_i \leq 4)$ denotes the rank of the ith event E_i and α, $(0 \leq \alpha \leq 1)$ is a coefficient to consider how large the difference of the rank affects the significance.

2) Event Occurrence Time: The score events occurring at the latter or final stage of the game largely affect the result. Thus, such events are of great significance. We define *occurrence time based significance degree* of an event I_t, $(0 \leq I_t \leq 1)$ as

$$I_t(E_i) = 1 - \frac{N - i}{N - 1} \cdot \beta$$

where N is the number of all events and β, $(0 \leq \beta \leq 1)$ is a coefficient to consider how large the occurrence time affects the significance.

3) Profile: Comparing the descriptions of the profile and the occurring event, we define *profile based significance degree* of an event I_p, $(0 \leq I_p \leq 1)$ as

$$I_p(E_i) = (1 - \gamma)^l$$

where l, $(l \geq 0)$ denotes the number of descriptions that do not coincide with each other. Also γ, $(0 \leq \gamma \leq 1)$ is a coefficient to consider how large the profile affects the significance.

As a consequence, *significance degree of an event* I, $(0 \leq I \leq 1)$ is given by

$$I(E_i) = I_r(E_i) \cdot I_t(E_i) \cdot I_p(E_i).$$

Changing the parameters of α, β, and γ enables us to control the composition of the video abstract. As α gets larger, it can be made by emphasizing the event rank. In this case, the events that turned the game would be certainly included. The other parameters behave in a similar manner. We assume that each parameter should be set by the viewer (user) when he/she starts making the abstract.

D. Selection of Highlights

To determine highlights, we concentrate on both *priority order* of shots and *significance degree* of events. We begin with segmentation of a shot. Some shots may have their partial segment which is still from the beginning. In general, such a shot can be lengthy and needs to be shortened in editing process. Therefore, we divide the shot into the first still segment and its subsequent motion segment. To find the still segment, we use block matching between the neighboring frames. From the initial frame of the shot, the matching is continued until we detect the motion frame where either some objects or the camera begins to move. Of course, there may be no still segment in the shot in some cases.

Let us now consider the priority order. The priority order of each shot segment is defined as follows: 1) motion live shot, 2) still live shot, 3) motion replay shot, 4) still replay shot, 5) motion pre-play shot, 6) still pre-play shot, 7) motion post-play shot, and 8) still post-play. In the above context, the motion live shot means the motion segment in the live shot. This order is based on the heuristics that live and replay shots revealing an identical event should be main components of the video abstract, as well as that the still segments from the beginning of a shot could be hardly informative.

The point to make good video abstracts is that more events with higher significance degree should be certainly included as well as that more shots related to such events should be included. An outline of highlight selection is as follows. First, all events are sorted in descending order of their significance degree. Similarly, all of the event shot segments are arranged according to

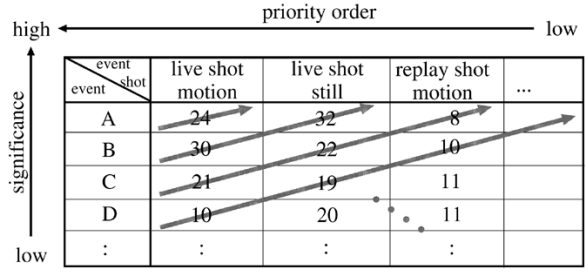


Fig. 5. Highlight selection using the shot length table.

their priority order. For each event, we form a table of each shot length (s), as illustrated in Fig. 5. From left-up to right-down and diagonally, the shots are selected in such order as the arrow in Fig. 5 indicates. The left-upper element is initially selected. The highlight selection is continued until the sum of the selected shot length exceeds the abstract length described in the profile. This way leads to selecting more shots of the event with higher significant degree.

In the following, we describe the *highlight selection algorithm*. Let $T = \{t_{m,n}\}, m = 1, \ldots, m_{\max}, n = 1, \ldots, n_{\max}$ be the table of shot length, where m and n denote the sorted events according to their significance degree and the event shot segments according to their priority order, respectively. Let $t_{m,n} = 0$ if no segment is given. In addition, we expediently set $t_{m,2} = 0, m = 1, \ldots, m_{\max}$ so as to avoid selecting the shot segments solely for a specific event. Accordingly, length values of the motion live shots and the still live shots are placed in the first and third columns of T, respectively.

Input: the shot length table T whose size is $m_{\max} \times n_{\max}$, and the abstract length A.
Output: the highlight list *LIST*.

```
Algorithm Highlight_Selection (T, A,
LIST)
begin
 sum := 0; LIST := empty;
 for k := 1 to m_max + n_max − 1 do
 for n := 1 to k do
  begin
  m := k + 1 − n;
  if m ≤ m_max and n ≤ n_max and t_{m,n} > 0 then
   begin
   sum := sum + t_{m,n};
   if sum ≥ A then Halt
   else Put t_{m,n} into LIST
   end
  end
end
```

E. Editing of Selected Highlights

According to the generation rules stated in Section IV-A, a video abstract is made. We rearrange the shot segments stored in the highlight list *LIST* in original temporal order, and then form a video clip by connecting them. In the video abstract, each highlight scene appears in order of the events that actually happened in the game.

TABLE I
EVENT SEQUENCES OF VIDEO1: SUPER BOWL XXXI, GREEN BAY (GB) VERSUS NEW ENGLAND (NE)

No.	Time		Event	Team	Player	Score	Rank
#1	1Q	3:32	TD	GB	Rison,Favre	7-0	1
#2	1Q	6:18	FG	GB	Jacke	10-0	4
#3	1Q	8:25	TD	NE	Byars,Bledsoe	10-7	4
#4	1Q	12:27	TD	NE	Coates,Bledsoe	10-14	1
#5	2Q	0:56	TD	GB	Freeman,Favre	17-14	1
#6	2Q	6:45	FG	GB	Jacke	20-14	4
#7	2Q	13:49	TD	GB	Favre	27-14	4
#8	3Q	11:33	TD	NE	Martin	27-21	4
#9	3Q	11:50	TD	GB	Howard	35-21	2

Further, video captions that are *annotations* about the highlight are attached before and after the highlight shot, as shown in Fig. 4, in order to enhance the viewer's comprehension. The captions shown before the highlight are the game time and the name of the offense team, while those shown after it are the current score after the event. Their contents are produced from the gamestats descriptions. In addition, the audio temporally corresponding to its shot is attached together.

V. EXPERIMENTAL RESULTS

In this section, we show some experimental results to verify the effectiveness of the proposed method. To evaluate the quality of generated video abstracts, we compare them with *man-made video abstracts* which are available at the WWW. Since they were produced by professional video editors, we can regard them as correct examples. Subsequently, we consider the effect of using the profile by investigating the difference between personalized and nonpersonalized abstracts. Lastly, we make discussions about the method.

A. Setup and Preliminary Experiments

We tested this method for five kinds of sample streams, designated by video1,…, video5, which were actual TV programs broadcasted in the U.S. Their total length was about 180 min. The gamestats for each game were obtained at [23], [24]. Tables I–V list event sequences of each stream. Totally, there were 45 events consisting of touchdown (TD) and field goal (FG). Since we here handled only score events, there was no event whose rank was three.

Using video1, we began by investigating the basic properties of the presence analysis of overlays and the recognition of the digits standing for the game time. The size of the image frame was 640×480, and the input frame rate was two frames/s. For all the frames in video1, the recall and precision of 99.8% and 100% for overlay presence analysis were obtained. The accuracy of the digit recognizer was 91.5%, and the processing time (SGI O2, 180 MHz) was 0.042 s on average.

After investigating the basic properties as above, we conducted the experiment of event detection for all the sample streams. The overlay model was obtained from the initial several image frames in which the overlay was present in the sample stream. The model was switched for each stream of

TABLE II
EVENT SEQUENCES OF VIDEO2: SUPER BOWL XXXIII, DENVER (DEN) VERSUS ATLANTA (ATL)

No.	Time	Event	Team	Player	Score	Rank
#1	1Q 5:25	FG	ATL	Andersen	0-3	1
#2	1Q 11:05	TD	DEN	Griffith	7-3	1
#3	2Q 5:43	FG	DEN	Elam	10-3	4
#4	2Q 10:06	TD	DEN	R.Smith,Elway	17-3	4
#5	2Q 12:35	FG	ATL	Andersen	17-6	4
#6	4Q 0:04	TD	DEN	Griffith	24-6	4
#7	4Q 3:40	TD	DEN	Elway	31-6	4
#8	4Q 3:59	TD	ATL	Dwight	31-13	2
#9	4Q 7:52	FG	DEN	Elam	34-13	4
#10	4Q 12:56	TD	ATL	Mathis,Chandler	34-19	4

TABLE III
EVENT SEQUENCES OF VIDEO3: SUPER BOWL XXXIV, ST. LOUIS (STL) VERSUS TENNESSEE (TEN)

No.	Time	Event	Team	Player	Score	Rank
#1	1Q 12:00	FG	STL	Wilkins	3-0	1
#2	2Q 10:44	FG	STL	Wilkins	6-0	4
#3	2Q 14:45	FG	STL	Wilkins	9-0	4
#4	3Q 7:40	TD	STL	Holt,Warner	16-0	4
#5	3Q 14:46	TD	TEN	George	16-6	4
#6	4Q 7:39	TD	TEN	George	16-13	4
#7	4Q 12:45	FG	TEN	Del Greco	16-16	1
#8	4Q 13:06	TD	STL	Bruce,Warner	23-16	1

concern because overlays with different configurations appeared at different locations of the image frame. We examined whether or not the overlays were sure to appear whenever the actual events happened. This can be checked by searching the overlay showing the event time. Two cases were missed for all of the 45 events in video1 to video5. In the detection, we tried to search each time within 2 s at the center of the event time. For example, we were seeking for 9:14, 9:15, 9:16, 9:17, and 9:18 when the event time was 9:16. As a result, we detected all the event frames when the overlays showing the event time were really present. Although the recognizer's ability was not perfect, the search of the event time was successful. Because a top-down strategy driven by target templates can be employed, the search of a specific digit is easier than the recognition of unknown digits. The missed two events when no overlay was present were given manually for the sake of the following experiments.

B. Comparison With Actual Video Abstracts

Actual man-made video abstracts for the same game, denoted by WWW-1,…, WWW-5, were available at [25], [26]. We think of these WWW abstracts as reference models of the abstracts this method generates. For video5, we found two kinds of WWW abstracts: WWW-5-MI for the MI fans and WWW-5-SL for the SL fans. They can be naturally regarded as instances of personalized abstracts. The WWW abstracts included some scenes that were not concerned with the score events like the scene of the stadium just before the start of the game. For fair comparison with this method, such scenes were disregarded. Then, each length of WWW-1,…, WWW-4, WWW-5-MI and WWW-5-SL was 47, 48, 65, 135, 68, and 70 s, respectively. We generated each abstract, denoted by OUR-1,…, OUR-5-MI (SL), and compared them with WWW-1,…, WWW-5-MI (SL). In this case, the length coinciding with that of each WWW abstract was set as the specification and no preference was specified in the profile.

TABLE IV
EVENT SEQUENCES OF VIDEO4: SUPER BOWL XXXV, BALTIMORE (BAL) VERSUS NEW YORK (NYG)

No.	Time	Event	Team	Player	Score	Rank
#1	1Q 8:10	TD	BAL	Stokley,Dilfer	7-0	1
#2	2Q 13:19	FG	BAL	Stover	10-0	4
#3	3Q 11:11	TD	BAL	Starks	17-0	2
#4	3Q 11:29	TD	NYG	Dixon	17-7	2
#5	3Q 11:47	TD	BAL	Ja.Lewis	24-7	2
#6	4Q 6:15	TD	BAL	Ja.Lewis	31-7	4
#7	4Q 9:33	FG	BAL	Stover	34-7	4

TABLE V
EVENT SEQUENCES OF VIDEO5: NFL2000, MINNESOTA (MI) VERSUS ST. LOUIS (SL)

No.	Time	Event	Team	Player	Score	Rank
#1	1Q 6:04	TD	SL	Faulk	0-7	1
#2	1Q 10:06	TD	SL	Faulk	0-14	4
#3	2Q 1:12	FG	SL	Wilkins	0-17	4
#4	2Q 7:46	TD	MI	McWilliams,Culpepper	7-17	4
#5	2Q 13:49	FG	SL	Wilkins	7-20	4
#6	3Q 3:40	TD	MI	Culpepper	14-20	4
#7	3Q 6:41	TD	SL	Faulk	14-26	4
#8	3Q 10:24	TD	SL	Faulk	14-33	4
#9	3Q 14:28	TD	MI	Carter,Culpepper	21-33	4
#10	4Q 6:50	TD	SL	Waston	21-40	4
#11	4Q 12:33	TD	MI	Moss,Culpepper	29-40	4

To evaluate the quality of the generated abstract, we introduce two measures as

$$R_W = \frac{N_{both}}{N_W}$$
$$R_O = \frac{N_{both}}{N_O}$$

where N_{both}, N_W, and N_O denote the number of highlights included in both WWW and our abstracts, that of highlights included in the WWW abstract, and that of highlights included in our abstract, respectively. As both R_W and R_O approximate to 1, the generated abstract gets similar to the WWW abstract. This implies that the higher the two measures, the better abstract is given. In generating the abstracts, we gave the control parameters with respect to the significance degree as follows. We chose $(\alpha, \beta) = (0.6, 0.4)$ from a set of parameters with which the product of R_W and R_O was maximal; $\gamma = 0$ was set because we took no account of personalization in this experiment.

Fig. 6(a)–(f) illustrates the temporal composition of both video abstracts for video1 to video5. In this figure, the event

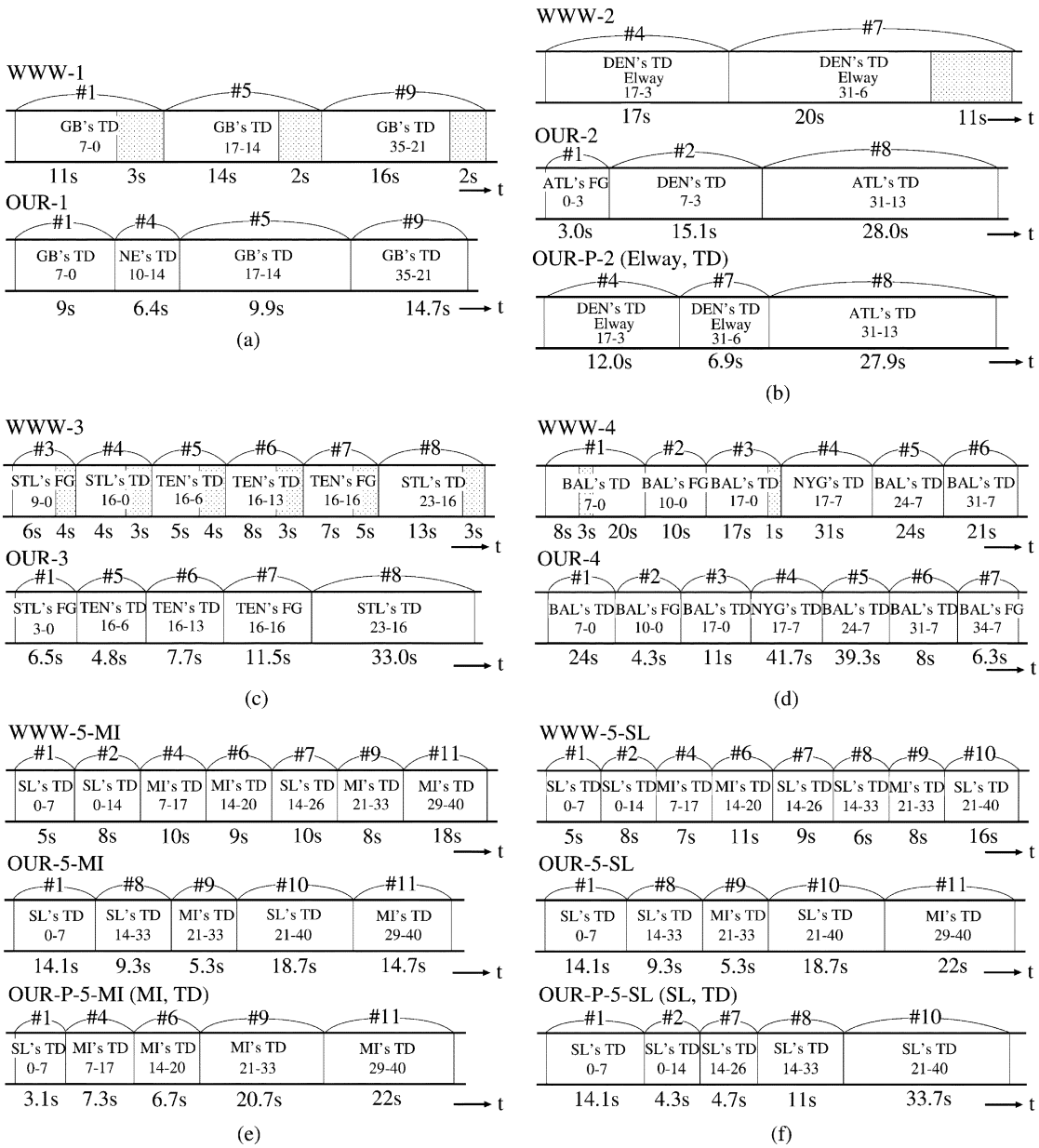


Fig. 6. Comparison between WWW and our abstracts: (a) video1, (b) video2, (c) video3, (d) video4, (e) video5 (MI), and (f) video5 (SL). The shaded portion in the WWW abstracts indicates the post-play shots.

number (#) and the length (s) of each shot are depicted. Each length of the resultant OUR-1,…, OUR-5-MI, OUR-5-SL was 40, 46, 64, 135, 58, and 69 s, respectively. The obtained R_W and R_O for each video are shown in the OUR column of Table VI. As indicated in Fig. 6 and Table VI, the selected highlights were satisfactory for video1, video3 and video4. This method tends to select the first event as a highlight because its rank is always equal to one.

In contrast to them, both abstracts showed quite a difference for video2. Events #1, #2, and #8 were included in OUR-2 due to their high rank. It should be noted that WWW-2 contains only events #4 and #7 about John Elway, who was once a star player in the NFL and retired after this game. This indicates that WWW-2 was produced according to the editor's special intention. In a sense, it was not merely the abstract of the game. Following the WWW abstracts for video5, we notice that the highlights in the abstract can be changed depending on each standpoint. Namely, WWW-5-SL (MI) was designed so that the SL (MI)s fans can see more SL (MI)'s active scenes. Our frame-

TABLE VI
RESULT OF R_W AND R_O FOR EACH VIDEO. OUR AND OUR-P REPRESENT OUR GENERATED ABSTRACT AND OUR PERSONALIZED ABSTRACT, RESPECTIVELY

	OUR		OUR-P	
stream	R_W	R_O	R_W	R_O
video1	3/3	3/4	–	–
video2	0/2	0/3	2/2	2/3
video3	4/6	4/5	–	–
video4	6/6	6/7	–	–
video5(MI)	3/7	3/5	5/7	5/5
video5(SL)	4/7	4/5	5/7	5/5

work is able to tackle this issue by means of specifying "Elway" or the name of the favorite team in the profile. We will discuss this in Section V-C.

We compare both abstracts from the shot aspects. The WWW abstracts effectively included the shots of players who rejoiced at their good plays or fans who cheered their team. These correspond to the post-play shots in this method. At the current stage, higher priority is given to the live and replay shots that show the event itself than the other shots. We have to develop a way of including more shots related to the most interesting highlight. Moreover, some lengthy shots in the WWW abstracts were shortened through edit process, but this method only has possibility to cut still segments from the beginning of the shot.

Let us discuss the quality of video abstracts. Recently, He *et al.* [4] have proposed the desirable attributes for video abstracts, characterized by the four C's: conciseness, coverage, context, and coherence. For conciseness and coverage, although they seem to have a tradeoff relationship, the abstract should have sufficient highlights in a compact form. For coherence, on the other hand, the highlights should appear in original temporal order. Due to this property, we can understand the game flow naturally. Therefore, we think that the generated abstract meets these requirements. If we generated abstracts with the existing methods, e.g., [2], [3], mainly on the basis of surface features of the video stream, satisfactory results could not be achievable. They would probably detect a lot of events besides the significant score events. Keep in mind, however, that this method is exclusively for the broadcasted sports video, not for the general video, utilizing the domain specific knowledge. It is worth noting that without the knowledge it is definitely impossible to produce good video abstracts.

C. Effect of Personalization

We first investigated the effect of personalization, changing the abstract length. It was assumed that the defeated team of the game was specified as the favorite team in the profile. Specifically, the teams, "NE", "ATL", "TEN", "NYG' ", and "MI" were given for video1 through video5, respectively. Because the defeated teams were thought to have less highlights in the game, it is of interest to examine how such preferences were reflected upon in the personalized abstracts we produced. These were compared with the nonpersonalized abstracts such that no favorite team was specified at all. In what follows, the control parameters were set $(\alpha, \beta, \gamma) = (0.6, 0.4, 0.7)$, where α and β were the same with the previous experiment; γ was given experimentally.

For video1 to video5, we measured the *inclusion ratio*, defined as the ratio of the length of shots which are concerned with the specified team to the total length, as we were increasing the abstract length. Fig. 7 shows the relationship between the inclusion ratio and the abstract length for the personalized and nonpersonalized abstracts. As can be seen, the personalized abstracts had more shots of the specified team than the nonpersonalized ones if its length was less than 350 s. The effect of personalization was outstanding while the abstract length was short. Finally, both abstracts became the same because all the highlights about the specified team were fully selected. Note that the ratio converged at about 0.35. Since the number of

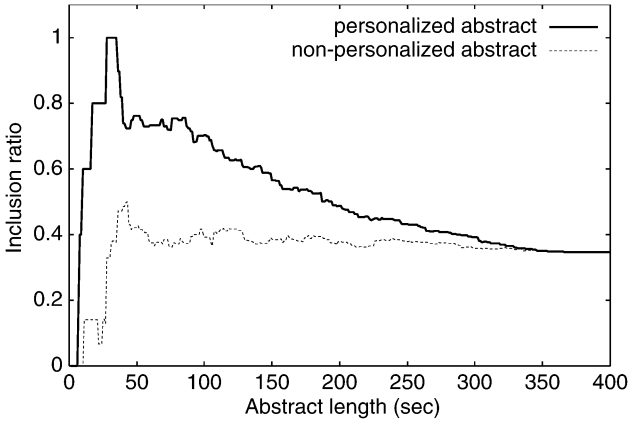


Fig. 7. Relationship between the inclusion ratio and the abstract length for the personalized and nonpersonalized abstracts.

TABLE VII
PROFILE DESCRIPTIONS

Preference	Spec2 & Spec4					Spec5
Player	Martin	Elway	Holt	Ja.Lewis	Carter	none
Team	NE	ATL	TEN	NYG	MI	none
Event	TD					FG
Length(sec)	60(Spec2)/120(Spec4)					60

events about the defeated team was 15 and that of all the events was 45, the final ratio agreed with that value assuming that the length of each event was constant. From this result, it is clear that the profile descriptions act as a bias to control highlight selection.

Next, we carefully examined the change of the abstract contents in accordance with the profile descriptions under setting their length 60 and 120 s. We provided Spec1,…, Spec4 for different profile descriptions. In Spec1 and Spec3, its length was 60 and 120 s; no preference was specified. In Spec2 and Spec4, all the profile descriptions were instantiated, as summarized in Table VII.

Tables VIII–XII indicate the contents of each abstract generated by this method according to Spec1 to Spec4. In each table, "Matched preference" means that the items about the event are matched with the profile descriptions. If the event is matched with them, we simply call it a matched event. The 4-symbol string in the cells of Tables VIII–XII represents each condition of the pre-play, live, replay and post-play shots for the event. The symbols "Y", "N", and "$-$" represent that "the shot is included in the abstract," "the shot is not included in it," and "the shot does not exist in the original video," respectively. For example, the string $NY - Y$ indicates that the pre-play shot is not included in the generated abstract; the live and post-play shots are included; the replay shot did not exist in the original video. As Tables VIII–XII indicated, the video abstract consisted mainly of the live shots of the matched events. In addition to the live shot, its replay shot was included if more than one item in the descriptions was matched. This case was often found for Spec3 and Spec4, where the abstract length was 120 s.

Subsequently, we investigated the temporal composition of the generated abstracts in further detail according to Spec1,

TABLE VIII
ABSTRACT CONTENTS FOR VIDEO1

No.	Matched preference	Spec1	Spec2	Spec3	Spec4
#1	TD	N Y N N	N N N N	N Y N N	N Y N N
#2	NE	N N – N	N N – N	N N – N	N N – N
#3	TD,NE	– N N N	– Y N N	– N N N	– Y Y N
#4	TD,NE	– Y N N	– Y N N	– Y N N	– Y Y N
#5	TD	– Y N N	– N N N	– Y N N	– Y N N
#6		– N – N	– N – N	– N – N	– N – N
#7	TD	– N N N	– N N N	– N N N	– N N N
#8	TD,NE,Martin	N N N N	N Y N N	N Y N N	N Y Y N
#9	TD	N Y N N	N Y N N	N Y Y N	N Y N N

TABLE IX
ABSTRACT CONTENTS FOR VIDEO2

No.	Matched preference	Spec1	Spec2	Spec3	Spec4
#1	ATL	– Y – N	– N – N	– Y – N	– N – N
#2	TD	– Y N N	– N N N	– Y Y N	– Y N N
#3		N N – N	N N – N	N N – N	N N – N
#4	TD,Elway	– N N N	– Y N N	– N N N	– Y N N
#5	ATL	N N – N	N N – N	N N – N	N N – N
#6	TD	N N N N	N N N N	N N N N	N N N N
#7	TD,Elway	N N N N	N Y N N	N Y N N	N Y N N
#8	TD,ATL	– Y N N	– Y N N	– Y Y N	– Y Y N
#9		N N N N	N N N N	N Y N N	N N N N
#10	TD,ATL	– Y N N	– Y N N	– Y N N	– Y N N

TABLE X
ABSTRACT CONTENTS FOR VIDEO3

No.	Matched preference	Spec1	Spec2	Spec3	Spec4
#1		N Y – N	N N – N	N Y – N	N Y – N
#2		– N – N	– N – N	– Y – N	– Y – N
#3		– N – N	– N – N	– Y – N	– Y – N
#4	TD,Holt	– N N N	– Y N N	– Y N N	– Y Y N
#5	TD,TEN	N N N N	N Y N N	N Y Y N	N Y Y N
#6	TD,TEN	– Y N N	– Y Y N	– Y Y N	– Y Y N
#7	TEN	– Y – N	– N – N	– Y – N	– Y – N
#8	TD	N Y Y N	N Y N N	Y Y Y Y	N Y Y N

Spec2 and Spec5. Their length was fixed as 60 s. As shown in Table VII, Spec5 indicated that the favorite event was specified as "FG"; no other preferences were given. Fig. 8(a)–(e) shows the temporal composition of each abstract. We understand that even though the abstract length was comparatively short, still live shots and motion replay shots were occasionally added to motion live shots when the event was actually specified in the viewer's profile. It should be worth noting that the contents of the abstract drastically change depending on the profile descriptions. In particular, the considerable effect of personalization was achieved when "FG" was specified. All the "FG" scenes in the sample streams were included in the personalized abstracts for Spec5. In contrast, only three "FG" scenes were included in the nonpersonalized abstracts for Spec1.

TABLE XI
ABSTRACT CONTENTS FOR VIDEO4

No.	Matched preference	Spec1	Spec2	Spec3	Spec4
#1	TD	N Y N N	N Y N N	N Y N N	N Y N N
#2		N N – N	N N – N	N N – N	N N – N
#3	TD	N Y N N	N N N N	N Y N N	N Y N N
#4	TD,NYG	N Y N N	N Y N N	N Y Y N	N Y Y N
#5	TD	N Y N N	N Y N N	N Y Y N	N Y N N
#6	TD,Ja.Lewis	N N N N	N Y N N	N Y N N	N Y Y N
#7	TD	N N – N	N N – N	N Y – N	N Y – N

TABLE XII
ABSTRACT CONTENTS FOR VIDEO5

No.	Matched preference	Spec1	Spec2	Spec3	Spec4
#1	TD	N Y Y N	N Y N N	Y Y Y N	N Y N N
#2	TD	– N N N	– N N N	– N N N	– N N N
#3		N N – N	N N – N	N N – N	N N – N
#4	TD,MI	N N N N	N Y N N	N N N N	N Y N N
#5		N N – N	N N – N	N Y – N	N N – N
#6	TD,MI	– N N N	– Y N N	– Y N N	– Y Y N
#7	TD	– N N N	– N N N	– Y N N	– N N N
#8	TD	– Y N N	– N N N	– Y N N	– Y N N
#9	TD,MI,Carter	N Y – N	N Y – N	N Y – N	N Y – N
#10	TD	– Y N N	– N N N	– Y Y N	– Y N N
#11	TD,MI	N Y N N	N Y N N	N Y Y N	N Y Y N

Now, let us recall the previously generated abstracts for video2 and video5 that were much different from the WWW abstracts, as stated in the last section [see Fig. 6(b), (e) and (f)]. With the profile describing the favorite player and event as "Elway" and "TD", we successfully obtained a personalized abstract for video2, represented as OUR-P-2 (Elway, TD), as shown in the lower figure of Fig. 6(b). It is more similar to the man-made WWW-2 than OUR-2. Its shots were really replaced with those related to Elway. Likewise, we produced the two personalized abstracts for video5, i.e., OUR-P-5-MI (MI,TD) and OUR-P-5-SL (SL,TD), which are shown in the lower figures of Fig. 6(e) and (f). The two measures of R_W and R_O evaluating the quality of the abstract were increased, as indicated in the OUR-P column of Table VI. We confirm that both abstracts have a lot of the same highlights. These results demonstrate that based on personal preferences and interests, this method offers a framework to generate valid personalized abstracts from diverse aspects.

D. Discussion

1) *Generality of this method*: This method is characterized by the use of the gamestats and the overlay. The gamestats are easy to obtain because they are nowadays available at various popular websites on the Internet. On the other hand, the overlays are sure to appear in the live sports programs. In addition, our inspection supports the assumption that it is present with high probability when an event actually happens. Accordingly, utilizing the textual information in the overlay is

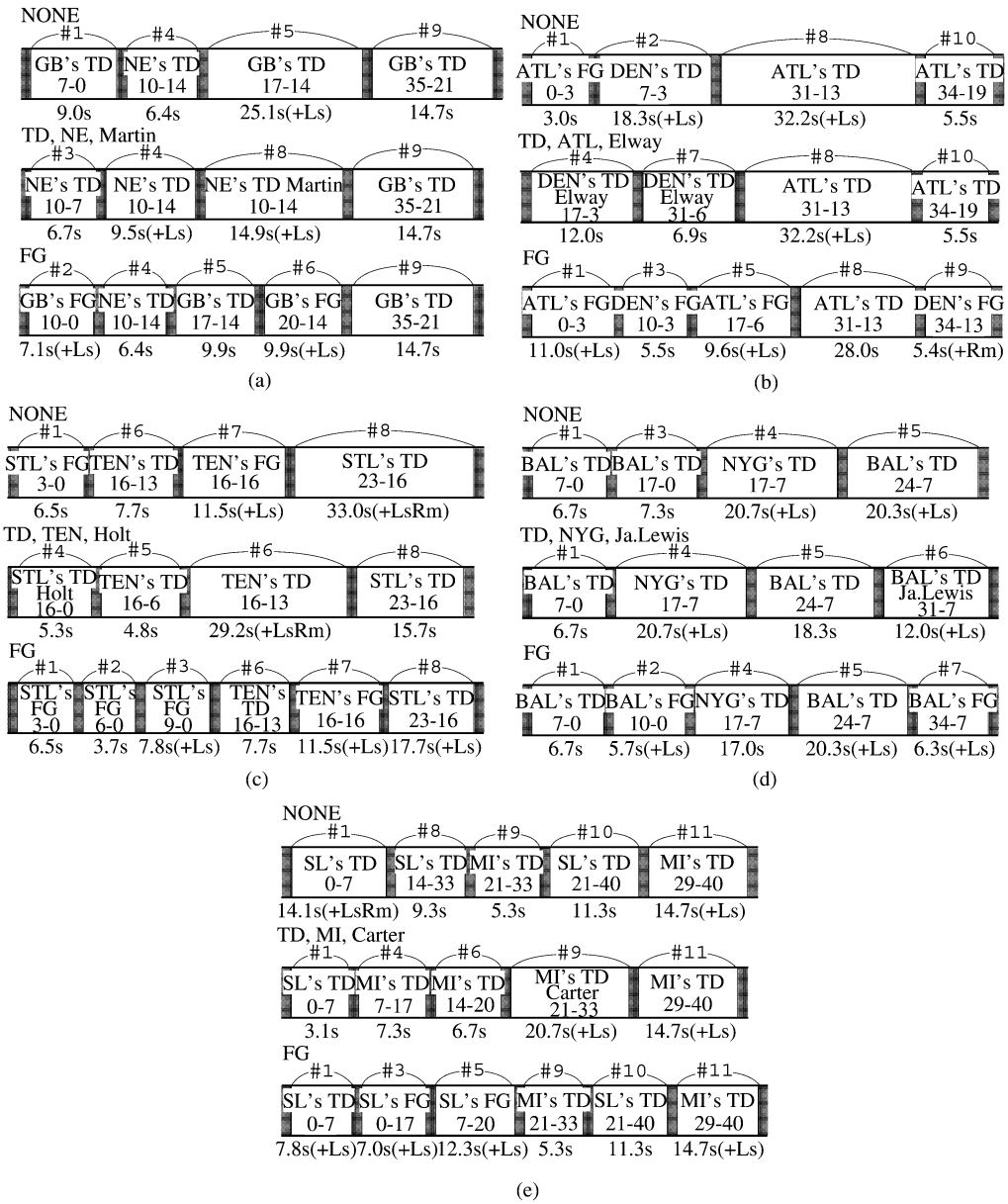


Fig. 8. Temporal composition of each abstract: (a) video1, (b) video2, (c) video3, (d) video4, and (e) video5. The upper, middle, and lower figures are the abstract for Spec1, Spec2, and Spec5, respectively. The shaded portion indicates the interval when video captions are superimposed. The symbols, + Ls and + Rm mean the still live shot and the motion replay shot are added to the motion live shot, respectively.

promising in handling broadcasted sports video. Its disadvantage is that currently we have to define the overlay model for each video stream to deal with the variations with respect to their configurations and appearing locations. For each of the different TV programs, it is necessary to describe the corresponding overlay model. To facilitate this task, we have provided an overlay model editor, which allows us to extract the layout structure of the overlay from the display of an image frame.

The idea of the significance degree for each event, as stated in Section IV-C, is readily applicable to the different two-team sports such as baseball and basketball. In our opinion, this method could have potential for wide applicability to a variety of sports. However, since the domain knowledge is taken into account, verification of its applicability through experimental considerations is our future work.

2) *Detection of significant events*: Focusing on the descriptions of the gamestats which can be viewed as external metadata, this method attempts to detect significant events through overlay recognition. Unfortunately, the gamestats cannot be obtained until the game is over. Consequently, when using them, we can point out the following problems: 1) The events that are not described in the gamestats are never detected, and 2) the abstracts cannot be generated in real time. To solve these problems, we should introduce a more flexible method to detect the significant events. We think methods to integrate the video analysis and the audio/text analysis [18]–[21], which have been extensively developed in recent years, promising.

3) *Selection of highlights*: Highlight selection is naturally formulated as the well known *knapsack problem*. Its brief definition is as follows: given items of different values and vol-

umes, find the most valuable set of items which fit in a knapsack of fixed volume. In this case, values, volumes, and fixed volume correspond to significance degree of the event, shot length, and abstract length, respectively. We implemented an alternative way of highlight selection with solution of the knapsack problem, but then we missed the most significant event in some cases. This is because the selection is based on an overall maximization strategy. In contrast to this, our method shown in Section IV-D is a simple greedy strategy. In our application, what is important is not maximization of the sum of significance degrees but proper selection of the most significant events. We think that the highlight selection algorithm is effective but needs to be tested in various application domains.

4) *Profile*: The notion of the profile is essentially derived from the research on information filtering. The profile we introduced into this method is one of the simplest versions to directly specify the preferences. It is something like a catalog of the viewer's preferences. All the items in the profile, i.e., favorite players, teams and events, were handled equally. To describe them in detail, some modifications will be needed. For instance, grades should be considered to describe each item in the profile. This allows us to represent how large we like the team or so. In addition, it is of great importance to learn the preferences while generating the video abstracts. Approaches to learning to personalize should be explored [27].

5) *Evaluation of video abstracts*: The task of video abstraction is considerably subjective. Accordingly, it is very difficult to evaluate the quality of the video abstracts objectively. It seems that there is no correct answer to video abstraction. Our solution is the comparison between man-made and machine-made abstracts. However, quantitative evaluation in terms of difference or similarity between them is an open problem. We have to continue to pursue an evaluation scheme taking account of human factors and subjective aspects.

VI. Concluding Remarks

This paper addressed a method of automatically generating abstracts of broadcasted American football video. Based on the detected significant events by recognizing the textual overlays, we can make a favorable video abstract in an arbitrary length, reflecting on the viewer's preferences and requirements. We believe that the use of the gamestats is a new idea to link the video contents with useful external metadata. From the experimental results, we verified that an hour-length video can be compressed into a minute-length personalized abstract, and that the profile can act as a bias to generate video abstracts. It was also clarified that this method is capable of selecting the highlights similar to those in the man-made abstracts.

Three sorts of significance degrees play a central role in highlight selection. The degrees about the event rank and the occurrence time suitably express our feeling or intuition that what is significant is the event that turns the game or that occurs at the final stage of the game. Also the degree about the profile relatively magnifies the event's significance to meet the viewer's preferences. We think this attempt to numerize semantical information contributes to video content processing.

The underlying idea of generating a video abstract is to connect the highlight scenes in original temporal order. In fact, actual game digests broadcasted in news TV programs are made up in the similar manner. Time-ordered highlight based video abstraction is appropriate particularly for the sports video. If we intend to make video abstracts like movie trails out of an original movie, this idea may have to be expanded because their scenes are not always arranged in such order.

Finally, let us mention the remaining problems to be explored. The first is to examine this method for different two-team sports such as baseball and basketball to verify its generality. It is also of interest to investigate whether it is applicable to other kinds of sports such as golf and athletics. The second is an editing procedure for shots. Compared with the man-made abstract, this abstract tends to have lengthy shots. We are interested in a tailoring mechanism for shots, adjusting for the total abstract length. In addition, visual effects derived from shot change operations such as dissolve, fade and wipe should be considered. The third is to seek for a sophisticated way of refining the profile. The relationship between abstraction and personalization should be made clear. The last point is to construct a scheme to evaluate the generated video abstracts from various viewpoints. In this case, we should take account of the subjective human factors, for example, global impression of the abstract.

References

[1] P. Aigrain, H. J. Zhang, and D. Petkovic, "Content-based representation and retrieval of visual media: A state-of-the-art review," *Multimedia Tools Applicat.*, vol. 3, pp. 179–202, 1996.

[2] M. A. Smith and T. Kanade, "Video skimming and characterization through the combination of image and language understanding techniques," in *Proc. IEEE CVPR*, 1997, pp. 775–781.

[3] R. Lienhart, S. Pfeiffer, and W. Effelsberg, "Video abstracting," *Commun. ACM*, vol. 40, no. 12, pp. 55–62, 1997.

[4] L. He, E. Sanocki, A. Gupta, and J. Grudin, "Auto-summarization of audio-video presentations," in *Proc. ACM Multimedia*, 1999, pp. 489–498.

[5] J. H. Oh and K. A. Hua, "An efficient technique for summarizing videos using visual contents," in *Proc. IEEE ICME*, 2000.

[6] N. Babaguchi, "Toward abstracting sports video by highlights," in *Proc. IEEE ICME*, 2000, pp. 1519–1522.

[7] M. M. Yeung and B.-L. Yeo, "Video visualization for compact presentation and fast browsing of pictorial content," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 5, pp. 771–785, Oct. 1997.

[8] S. Uchihashi, J. Foote, A. Girgensohn, and J. Boreczky, "Video manga: Generating semantically meaningful video summaries," in *Proc. ACM Multimedia*, 1999, pp. 383–392.

[9] S.-F. Chang and H. Sundaram, "Structural and semantic analysis of video," in *Proc. IEEE ICME*, 2000.

[10] C. Toklu, S.-P. Liou, and M. Das, "VIDEOABSTRACT: A hybrid approach to generate semantically meaningful video summaries," in *Proc. IEEE ICME*, 2000.

[11] D. Riecken, "Personalized views of personalization," *Commun. ACM*, vol. 43, no. 8, pp. 27–28, 2000.

[12] H. Hirsh, C. Basu, and B. D. Davison, "Intermediaries personalize information streams," *Commun. ACM*, vol. 43, no. 8, pp. 96–101, 2000.

[13] B. Merialdo, K. T. Lee, D. Luparello, and J. Roudaire, "Automatic construction of personalized TV news programs," in *Proc. ACM Multimedia*, 1999, pp. 323–331.

[14] B. Smyth and P. Cotter, "A personalized television listings service," *Commun. ACM*, vol. 43, no. 8, pp. 107–111, 2000.

[15] R. Jasinschi, N. Dimitrova, T. McGee, L. Agnihotri, and J. Zimmerman, "Video scouting: An architecture and system for the integration of multimedia information in personal TV applications," in *Proc. IEEE ICASSP*, 2001, pp. 1405–1408.

[16] T. Echigo, M. Kurokawa, A. Tomita, H. Miyamori, and S. Iisaku, "Video enrichment: Retrieval and enhanced visualization based on behaviors of objects," in *Proc. ACCV*, 2000, pp. 364–369.

[17] D. Zhong and S.-F. Chang, "Structure analysis of sports video using domain models," in *Proc. IEEE ICME*, 2001, pp. 920–923.
[18] Y. Chang, W. Zeng, I. Kamel, and R. Alonso, "Integrated image and speech analysis for content-based video indexing," in *Proc. IEEE ICMCS*, 1996, pp. 306–313.
[19] N. Babaguchi, S. Sasamori, T. Kitahashi, and R. Jain, "Detecting events from continuous media by intermodal collaboration and knowledge use," in *Proc. IEEE ICMCS*, vol. 1, 1999, pp. 782–786.
[20] Y. Rui, A. Gupta, and A. Acero, "Automatically extracting highlights for TV baseball programs," in *Proc. ACM Multimedia*, 2000, pp. 105–115.
[21] N. Babaguchi, Y. Kawai, and T. Kitahashi, "Event based indexing of broadcasted sports video by intermodal collaboration," *IEEE Trans. Multimedia*, vol. 4, pp. 68–75, Mar. 2002.
[22] N. Babaguchi, Y. Kawai, Y. Yasugi, and T. Kitahashi, "Linking live and replay scenes in broadcasted sports video," in *Proc. ACM Multimedia 2000 Workshop on Multimedia Information Retrieval*, 2000, pp. 205–208.
[23] [Online]. Available: http://www.superbowl.com/xxxvi/history/box-scores/
[24] [Online]. Available: http://www.nfl.com/gamebooks/minstl.pdf
[25] [Online]. Available: http://www.superbowl.com/xxxvi/multimedia/past-superbowls.html
[26] [Online]. Available: http://www.nfl.com/nflfilmstv/
[27] H. Hirsh, C. Basu, and B. D. Davison, "Learning to personalize," *Commun. ACM*, vol. 43, no. 8, pp. 102–106, 2000.

Noboru Babaguchi (M'90) received the B.E., M.E., and Ph.D. degrees in communication engineering from Osaka University, in 1979, 1981, and 1984, respectively.

He is currently a Professor, Department of Communication Engineering, Osaka University. From 1996 to 1997, he was a Visiting Scholar at the University of California at San Diego, La Jolla. His research interests include image analysis, multimedia computing and intelligent systems, currently content based video indexing and summarization. He has published over 100 journal and conference papers and several textbooks. He is on the editorial board of *Multimedia Tools and Applications*, *New Generation Computing*, and the *Journal of Information Processing Society of Japan*.

Dr. Babaguchi is a member of the ACM, IEICE, IPSJ, and JSAI. He served as a workshop Co-chair of 3rd International Workshop on Multimedia Information Retrieval (MIR2001), and on the program committee of international conferences in these fields.

Yoshihiko Kawai received the B.E. and M.E. degrees in informatics and mathematical science from Osaka University, in 1999 and 2001, respectively.

He is now with NHK (Japan Broadcasting Corporation), Tokyo, Japan. His research interests include video content analysis.

Takehiro Ogura received the B.E and M.E. degrees in informatics and mathematical science from Osaka University, Osaka, Japan, in 2001 and 2003, respectively.

He is now with NHK, Japan. His research interests include design of video portals and personalized media systems.

Tadahiro Kitahashi (M'70) received the B.E., M.E., and D.E. degrees in communication engineering from Osaka University, Osaka, Japan, in 1962, 1964, and 1968, respectively.

He is currently a Professor at Kwansei Gakuin University, Hyogo, Japan, and a Professor Emeritus of Osaka University. His recent research interests include media engineering, especially conversion between image media and linguistic media, computer human interface for personal navigation systems and generating virtual images.

Dr. Kitahashi is a Fellow of the IEICE, and a member of the IPSJ, the JSAI, and the AAAI.