# PLAYWATCH: CHART-STYLE VIDEO PLAYBACK INTERFACE

*Kiyoshi Tanaka, Tsutomu Sasaki,*
*Yoshinobu Tonomura, Tadashi Nakanishi*

NTT Cyber Solutions Laboratories
Yokosuka-shi, Kanagawa, Japan   239-0847

*Noboru Babaguchi*

Osaka University
Graduate School of Engineering
Suita-shi, Osaka, Japan   565-0871

## ABSTRACT

This paper proposes the chart-style video playback interface PlayWatch; it displays a chart of semantic indices for locating video scenes. The main features of PlayWatch are: 1) the user understands the distribution of the scenes because PlayWatch shows the indices in order. 2) The user can access a desired scene directly through the indices since they also act as link buttons. This paper also describes the evaluation of PlayWatch. Experiments on scene searching show that PlayWatch is effective in accessing precisely indexed scenes.

## 1. INTRODUCTION

Due to the popularity of digital video equipment, such as DVD players, we need tools that can index scenes and make it easy to locate the desired scenes. Current indexing methods are not really useful as they are too coarse; they show only the beginning of each chapter and so miss the segments in each scene.

Even if detailed indices are available, there is the problem of how to display them. DVDs show the indices as fixed thumbnails of the indexed scenes. If the detailed indices are displayed using thumbnails, they would be so numerous that the user would take too long to find anything.

The scene index is essentially composed of the time position of the scene in a video program and the scene description. The scene description is written in the characteristic keyword that expresses the scene, such as names of actors or events in the scene. If the descriptions are limited to the several keywords about some specific scene features, the scene indices can be categorized to the same number of categories as the keywords. Displaying the indices by putting them in order of the categories yields an easy to understand interface even if there are many scene indices.

In this paper, we propose a chart-style video playback interface, called PlayWatch, that visualizes indices as points on a chart, the x-axis of which shows the time line while the y-axis shows representations of the indices, e.g. scene events. Because this chart orders the indices, users can better understand the video program from the distribution of the indices on the chart. Moreover, since the index points are buttons that link to the indexed scenes in the video, users can access the desired scenes rapidly. In this paper, we also report the evaluation results of PlayWatch in terms of search time.

## 2. RELATED WORK

Visual interfaces of indices have been researched for video retrieval [1, 2]. They use scene or shot thumbnails to inform the users of what the video consists of.

There are the other video browsing interfaces [3, 4, 5] proposed for summarizing video programs. The newspaper-like interface called Video Paper [5] is one of such interfaces. Video Paper expresses the scenes using indices, each of which consists of a thumbnail and a descriptive text. Video Paper can be useful for structured programs that consist of many different scenes such as news programs, but it is not suited for video programs that are not structured or that have many detailed indexed scenes.

The user interface of video authoring systems [6, 7] and video analysis systems [8, 9, 10] uses scene indexing to help the user in understanding what each scene shows. In these systems, the index thumbnails are mainly used to check scene content, not to access the scenes.

Several interfaces that arrange semantic representations on a time line have been proposed [11, 12, 13]. They are intended to show the structures of the video programs and are not designed to support retrieval by users.

## 3. INDEX CHART FOR VIDEO PLAYBACK

The index of the scene $i(t, D)$ is represented by time position $t$ and scene description $D$. A scene description is an explanation of the scene based on the 5W1H (what, who, where, when, why and how) approach.

$$
\begin{aligned}
i(t, D) \quad & : \text{index of the scene} \\
t \quad & : \text{time position of the scene} \\
D = \quad & \{d_{what}, d_{who}, d_{where}, d_{when}, d_{why}, d_{how}\} \\
& : \text{description of the scene}
\end{aligned}
$$

**Fig. 1**. Chart of indices: PlayWatch



**Fig. 3**. System configuration
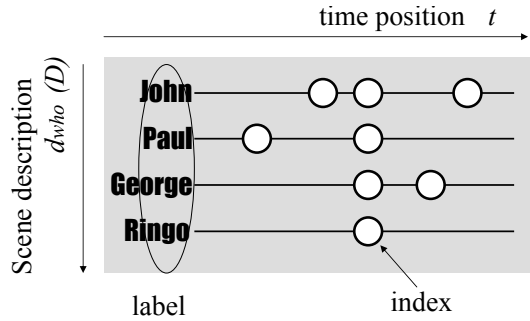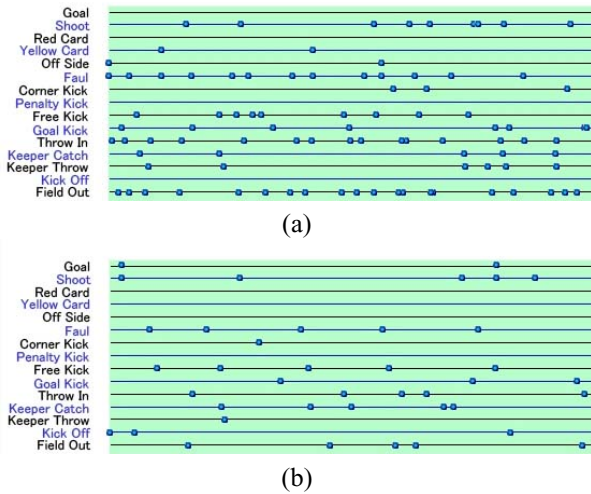


(a)



(b)

**Fig. 2**. Charts of two different soccer games. (a) is a more exciting game than (b). This can be found from the distribution of the indices.

PlayWatch presents a chart of the indices: the x-axis plots $t$ while the y-axis shows scene description $D$ as shown in Figure 1. Figure 1 is an example that uses the "who" item $d_{who}$ from $D$ as a scene description. PlayWatch also displays each scene description as a label for notifying user's classification of the indices.

The scene description is a keyword that determines the characteristic of the scene; examples include the name of an actor and an action such as "goal" in a soccer game. It is clear that the suitable scene descriptions change with the type of program. Moreover, while the user switches to a different scene description set for the same program if his/her intention changes, the index chart needs to change its composition with the selected scene description set.

PlayWatch has two unique features. First, the PlayWatch chart, like a musical score, allows users to acquire a lot of information about the program such as which actors are present and the frequency of their appearance. It is especially useful for 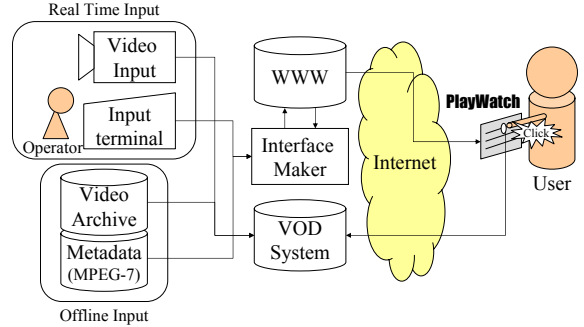sports programs; the index chart shows whether the game is exciting or not; if the index points are widely separated, the events are not frequent. Figure 2 shows two charts of different soccer games. From this figure, it is clear that (a) is a more exciting game than (b).

The second feature of PlayWatch is direct access to the scenes because the index entries are buttons that allow the user to jump to particular scenes.

These two features lead users to find the desired scene rapidly because the users search the scene from scene distribution in the program and access the scene directly with a simple button click.

## 4. SYSTEM CONFIGURATION

We constructed PlayWatch as an Internet application for broadband networks. PlayWatch will be applied to video programs including live sports events so it was developed as a real time system.

Figure 3 shows the system configuration. The operator initially inputs the scene descriptions from the input terminal. Since the scene descriptions are semantic information, they are difficult for automatic methods to acquire precisely. Moreover, the scene descriptions may need to be input in real time or quasi-real time to handle live events such as the replacement of one player with another.

The operator inputs one or more of the 5W1H items of each scene. If the candidates of scene descriptions are predefined, the description is input from such candidates by selecting it from a menu using our input tool [14]. The description input together with the input time yields the index data.

The scene descriptions of an offline program can be derived from the metadata of the program such as MPEG-7 [15]. The text annotation of MPEG-7 and its media time position are converted into a scene index.

The system was implemented on the WWW. Here, the chart is created in layered HTML using the interface maker of a CGI program written in perl. The coordinates of index points are determined by the time position and the scene

**Table 1**. Examples of retrieval tasks.

| Type | Examples |
|---|---|
| Search for a scene | Find the first goal scene.<br>Find the starting kick-off scene of the match.<br>Find the final scene of the match. |
| Jump to a particular time position | What is the name of the player who enters the game as a replacement 8 minutes into the program?<br>What is the name of the player who make the goal about 10 minutes after the program started?<br>What is the name of the player who make the first goal about 4 minutes into the program? |

description of the scene index data. The label is also written by the scene description. These objects are overlaid on the chart in HTML. The index points are set up as hyper links to call each scene from the VOD (Video on Demand) system. Therefore, the chart is made dynamically from the descriptions, different sets of which can be selected by the user according to his/her intention.

## 5. EVALUATION

We compared PlayWatch to conventional interfaces in an experiment. Several subjects searched for scenes in a futsal[1] program. In the experiment, we gave the subjects two types of problems and timed them until the answers were given. The subjects were three males and six females ranging in age from 20 to 40 years old. Their occupations included company employees, housewives, and students. They had no experience in using PlayWatch, only basic skills in operating personal computers running the Windows PC operating system.

Considering that the main purpose of scene searching is to find a particular scene and then jump to that scene, we gave the subjects two kinds of problems as shown in Table 1. The tasks consisted of; "search for a scene" task (scene retrieval) and "jump to a particular time position" task (time access).

The same tasks were also conducted using the movie player shown in Figure 4 (a) and the scene thumbnail interface in Figure 4 (b). The movie player offered a time slider for jumping to an arbitrary time position and the thumbnail interface offered scene thumbnails for showing the scene image and jumping to the scene. Since thumbnails were generated from the images at scene changes, the time positions of indices differed from those of the thumbnail interface and PlayWatch. We compared the search times required by the three methods.

According to Figure 5, PlayWatch was most effective in finding scenes from keywords (scene retrieval). Moreover, five of the subjects could execute this task within 15 seconds with PlayWatch, and no less than three could do it within
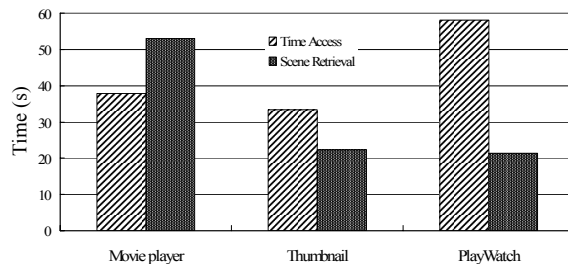


(a)



(b)

**Fig. 4**. Movie player (Apple QuickTime player) and thumbnail interface compared to PlayWatch.



**Fig. 5**. Retrieval result with the scene retrieval tasks.

---

[1]Futsal is the mini-version of soccer that is played between two teams of 5 players.

only 10 seconds. This is because PlayWatch showed semantic indices that allowed the subjects to access the scene directly.

However, PlayWatch was the slowest at jumping to a particular time position (time access). Although the name of player was narrated in the video, there was no index to indicate a time position of the narration with PlayWatch. Moreover, the PlayWatch chart used in the experiments allowed jumping to only a selected index or one minute intervals, so the subjects had to wait to hear the narration to answer the problem after jumping.

Although the thumbnail interface has the same waiting restriction as PlayWatch, it was fastest at finding the answer. This is considered to be because it allowed the subjects to jump to specific scenes while checking the time position by character and the scene by image; PlayWatch did not show the time position by character.

This result shows that visual and character information of a scene is important in locating scenes. Although the effectiveness of PlayWatch was confirmed, we believe it is necessary to show visual and character information such as thumbnails and to add fast forwarding and rewinding playback controls to make it easier to find scenes with Play-Watch.

## 6. SUMMARY AND CONCLUSIONS

This paper proposed the chart-style video playback interface PlayWatch. PlayWatch plots scene indices on a 2D chart. The chart allows users to realize the intensity of action in a video and its contents. The user can jump directly to a scene indicated by an index point by simply clicking on it.

From an experiment conducted to evaluate how quickly users could locate a scene, PlayWatch was found to be highly effective in accessing scenes by precise keywords. We intend to extend PlayWatch by adding thumbnails as well as nonlinear playback functions.

## 7. ACKNOWLEDGMENTS

## 8. REFERENCES

[1] Y. Tonomura, A. Akutsu, Y. Taniguchi, and G. Suzuki, "Structured video computing," *IEEE Multimedia*, vol. 1, no. 3, pp. 34–43, 1994.

[2] S. W. Simoliar and H. Zhang, "Content-based video indexing and retrieval," *IEEE Multimedia*, vol. 1, no. 2, pp. 62–72, 1994.

[3] Y. Taniguchi, A. Akutsu, Y. Tonomura, and H. Hamada, "An intuitive and efficient access interface to real-time incoming video based on automatic indexing," in *Proc. ACM Multimedia '95*, pp. 25–33, 1995.

[4] J. Boreczky, A. Girgensohn, G. Golovchinsky, and S. Uchihashi, "An interactive comic book presentation for exploring video," in *Proc. ACM CHI2000*, pp. 185–192, 2000.

[5] J. Graham and J. J. Hull, "A paper-based interface for video browsing and retrieval," in *Proc. ICME 2003*, vol. II, pp. 749–752, 2003.

[6] F. Shipman, A. Girgensohn, and L. Wilcox, "Hyper-Hitchcock: Towards the easy authoring of interactive video," in *Proc. Interact 2003*, pp. 33–40, 2003.

[7] A. C. Long, J. Casares, B. A. Myers, R. Bhatnager, S. M. Stevens, L. Dabbish, D. Yocum, and A. Corbett, "SILVER: Simplifying video editing with metadata," in *Proc. CHI 2003*, pp. 628–629, 2003.

[8] C. Taskiran, J. Y. Chen, A. Albiol, L. Torres, C. A. Bouman, and E. J. Delp, "ViBE: A compressed video database structured for active browsing and search," *IEEE Trans. Multimedia*, vol. 6, no. 1, pp. 103–118, 2004.

[9] J. Vendrig, M. Worring, and A. W. M. Smeulders, "Components and systems for interactive video indexing," in *Proc. ICME 2003*, vol. II, pp. 773–776, 2003.

[10] B. Merialdo, K. T. Lee, D. Luparello, and J. Roudaire, "Automatic construction of personalized TV news programs," in *Proc. ACM Multimedia '99*, pp. 323–331, 1999.

[11] N. Vasconcelos and A. Lippman, "Bayesian modeling of video editing and structure: Semantic features for video summarization and browsing," in *Proc. ICIP 98*, vol. 3, pp. 153–157, 1998.

[12] M. Davis, "Media Stream: An iconic visual language for video representation," in *Human-Computer Interaction: Toward the Year 2000*, pp. 854–866. Morgan Kaufmann Pub., 1995.

[13] A. Haubold and J. R. Kendar, "Analysis and interface for instructional video," in *Proc. ICME 2003*, vol. II, pp. 705–708, 2003.

[14] K. Tanaka, T. Taura, A. Akutsu, H. Jinzenji, and Y. Tonomura, "LiveWatch: The interactive live casting system," *Systems and Computers in Japan*, vol. 35, no. 6, pp. 31–41, 2004.

[15] B. S. Manjunath, P. Salembier, and T. Sikora, *Introduction to MPEG-7 Multimedia Content Description Interface*, John Wiley & Sons Ltd., 2002.