# Synchronized Video: An Interface for Harmonizing Video with Body Movements

*Yoshihiro Watanabe, Hiroaki Ohno, Takashi Komuro, Masatoshi Ishikawa*
Graduate School of Information Science and Technology, University of Tokyo
Hongo 7-3-1, Bunkyo-ku, Tokyo 113-8656, JAPAN
Yoshihiro_Watanabe@ipc.i.u-tokyo.ac.jp

## ABSTRACT

We propose a new media reproduction technology called *Synchronized Video* that supports the learning of new body movements. The proposed technology plays video in synchronization with human body movements. Doing so enables easy video control that allows mapping between action elements contained in the video and the user's actual movements in a harmonious way. In this paper, we describe its metaphor and technological challenges. We also describe a prototype developed to show its effectiveness.

**ACM Classification:** H5.2 [Information interfaces and presentation]: User Interfaces. - Interaction styles.

**General terms:** Design, Human Factors, Measurement

**Keywords:** Video interfaces, Body motion, Exercise.

## INTRODUCTION

The advances made in the fields of sensors, networks, and storage have enabled sharing of various video content around the world. With the diversification of video content, the importance of video interface designs is increasing. The videos that we focus on in this paper are the kinds of content where user interaction with the video is critical. A typical example is video content introducing new movements to users. The purpose of this type of video is to generate interaction that leads users to map displayed action elements to their bodies adaptively.

However, conventional video browsing technology has interfered with the potential of such videos. The problem lies in designs that assume that users browse videos passively and do not consider such interactions. Ineffective methods for providing explicit commands, for example, by operating the video equipment or by giving gestures or voice commands, interfere with the user's concentration on the interaction. There has been some work to try to break out of such constraints [1, 2]. They provide intuitive controls where video playback is based on tracing an object's motion within the video. These studies have shown that the design should take a direction that harmonizes video control and the user's active intentions. However, the proposed scenarios are limited to operations and manipulations carried out by the user on digital terminals, rather than via actual body movements that are captured.

Here we propose a video interface that controls the content automatically based on real-world actions. In particular, we concentrate on harmonizing video with the user's body movements, via a technology that we call *Synchronized Video*.

## SYHNCHRONIZED VIDEO

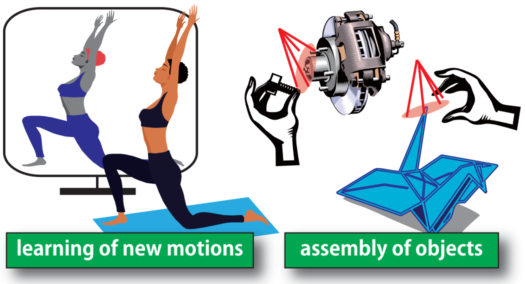

Figure 1: Metaphor of *Synchronized Video*: a video control technique for showing video scenes in synchronization with real-world actions without any explicit commands.

*Synchronized Video* is a video control technique that shows frames in synchronization with real-world actions. Figure 1 depicts the metaphor of this technology. Suitable video content for this type of video interface includes learning forms of exercise and assembly of structural objects. In this paper, we introduce a prototype for learning new movements. In the proposed system, the user feels as if he or she is standing in front of a strange mirror where there is a virtual human performing the correct actions in synchronization. Also, the proposed technology is effective for projecting target images onto real-world objects serving as a screen, for recreating the same situation as in prepared video content.

The synchronization in this proposed interface is achieved without the user performing any explicit operations. The state created in the real world by the user is what directly controls the video. This design allows transparent video control for users. The intended frames could be provided intuitively in a continuous form.

The key to realize the proposed environment is identity recognition for finding similar states between the user and the prepared video. This involves defining diversity of similar states in video content, as well as extracting the required states from scenes including differences caused by body size and the skill-level of the movement and identifying these states continuously. Also, we would like to convert existing videos to a suitable format for *Synchronized Video*. In addition,

Copyright is held by the owner/author(s).
*UIST'09,* October 4-7, 2009, Victoria, BC, Canada.

assuming that a camera is the tool used to capture the real world, the system must compensate for variations caused by differences in viewing locations of the prepared video and the camera.

These technical challenges can be summarized as two issues: what kind of sensing feature is effective and how should the previous knowledge be accumulated in advance? The related technology is considered to include motion capture by a single camera using constraints determined by the structure of the human body, generalized motion models for creating candidates for a specific motion, image retrieval, and so on.

## PROTOTYPE IMPLEMENTATION AND DEMONSTRATION

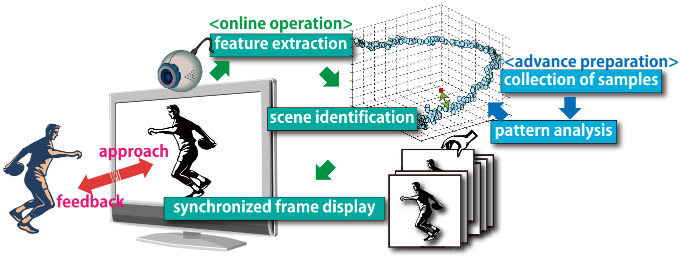

Figure 2: Prototype configuration. A synchronized frame is selected based on the features extracted from a captured scene.

We developed the prototype based on the metaphor described in the previous section. Figure 2 shows its configuration. A single camera captured the user's state. Also, as an initial condition, it was assumed that the user's position was adjusted to be captured as a pose close to that in the prepared video. In this prototype, multiple markers were placed at the same points on the user and the human model in the video in order to simplify the identity recognition. The positions of the markers captured in an image defined image features $\boldsymbol{f}_i$ in a frame $i$. To compensate for a translational gap in an image, the detected positions were converted to relative ones $\bar{\boldsymbol{f}}_i$, with the origin at a marker at the center of the body.

In advance preparation, we collected samples of multiple persons, including actual users and people shown in videos. Principal component analysis (PCA) was applied to the set of image features $\bar{\boldsymbol{f}}$ to obtain the projective transformation $\boldsymbol{U}$ for dimensional reduction with improved recognition robustness. The value $\boldsymbol{g}_i = \boldsymbol{U}\bar{\boldsymbol{f}}_i$ was used as a recognition feature. As the video source, a single representative motion was selected from samples, and the set of its corresponding features $\boldsymbol{g}$ was stored for subsequent online operation.

These data were used to synchronize the displayed video with a user's motion. First, the marker positions $\tilde{\boldsymbol{f}}_t$ were detected from a captured image and converted to recognition feature $\tilde{\boldsymbol{g}}_t$ by using a projective transform $\boldsymbol{U}$ obtained in advance. Next, the nearest feature $\boldsymbol{g}_j$ to the obtained recognition feature $\tilde{\boldsymbol{g}}_t$ was acquired from the video data $\boldsymbol{g}$, and the corresponding image was shown. These two steps were repeated for every captured image.

We demonstrated the operation of the developed prototype based on the described configuration. The video content was swings of table tennis players. The same camera was used for the advance preparation and online operation. The camera resolution was 320 × 240. The frame rate was 200 fps for sample collection and 60 fps for online operation. In this case, several swings of two persons were sampled. PCA was applied to all samples. The swing of one person was used as the displayed video. The number of markers was five, placed at the wrists, thighs, and navel. The image features $\bar{\boldsymbol{f}}_i$ were 8-dimensional, and the recognition features $\boldsymbol{g}_i$ were 3-dimensional. The contribution rate was 95%.

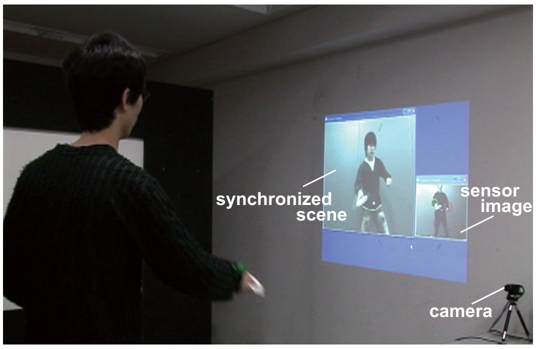

Figure 3: Demonstration. Table tennis swings of the user and a model in the video were synchronized.

Figure 3 shows the demonstrations. We confirmed that the prototype synchronized the video and the users' motions accurately. Users could quickly and intuitively understand the method of using this system, thanks to the simple interactive design. Also, we obtained qualitative confirmation that the users could control the given video freely, recognize the correct form, and smoothly make their actions conform to the displayed video.

## CONCLUSION

In this paper, we have proposed *Synchronized Video*, a new video control technique for harmonizing displayed video with user body movements. This is an interface that allows users to utilize video media distributing new motions and operations. We described the details of its metaphor model and the technological challenges involved. The developed prototype showed the effectiveness of this approach.

As a next step, we plan to focus on extension of identify recognition and interaction evaluation. Also, the importance of videos containing objects and their direct projections to the real world will be increased in the proposed system. Related work that may be applicable includes virtual sculpting [3] and a system that captures 3D scenes and projects images onto a non-flat surface with no distortion in real time [4].

## REFERENCES

1. T. Karrer, et al.: "DRAGON: a direct manipulation interface for frame-accurate in-scene video navigation," CHI'08, pp. 247–250, 2008.
2. P. Dragicevic, et al.: "Video browsing by direct manipulation," CHI'08, pp. 237–246, 2008.
3. C. Skeels, et al.: "ShapeShift: A Projector-Guided Sculpture System," UIST'07, 2007.
4. Y. Watanabe, et al.: "The deformable workspace: A membrane between real and virtual space," IEEE Tabletops and Interactive Surfaces'08, pp. 155–162, 2008.