

# シンクロナイズドビデオ：身体動作と調和するビデオ操作

Synchronized Video: An Interface for Harmonizing Video with Body Movements

渡辺義浩<sup>1)</sup>, 大野紘明<sup>1)</sup>, 小室孝<sup>1)</sup>, 石川正俊<sup>1)</sup>

Yoshihiro Watanabe, Hiroaki Ohno, Takashi Komuro and Masatoshi Ishikawa

1) 東京大学 情報理工学系研究科

(〒113 東京都文京区本郷 7-3-1, 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.

**Key Words:** *Media reproduction, Video interface, Body movements*

## 1. はじめに

センサ、ネットワーク、ストレージ技術の発達とともに、様々なビデオコンテンツが世界中で共有されるようになった。このようなビデオコンテンツの多様化とともに、そのインターフェースデザインが重要な課題となっている。特に、本稿では、ビデオとユーザのインタラクションが鍵となるものに焦点を当てる。代表的な例として、ユーザに新しい動作を紹介するビデオが挙げられる。このタイプのビデオの目的は、ユーザが、提示される動作要素を身体へ適応的にマッピングすることを促すインタラクションを生み出すことである。

しかし、既存のビデオ閲覧技術は、そのようなビデオの潜在的価値を阻害するものであったと考えられる。問題は、ユーザが受動的にビデオ閲覧を行うことを前提とするデザインにある。これまでの操作には、付属機器や、ジェスチャ、音声などに基づくものがあるが、いずれも明示的な指令を提供する非効率なデザインであり、ユーザがビデオとのインタラクションに集中する構成をとっていなかった。一方、このような制約からの脱却を図った技術が提案されている [1, 2]。同技術は、ユーザが映像内の対象物の動きをトレースすることで映像を再生する、直感的なビデオ操作を実現するものである。このように、ビデオ操作とユーザの能動的意図を調和する方向性は強力であるが、同成果はデジタル端末内での操作を前提としたものであった。

本稿では、コンテンツを実世界の状態に応じて、自動的に制御するビデオインターフェースを提案する。特に、我々はビデオとユーザの身体動作を調和させることに重点を置いた。この提案技術を、シンクロナイズドビデオと呼ぶ。

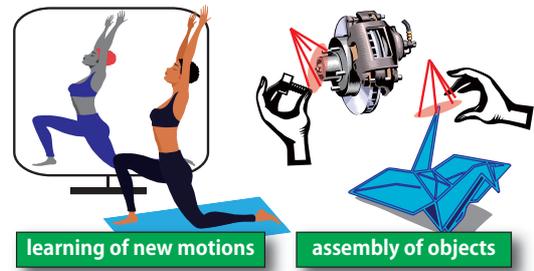


図 1: シンクロナイズドビデオ

## 2. シンクロナイズドビデオ

シンクロナイズドビデオは、実世界の状態と同期して、映像を提示するビデオ操作技術である。図 1 に概念図を示す。このようなビデオ操作に有効なコンテンツには、体操などのフォームや構造物の組み立ての学習などが考えられる。特に、本稿では新しい身体運動を獲得するためのプロトタイプを報告する。提案するシステムにおいて、ユーザは、自身とシンクロして正しい動作を行う仮想人物が眼前に存在する感覚を得ることができる。また、物体が介在する場合には、用意されたビデオコンテンツと同じ状況を再現するために、実世界の物体をスクリーンとして映像を投影する技術の導入も、提案技術では有効であると考えられる。

提案するインターフェースにおける同期は、機器などの明示的な指令を介さずに達成される。ユーザが創出する実世界の状態が、ビデオを直接操作する構成となっている。これによって、ユーザに対して、ビデオ操作の感覚を強く意識させず、意図した映像を直感的に取得する連続的なインタラクションを形成することができる。

提案する情報環境を実現するためには、ユーザの動作状態と、保持する映像の間で、類似の状態を認識する技術が鍵となる。これは、身体差や動作獲得レベルによって生じる状態の多様性を定義し、本質的に同一の状態を識別する

ものである。また、既存のビデオを、シンクロナイズドビデオが利用できる形式へ変換することも重要である。さらに、カメラを実世界を捉えるためのツールとして想定した場合、実世界と提示映像との間で生じる視点の違いが影響する変動を補償する必要がある。

このような課題は、センシングされたデータからどのような特徴量を抽出するか、動作に関する事前知識をどのように蓄積するか、の2点に集約される。特に、身体構造を拘束条件とする単眼カメラによるモーションキャプチャや、基底動作に対する多様性を自動抽出する身体動作の一般モデル、画像検索などが関連技術になり得ると考えられる。

### 3. プロトタイプ構成と検証

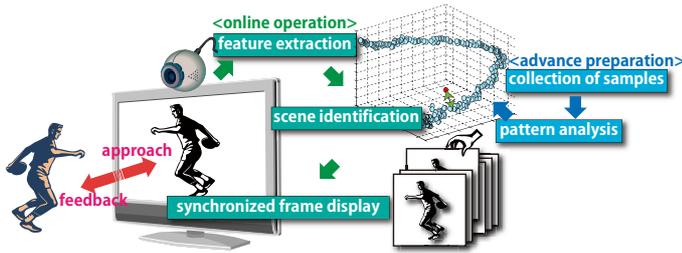


図2: プロトタイプの構成

シンクロナイズドビデオのコンセプトモデルを元に、プロトタイプを開発した。図2に構成図を示す。本構成では、単一のカメラでユーザの状態を捉える。また、初期条件として、映像内とほぼ同じ姿勢で、センシングされるように位置を調整した。今回のプロトタイプでは、映像と身体状態の同一性の認識を簡単化するために、ユーザ、及び教師データとなる映像内の人物には複数のマーカを同じ部位に設置した。同マーカをカメラで捉えた位置がフレーム  $i$  の画像特徴量  $f_i$  となる。特に、画像内の併進ずれを補償するために、身体中央部に設置されたマーカの位置を原点として相対位置  $\tilde{f}_i$  を用いた。

事前準備では、ユーザと教師データのサンプルを集め、画像特徴量の集合  $f$  に対して主成分分析を適用し、射影変換行列  $U$  を取得する。これを用いて、次元削減を行ったものを認識特徴量  $g_i = U\tilde{f}_i$  として利用する。ユーザに提示する標準ビデオとして、サンプルからある動作を選び、同映像と対応する特徴量  $g$  を次のオンラインでの処理のために記憶する。

以上のデータを用いて、提示映像をユーザと同期させる。まず、撮像画像からマーカ位置  $\tilde{f}_t$  を検出し、事前に得た射影変換行列  $U$  を用いて、認識特徴量  $\tilde{g}_t$  へ変換する。次に、この取得された認識特徴量  $\tilde{g}_t$  と最も近い特徴量を標準ビデオのデータ  $g$  から検索し、対応する画像を表示する。この2ステップを画像を撮像する毎に繰り返す。

以上の構成で、動作検証を行った。ビデオコンテンツは卓球のスイングとした。検証では、事前準備とオンライン動作の際に、同じカメラを用いた。カメラの解像度は  $320 \times 240$  である。また、フレームレートはサンプル収集時に 200fps、オンライン処理時に 60fps とした。今回の人物の数は2人である。サンプルとして、それぞれから2~3スイング程度

のサンプルを収集した。主成分分析は、全サンプルに対して適用した。また、標準ビデオには、一方の人物のあるスイングを用いた。マーカは、両手首、両大腿部、へその合計5か所に設置した。画像特徴量  $f_i$  の次元数は8次元、認識特徴量  $g_i$  は3次元とした。この場合の累積寄与率は95%であった。



図3: 動作検証

図3に動作検証の様子を示す。検証の結果、ユーザの動作と同期して、ビデオが再生されることが確認された。シンプルなインタラクションの構成であったため、ユーザは本システムの利用方法を即座に理解し、使用することができた。また、映像を自在に操り、正しいフォームを認識し、提示される映像へ自身を近づける取り組みがスムーズに実施可能であることを定性的に確認した。

### 4. まとめ

本稿では、身体動作と調和するビデオ操作を提供するシンクロナイズドビデオを提案した。これは、新しい動作や作業を映像で配信するメディアを有効利用するためのインターフェースの新しい形態を実現するものである。本稿では、そのコンセプトモデルと技術課題を具体的に述べた。また、プロトタイプを開発し、提案技術が実現可能であることを実証した。今後は、マーカレス画像認識や、状態の同一性を認識する技術の拡張、技術の有効性の評価などが重要であると考えている。また、モノが介在するシーンへの取り組みや、実世界へ直接投影するアプローチも重要な課題となる。これまでに、彫刻作業を支援するものが開発されている [3]。その他、要素技術として、3次元の形状を認識し、リアルタイムに歪みのない映像を物体へ投影するシステムが開発されている [4]。

### 参考文献

- [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.