

## 推薦論文

## 遠隔実習教育のための多視点映像同期再生システム

市村 哲<sup>†</sup> 中村 亮太<sup>††</sup> 伊藤 雅仁<sup>†</sup>  
 宇田 隆哉<sup>†</sup> 田胡 和哉<sup>†</sup> 松下 温<sup>†</sup>

専門学校や企業内研修では、講義よりも実習を中心とした授業が実施される場合が多い。近年、E-ラーニングに対する関心の高まりにともない、実習主体の教育に対しても E-ラーニングを導入したいという要求が高まっている。しかしながら、講師映像と講義スライド資料とを同期配信する従来の E-ラーニングがカバーできるのは講義部分のみであり、実習部分に関してはほとんど支援手段とはなりえないのが現状である。著者らは、従来は集合教育でしかできないと考えられていた実習を含む教育を、リモート環境または独学において可能とすることを旨として研究を実施している。実習を支援するためには、「見たい箇所を多角度から見られるコンテンツ」または「各部の動きが正しく把握しやすいコンテンツ」が望まれる。本論文では、各撮影者がビデオカメラで個別に撮影したビデオ映像中の撮影情報を解析し、複数ビデオ映像から関連のある映像箇所を抽出して同期再生する MVP システムを提案する。通常、個別に撮影した複数ビデオ映像を同期再生するためには、非常な労力を用いてビデオ編集作業をする必要があるが、MVP を使用すればこの作業の大部分を自動化できる。作業効率を測定する実験を行い、MVP システムの有効性を確認した。

## Generating Synchronized Multi-Video Contents for Distance Training

SATOSHI ICHIMURA,<sup>†</sup> RYOTA NAKAMURA,<sup>††</sup> MASAHITO ITO,<sup>†</sup>  
 RYUYA UDA,<sup>†</sup> KAZUYA TAGO<sup>†</sup> and YUTAKA MATSUSHITA<sup>†</sup>

At vocational schools and corporate in-service training facilities, greater emphasis is usually put on giving a practical training than giving a lecture. However, traditional e-learning systems that provide learners with lecturer's audio/voice with synchronized presentation slides can hardly support the practical training. Our study focused on the creation of distance training environment in which people can learn physical operations, e.g., machine operations or instrument operations, in a remote or asynchronous manner. As a part of the study, we developed MultiVideoPlayer (MVP for short), a software playing back multiple video clips synchronously, each of which is spontaneously videotaped by different digital video camcorder. The software can extract the date and time information of video-recording from the digital video clips, and synchronize them automatically. Since it is well-known that synchronizing multiple videotapes needs an enormous amount of time and effort, MVP can greatly contribute to the creation of multi-angle video contents. Through evaluations of the system, we verified the effectiveness of our system.

## 1. はじめに

常時接続とブロードバンドの浸透にともない、職場・学校・公共施設においてインターネットサービスを利用できる環境が整いつつある。このような状況下、E-ラーニング（遠隔学習）が、ブロードバンドを用いたアプリケーションの 1 つとしておおいに期待を集めて

いる。しかしながら、現在効果が実証されている E-ラーニングは、マナー研修、英語学習、座学講義など、従来からテキストとビデオテープによる独学環境が存在したような分野に関するものに限られている。これらの E-ラーニングシステムは、講師映像と講義スライド資料とを同期させてインターネットを介して受講者に配信するというものがほとんどである。

<sup>†</sup> 東京工科大学  
 Tokyo University of Technology

<sup>††</sup> 慶應義塾大学  
 Keio University

本論文の内容は 2003 年 6 月のマルチメディア、分散、協調とモバイル (DICOMO2003) シンポジウム-63 にて報告され、DICOMO2003 プログラム委員会委員長により情報処理学会論文誌への掲載が推薦された論文である。

このような現状をふまえ、著者らは、リモートで行う実習教育を「ディスタンストレーニング」と呼び、そのための支援システムの研究を実施している。従来は集合教育でしかできないと考えられていた「実習」を含む教育を、リモート環境または独学において可能とすることを目的としている。実習を支援するためには、詳細な機器操作映像や、その映像についての解説などを学習者に正確に伝達することが重要である。そのためには、「見たい箇所を多角度から見られるコンテンツ」、「各部の動きが正しく把握しやすいコンテンツ」などが望まれる。

本論文では、ビデオカメラを複数用いて機器操作映像等を撮影し、その多面的ビデオ映像を同期再生するシステム MultiVideoPlayer (以下、MVP) についての提案を行う。MVP は、複数のデジタルビデオカメラ (以下、DV カメラ) で別々に撮影したビデオ映像中の映像撮影情報を解析してそれら複数のビデオ映像の関連性を自動検出する機能を有しており、別々に撮影した複数のビデオ映像から関連のある映像箇所を抽出して同期再生できる。

## 2. 遠隔実習支援

### 2.1 要求分析

たとえば専門学校では、実習を主体とした職業訓練が授業の中で実施されている。このような授業は、講義 30%、実習 70% の割合で構成されるような場合が多く、実習が重視されている<sup>1)</sup>。しかしながら、このような実習中心の授業を対象とした場合、講師映像と講義スライド資料とを同期させて配信する従来の E-ラーニングシステムがカバーできるのは講義部分に関してのみであり、肝心の実習部分に関してはほとんど支援手段とはなりえないのが現状である。

日本工学院専門学校 (著者らが所属する大学と同一学校法人に属する専門学校) の臨床工学専攻科や理学療法学科においては、日進月歩で進化する手術用医療機器や、実際のリハビリテーションに利用される治療器を日常的に用いて学生に実習させている (図 1)。しかしながら、講師映像と講義スライド資料とを同期配信するという従来の E-ラーニングシステムは、以上のような実習主体の教育においてほとんど支援手段とはなりえない。

実習を支援するためには、詳細な機器操作映像や、その映像についての解説などを学習者に正確に伝達することが重要である。そのためには著者らは、単体のビデオ映像を再生できるようにするだけでなく、多面的にビデオ映像を撮影し、それら複数ビデオ映像を

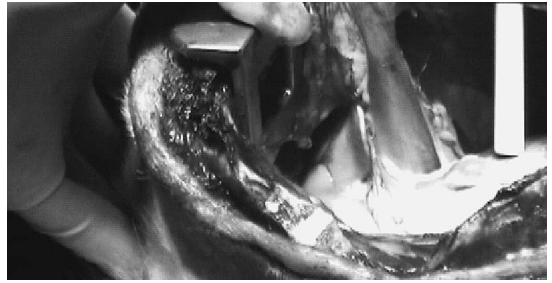


図 1 臨床工学専攻科の演習風景 (日本工学院専門学校)  
Fig. 1 Clinical instruction at a vocational school.

整理して提示できる機能が必要であると考えた。

特に、機器操作風景を撮影するような場合、ある人は望遠レンズで人物を大きく撮影しようとし、ある人は標準レンズで機器全体を撮影しようとするのが普通である。また、撮影者各自は、任意のタイミングで録画を開始または停止するであろう。また、ある人はビデオカメラを三脚に固定して撮影し、ある人はビデオカメラを手持ちで移動しながら撮影する。さらに、ビデオテープの最後まで撮影した場合には、新しいビデオテープに取り替えて撮影を継続するであろう。

すなわち、上記のように、複数人がきわめて自由に撮影したビデオ映像であっても、関連のある映像箇所を抽出して多視点の同期映像を自動作成できるようにすることが重要と考えられる。

## 3. 従来技術の問題点

同じ対象物を同時刻に撮影した映像であっても、複数のビデオカメラで別々に撮影した映像には関連性がなく、再生する際にはそれぞれの映像を独立して再生するしか方法がないのが普通である。同時刻に撮影した複数映像を探し出すためには、複数のテープから早送りや巻き戻しをして探し出す必要があり、非常に手間がかかる。

### 3.1 固定カメラを用いる方式

従来、多視点ビデオ映像を作成するためには、固定カメラを複数台設置し、同時に記録スタートまたは記録終了させるための特別な装置を付加して用いることが一般的である<sup>5)</sup>。しかしながら、ビデオ撮影システムが大規模化してしまい可搬性が損なわれるという問題があった。さらに、手元の拡大映像が撮影できない、機器や作業者の陰になる部分の撮影が困難、作業者の視点の映像を撮影できない等の問題や、システムの組み立て、配線、テープの頭出しなどの撮影準備に膨大な手間がかかるという問題があった。

たとえば、日本工学院専門学校には、大教室に複数

台の固定カメラを設置し、この大教室で行われた授業を映像キャプチャして保存するための講義収録システムが導入されている<sup>1)</sup>。しかしながら、本講義収録システムは大規模であり、かつ、ビデオカメラを定位置に固定する必要があるため、当大教室以外では使用できないのが現状である。たとえば、実習授業の風景を撮影したいというような場合、ほとんどの実習が小教室で行われるという事情があるため、この講義収録システムをそれらの教室に導入することはきわめて難しいという課題をかかえていた。

### 3.2 手持ちカメラを用いる方式

複数の手持ちビデオカメラを使って撮影し、後からビデオ編集で複数ビデオ映像間の同期を合わせることで多視点映像を作成するという方法も考えられるが、実際は、同時刻に撮影した映像を複数ビデオから探すためには、シーンの早送りや巻き戻しを繰り返したり、撮影カットの切れ目を、目で見て判断したりする必要があり、非常に手間がかかる。

また、タイムコード編集可能なノンリニア編集器<sup>6)</sup>を用いた場合においても、撮影を開始するタイミングおよび撮影を終了するタイミングをそのビデオカメラの撮影者が独自に決定するような場合には、ビデオテープのタイムコードと撮影時刻とを一致させることができず同期がとれない。また、複数のカットを連続して撮影する場合、いちいちカットごとにビデオテープを最初に巻き戻すようなことはせず、ビデオテープの続きから撮影することが普通であるため、この点からも、ビデオテープのタイムコードと撮影時刻とを一致させることが難しい。このため従来は、ビデオ撮影するたびに、撮影内容と、撮影日時と、タイムコードとの対応関係を紙に書きとめておき、ビデオテープと一緒にこの紙を管理するというのが普通となっている。このための手間は非常に大きく、多視点映像を作成するうえで大きな課題となっている。

この問題に対処するための方法として、タイムコードを音声信号として生成するタイムコードジェネレータという装置を用い、撮影前にあらかじめこのタイムコードジェネレータが発生する音声信号をすべてのビデオテープに録音しておくことで、複数ビデオテープのタイムコードを統一させるという方法が提案されている<sup>7)</sup>。たとえば、電子音発生器によって、ビデオカメラの音声トラック部分に同期信号を刻印する。

しかしながら、各撮影カットの撮影開始時にすべてのビデオカメラに対してタイムコード音声信号を録音する必要があり、ビデオカメラの数が多き場合には非常に手間がかかるという問題があった。また、タイム

コード音声信号発生装置、および、その音声信号をビデオテープから検出する音声信号検出装置が必要となり、コストが増大するという問題があった。

## 4. システムの機能

MVP は、複数の DV カメラで別々に撮影したビデオ映像中の映像撮影情報を解析してそれら複数のビデオ映像の関連性を自動検出する機能を有している。これは、後述する「同一時刻撮影映像の同期再生機能」によって実現されている。この方法を基本として本システムは、別々に撮影した複数のビデオ映像から関連のある映像箇所を抽出して同期再生する機能を提供している。

特に MVP は、多視点映像コンテンツを撮影する際の制約が非常に少ないのが特徴である。ビデオカメラの現在時刻を標準時刻に一致させておきさえすれば、各撮影者個人が自由なタイミングで撮影開始または終了してもかまわない。また、特殊な撮影装置が不要で、民生用 DV カメラのみを使用して即座に多視点映像撮影ができるという特徴がある。

MVP を用いれば、多視点映像を作成する場合でも、以下のようにビデオ撮影時の制約がきわめて少ない。

1. 民生用 DV カメラのみを使用し外部制御装置が不要。
2. 手持ちカメラを利用でき、撮影位置を固定されない。
3. 各人が自由なタイミングで撮影の開始・終了が可能。
4. ビデオテープの途中から撮影開始が可能。
5. 複数ビデオテープにまたがった撮影が可能。

なお、本システムにおいては、撮影者が用いるビデオカメラは、民生用ミニ DV カセットテープ（以下、DV テープ）を記録媒体として映像と音声を記録するデジタルビデオカメラである。通称、DV テープと呼ばれる記憶媒体は、正確には ISO/IEC 61834 規格<sup>3)</sup>に基づいた記憶媒体である。今までにも、この規格を用いて、DV 映像を IP ネットワーク上でストリーミング転送する DVTS<sup>9)</sup> (Digital Video Transport System) が開発された例がある。

### 4.1 同一時刻撮影映像の自動検出

本方式は、複数のビデオ映像から、同一時刻に撮影された映像を自動検出する技術が基本となっている。具体的には、DV カメラで撮影されているライブ映像、DV テープに記憶されたデジタル映像、または、DV キャプチャされた映像ファイルから、その映像が撮影された日時情報を取得し、それに基づいて関連映像の

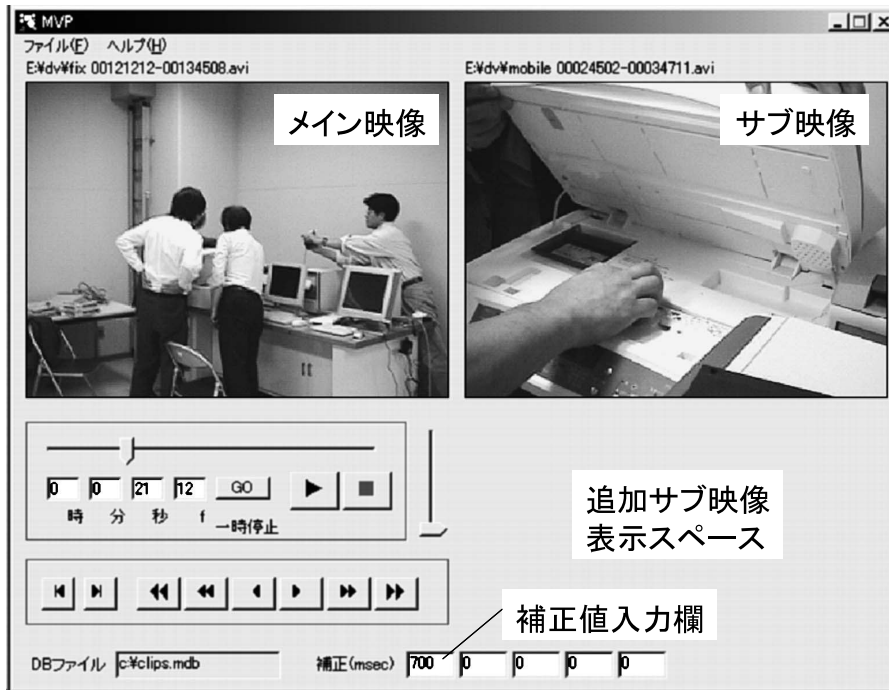


図 2 撮影者 2 名が撮影した映像の同期再生  
Fig. 2 MVP is playing two separate video files synchronously.

解析を行う。

DV システムの規格である ISO/IEC 61834 規格では、撮影日時情報（従来ビデオ編集等で用いられているタイムコードとは異なる情報）をすべての動画フレームのサブコードエリアに対して埋め込むことが規定されており、本システムはこの情報を利用している。この情報は、国内主要メーカーが販売する DV カメラから共通に取得可能である。本システムでは、DV カメラを IEEE1394 規格のケーブルによって PC と接続し、デジタルビデオキャプチャ（以下、DV キャプチャ）して PC の動画ファイルとして保存するようになっている。また、一般にはあまり知られていないが、DV キャプチャして作成された動画ファイルにも、DV テープに記録されている撮影日時情報がコピーされており、すべての動画フレームの中に撮影日時情報が記録されている。本システムは、Windows オペレーティングシステムの DirectShow の仕組みによって、動画ファイルの中から、動画フレームを 1 フレームずつ抽出し、撮影日時情報を取得するようになっている。すなわち、撮影に用いる DV カメラの内部時計が標準時刻に合わせてありさえすれば、DV テープまたは DV キャプチャファイルに記録された映像からその撮影日時を取得し自動的に利用できる。

撮影日時情報を解析するために要する時間は、CPU

Pentium4 1.6 GHz の PC を用いて処理した場合、10 分長の動画ファイルに対して約 1 分である（全フレームを対象に解析を行った場合）。処理速度が比較的速い PC を用いる場合は、DV キャプチャ処理と並行して撮影日時情報を取得することも可能である。

#### 4.2 同一時刻撮影映像の同期再生

図 2 は、MVP のユーザインタフェースを示している。左側のパネルにはメイン映像が、右側のパネルにはメイン映像に同期したサブ映像が適宜表示される。映像を再生する場合、ユーザは、メイン映像として再生すべき動画ファイルを、ファイルダイアログやドラッグアンドドロップ等で指定する。ここで、撮影者 A が撮影した動画ファイル A をメイン映像として指定した場合には、動画再生部の左側のパネルに動画ファイル A の映像が再生表示される。このとき、動画ファイル B および動画ファイル C に、動画ファイル A と同一時刻に撮影された映像が含まれていたとすると、システムが動画ファイル B および動画ファイル C を検出し、サブ映像として右パネルに同時再生する。理論上、同時再生するサブ映像数に上限はないが、現状の実装では、サブ映像を 4 つまで同時再生できる。

このとき MVP は、動画ファイルの再生箇所の撮影時刻が同一となるように、すべての動画ファイルの再生タイミングをシンクロさせて動画再生する。すなわ

ち、メイン映像である動画ファイル A が、撮影時刻  $T_1$  に撮影した映像を再生しているときには、他の動画ファイル B および動画ファイル C も撮影時刻  $T_1$  に撮影した映像を再生する。このとき音声は、動画ファイル A の音声のみ再生し、動画ファイル B および動画ファイル C に記録された音声は再生しないようになっている。なお、撮影に用いた DV カメラの内部時計が多少ずれていた場合に対処するため、その誤差をユーザが補正するための GUI も提供している。通常、DV テープ 1 本に対し 1 度だけ補正を行えばよい。

当然、撮影者 B が撮影した動画ファイル B をメイン映像として指定することも可能である。この場合には、動画再生部の左側のパネルに動画ファイル B の映像が再生表示され、右側のパネルには、動画ファイル A および動画ファイル C 中の、同時刻に撮影された映像部分が同期再生表示される。

なお、本システムでは、DV キャプチャされたビデオ映像を解析してインデックス情報をあらかじめ作成するインデクシングソフトウェア「MVP アナライザ」を提供している。MVP アナライザは、DV キャプチャしたビデオ映像中の撮影日時情報を解析することによって、撮影カット（DV カメラの映像記録開始ボタンが押されてから次に記録停止ボタンが押されるまでの区間の映像）ごとの撮影時刻インデックス（撮影カット OP は  $T_o$  から  $T_p$  までの映像、撮影カット QR は  $T_q$  から  $T_r$  までの映像、というようなインデックス情報）を作成する。これによって、MVP システムは、映像再生時に、同期場面を他の動画ファイルから高速に検索できる。

また、付加機能として、メイン映像が終了した場合には、それに連動して再生しているすべてのサブ映像の再生を終了させる機能を有している。加えて、メイン映像を一時停止、スロー再生、コマ送り再生、または、早送り再生した際に、すべてのサブ映像を同じモードで再生できるようになっている。

## 5. システム評価

本論文では、複数ビデオ映像を同期させるための手間を削減することによって、多視点映像コンテンツ作成作業の効率化を目指す MVP システムを提案した。この指針に基づいた評価を行うため、著者らは、複数ビデオ映像の同期作業が、本システムによってどれだけ効率化されるかについて測定実験を行った。

### 5.1 方法

実験に用いた映像は、機器操作映像（プリンタ取扱い映像）であり、主な被写体は小型プリンタとその取

扱いを学習する学習者数名である。2 台のビデオカメラによって並行撮影されたものであり、1 台（ビデオカメラ A）は作業風景全体を主にとらえ、ほかの 1 台（ビデオカメラ B）は作業者の手元をアップでとらえたものである。ビデオカメラ A で撮影した映像 A は長さ 45 秒、カット数 3 であり、ビデオカメラ B で撮影した映像 B は長さ 30 秒、カット数 4 である。いっせいに記録開始された映像ではないため、DV テープのタイムコードは参照できない状況である。

上記、映像 A と映像 B を題材とし、両映像を同期させるために要する処理時間を、従来のノンリニア編集ソフトを用いた場合と MVP アナライザを用いた場合とで比較する実験を行った。MVP アナライザを用いた場合の処理時間は、複数のビデオ映像から撮影日時情報を抽出し、対応する時間関係をすべてファイルに出力し終えるまでの時間であり、コンピュータにより全自動で行われる処理の時間である。Pentium4 1.6 GHz の市販デスクトップ PC を用いて処理した。

一方、ノンリニア編集ソフトを用いた場合の処理時間は、ノンリニア編集ソフト使用経験のある被験者 5 名（学部学生 3 名、大学院学生 1 名、大学教官 1 名）の手作業により行われる映像同期作業時間である。被験者らには、複数台のビデオカメラで撮影したビデオ映像（DV キャプチャファイル）を素材映像として、ノンリニア編集ソフトを用いてタイムチャート上で同期する作業が課せられた。実験に用いたノンリニア編集ソフトは、キーボード操作またはマウス操作により、早送り、巻き戻し、コマ送り再生、コマ送り逆再生がスピーディーにできる機能を備えており、通常の映像再生ソフトより映像の頭出しがはるかにしやすいソフト<sup>4)</sup>である。この実験の際、被験者らには、何を手掛かりに映像を同期しようとしているか、また、期待していたとおり同期できたか等について、つねに声に出しながら作業をしてもらうようにした。

また、本システムの場合、複数映像を並べて表示することが目的であるため、わずかでも再生タイミングがずれていると視覚的に違和感を与えてしまうと予測される。このことを確認するため、本実験を開始する前に、被験者らに、どれだけ同期タイミングがずれると違和感を持つかサンプル映像を見せて聞き取り調査を行った。この結果、テニスストローク、ゴルフスイングなどのスポーツ映像を題材とした場合、わずか 5 コマ程度（約 165 ms）ずれるだけで、ほとんどの人が映像のズレを違和感としてとらえることが分かった。そこで、タイミングのズレが重要な指標になると考え、本実験においては、平均誤りコマ数（1 コマ約 33 ms）

表1 ノンリニア編集ソフト使用時の着目点

Table 1 Used clues for synchronizing multiple videos.

## (A) 対象物別

人に着目	人の頭が最も沈み込んだ瞬間
	腕が沈み込んだ瞬間
	指が機械から離れる瞬間
物に着目	「どう、これ」という音声の始まり
	紙の折れが最大となった瞬間
	プラテンが閉まった瞬間
	バネが立ち上がった時の音の始まり

## (B) メディア別

画像に着目	人の頭が最も沈み込んだ瞬間
	腕が沈み込んだ瞬間
	指が機械から離れる瞬間
	紙の折れが最大となった瞬間
音声に着目	「どう、これ」という音声の始まり
	バネが立ち上がった時の音の始まり

もあわせて測定することとした。

なお、本実験は、自動処理速度と手動処理速度とを比較する内容であるため両者間に有意な差が生じることとは自明である。本実験の趣旨は、本システムの導入によって映像編集作業がどれだけ効率化されるかを測定することにある。

## 5.2 結果

ノンリニア編集ソフトを用いた作業中に被験者らが発声した内容を表1に示す。表1(A)は、着目した対象物の違いによって分類したものである。「人の頭が最も沈み込んだ瞬間」、「紙の折れが最大となった瞬間」など、見る角度によっては違うコマが選定されかねない不安定な指標が選定されていることが分かる。表1(B)は、着目したメディアの違いによって分類したものである。画像情報を手がかりとして検索する場合と、音声情報を手がかりとして検索する場合とが認められた。実際は、映像Aに現れた物体が期待したように映像Bに現れなかったり、映像Aで聞き取れた音声映像Bで聞き取れなかったりして、同期点を探す作業は困難を極めた。

表2(A)は、映像中の1カットの同期をとるための平均所要時間、および、1カットあたりの平均誤りコマ数の測定結果である。所要時間の比較、誤りコマ数の比較の両方について、ノンリニア編集ソフトを用いた場合に比べ、MVPアナライザを用いた場合の方が格段に好ましい値となっており、所要時間が短く誤りが少ないことが分かる。なお、表2(B)は、ノンリニア編集ソフトを用いた実験結果の被験者ごとの内訳である。

所要時間の比較に関し、ノンリニア編集ソフトを用

表2 映像同期作業の平均所要時間と誤りコマ数

Table 2 Errors and necessary time for synchronizing multiple videos.

## (A) 全被験者平均(1カット同期作業平均)

	平均所要時間 (秒)	誤差平均 (コマ)
MVPアナライザ (全自動)	7.0	0.0
ノンリニア編集 (手作業)	587.2	79.7

## (B) 被験者別の内訳(1カット同期作業平均)

	所要時間 (秒)	誤差 (コマ)
被験者A	530	2.0
被験者B	443	25.7
被験者C	515	2.7
被験者D	380	367.3
被験者E	1068	0.7
平均	587.2	79.7
標準偏差 (自由度4)	275.4	161.1

いた場合およびMVPアナライザを用いた場合は、それぞれ平均587.2秒および7.0秒であり、10分弱の差があることが分かった。これは、カット数4の映像を編集した場合の差であり、本システムの導入によって、映像同期作業時間が1カットあたり約2分半短縮できたことを示している。

一方誤りコマ数の比較に関し、被験者Dが突出して値が大きく(367.3コマ)、この値が被験者全体の平均誤りコマ数を大きな値(79.7コマ)にしていることが分かった。仮にこの値を除外すると、被験者全体の平均誤りコマ数は7.8コマ(約0.26秒)となる。人は5コマ程度ずれていると違和感を持つという前述の実験結果と照らし合わせると、被験者Dの誤差は異常な値である。そこで、被験者Dにヒアリングを行った。その結果「どう、これ」という音声が発せられた箇所映像中に2カ所あり、被験者Dは、映像Aでは1回目の発声箇所を同期点とし、映像Bでは2回目の発声箇所を同期点として取り違えたことが分かった。被験者5名という少人数による実験、かつ、40秒程度の短い映像を対象とした実験という状況でもこのような誤りが発生したことから推測すると、手作業の場合には今回の実験結果のような人為的誤りが発生する可能性が高い。この問題は、手作業による映像同期処理の本質的な問題であると考えられる。

以上の測定実験によって、MVPアナライザが、作業時間の短縮、および、同期誤りの防止に貢献することが分かった。ただし、上記実験では、MVPアナライザを用いた場合において複数DVカメラの内部時計のズレを補正する作業時間を考慮していなかった。MVPを用いる場合、この補正作業がDVテープ1本に対し通常1度だけ必要となる。そこで前記被験者5

名に、2台のDVカメラの内部時計の差を補正する下記タスクを課し、所要時間を計測する実験を追加して行った。

1. 両DVカメラを起動して、メニューから「日時あわせ」項目を選択する。これにより、液晶パネルにDVカメラの内部時計が秒単位で表示される。
2. 両DVカメラの液晶パネルを見比べ、進んでいる方の時計がある時刻Tになる時点でストップウォッチをスタートし、遅れている方の時計が時刻Tになった時点でストップウォッチをストップする。
3. ストップウォッチで計測した時間差を、MVPの撮影日時補正値入力欄に入力する。

以上の作業を被験者1名あたり2回ずつ実施した結果(サンプル数10)、所要時間の分布は平均56.1秒(標準偏差=13.1)となった(DVカメラの日時あわせ項目の表示方法については全被験者があらかじめ知っている状態で実験を実施)。表2(A)中のMVPアナライザ処理時間7.0秒にこの56.1秒を加算したとしても63.1秒であり、手作業による作業時間587.2秒よりきわめて短いことが分かる。加えて、この63.1秒の値は、カット数が増えた場合でも増加しない値である。なお、この実験では、内部時計に、9.40秒の差がある2台のDVカメラを用いて行った。ズレを±5コマ(±165ms)以内におさめるためには、ズレの計測値が9.235秒以上9.565秒以下となる必要がある。実験の結果、被験者がストップウォッチで計測した内部時計のズレの分布は平均9.35秒(標準偏差=0.0785)となり、95%以上の確率(正規分布と仮定した場合)で誤差±5コマ以内に内部時計を補正できた。

加えて著者らは、実用的な演習用コンテンツを作成する場合に、どの程度のカットが撮影され、また、どの程度のカットが同期映像として用いられるかを実験によって調べた。この実験では、圧着工具を用いてLANケーブルを作成する方法(リールからケーブルを切り分ける作業から、ケーブルの端にRJ45コネクタを圧着する作業まで)を映像化することを題材とし、それを2名のカメラ撮影者(AとB)に撮影させた。撮影者A(固定位置から撮影)には、作業工程の中で重要と思うシーンを漏らさずに撮影するように指示し、一方撮影者B(手持ち撮影)には、撮影者Aが撮影できない角度からの映像や、手元ズーム映像、工具ズーム映像を適宜撮影するように指示した。なお、両撮影者には、自由なタイミングでカット撮影の開始・終了を行うように指示し、また、LANケーブル作成者には、作業の要点を声に出しながら作業するように指示

した。

LANケーブル作成作業は約6分の作業であった。撮影者Aが撮影した映像は約2分であり、この中に13個のカットが含まれていた。一方撮影者Bが撮影した映像は約1分半であり、この中に11個のカットが含まれていた(この11個のカットには、撮影はされたが不要と判断されたカット5個が含まれる)。前述した実験結果に基づくと、1カットを手作業で映像同期させるために必要な時間は平均2分半であることから、11個のカットを同期させたとすると約27分半かかると予測される。また、不要と判断されたカット5個を取り除いた6個のカットのみを同期させたとしても約15分かかると予測できる。6分間という短い撮影時間に対し、映像同期にかかった作業時間は十分に大きく、MVPによってこの作業時間をなくすことができることは有意義であると思われる。

なお評価実験では、「両映像のカット数、映像長があらかじめ分かっていた」「両映像に同じ物体が映っていることがあらかじめ分かっていた」「両映像に適度な動きがあり、音声も含まれていた」「比較する映像が2つに限定されていた」等の制約が与えられており、これらの制約がない実際の映像同期作業では、手作業によるタイミング合わせ作業はさらに難しいと予測される。

## 6. まとめ

本論文では、だれでも簡単に多視点映像を同期再生できるMVPシステムを提案し、評価実験を通じてシステムの有効性を確認した。

撮影者各人が自由なタイミングで撮影を開始・終了でき、撮影や編集のための手間がきわめて少ないことが本システムの特徴である。今後は、実際の教育の現場にシステムを導入し、多視点映像コンテンツの実利用者を対象とした評価を行い、そのフィードバックに基づいてシステムを改良してゆきたい。専門のビデオ編集業者に頼らなくても、教官ら自らが講義用の多視点ビデオ映像を素早く手軽に作成できるようになることが期待できる。

また、多視点でキャプチャした映像にインタラクティブ性を与えるための方法、および、多視点映像教材を遠隔で共有する場合のコミュニケーション技術に関して研究を実施したいと考えている。さらに、本論文で提案した方法は、ノンリニアビデオ編集における複数映像ソースの同期化処理に応用できるため、複数ビデオから1つの映像を作成する工程でも利用できるようにしていきたい。

## 参 考 文 献

- 1) 市村, 宇田, 伊藤, 田胡, 松下: 遠隔実習教育のための多視点映像コンテンツ作成システム, 情報処理学会 DICOMO 2003 論文集, pp.249-252 (2003).
- 2) ジャレミーヴィンヤード: 映画技法完全レファレンス, フィルムアート社.
- 3) <http://www.iec.ch/>, ISO/IEC 61834: Recording — Helical-scan digital video cassette recording system using 6, 35 mm magnetic tape for consumer use (525-60, 625-50, 1125-60 and 1250-50 systems).
- 4) <http://www.area61.net/>
- 5) 石塚, 亀田, 美濃: 講義の自動撮影系における音声・映像インデキシング, 電子情報通信学会技術研究報告, PRMU99-258, pp.91-98 (2001).
- 6) <http://www.avid.co.jp/products/post/express/>
- 7) <http://www.xecoo.co.jp/muve2000/muveNL.htm>
- 8) 市村, 宇田, 伊藤, 田胡, 松下: 散在する多視点ビデオ映像を用いたユビキタスビデオ網構想, 情報処理学会ユビキタスコンピューティングシステム研究会報告, UBI-1-4, pp.23-30 (2003).
- 9) <http://www.sfc.wide.ad.jp/DVTS/>

(平成 15 年 10 月 16 日受付)

(平成 16 年 9 月 3 日採録)

## 推 薦 文

本論文は, 遠隔実習向けの映像コンテンツを作成する際に, 複数のビデオカメラで撮影した映像を時刻で同期させる方式に関するものである. 単なる「講義」にとどまらず, 機器操作など「実習」に主眼を置いた遠隔教育を目的としたコンテンツ作成にまで踏み込んだ内容であり, テーマ的に重要で本手法の適用価値は大きい. 多角的な視点からの情報を必要とするという対象コンテンツの特性に対し, 比較的簡易ながら効果が大きい手法を提案しており, 有用性も高いことから技術論文として評価できる.

(DICOMO2003 プログラム委員会委員長 高橋 修)



市村 哲 (正会員)

1966 年生. 1989 年慶應義塾大学理工学部計測工学科卒業. 1994 年同大学大学院理工学研究科博士後期課程修了. 博士 (工学). 同年富士ゼロックス (株) 入社. 1997 年 ~ 1999 年富士ゼロックスパロアルト研究所駐在. 2002 年 4 月より東京工科大学助教授. グループウェア, ネットワークサービス, 生体情報活用等の研究に従事. “IT TEXT 基礎 Web 技術”, “IT TEXT 応用 Web 技術” (オーム社). DICOMO 2003 優秀論文賞受賞. ACM 会員.



中村 亮太 (学生会員)

1979 年生. 2002 年東京工科大学工学部情報通信工学科卒業. 2004 年同大学大学院工学研究科博士前期課程修了. 現在, 慶應義塾大学大学院理工学研究科開放環境科学専攻後期博士課程在学中. 無線通信工学, ヒューマンインタフェースに興味を持つ.



伊藤 雅仁 (正会員)

1998 年慶應義塾大学理工学部計測工学科卒業. 2000 年同大学大学院理工学研究科修士課程修了. 2003 年同大学院理工学研究科後期博士課程修了. 博士 (工学). 現在, 東京工科大学コンピュータサイエンス学部専任講師. モバイル・コンピューティング, 情報家電, デジタル放送の研究に従事. 2000 年情報処理学会高度交通システム研究会優秀研究報告賞受賞. IEEE, 電子情報通信学会 会員.



宇田 隆哉 (正会員)

1998 年慶應義塾大学理工学部計測工学科卒業. 2000 年同大学大学院理工学研究科計測工学専攻前期博士課程修了. 2002 年同大学院理工学研究科開放環境科学専攻後期博士課程修了. 2003 年 4 月より東京工科大学コンピュータサイエンス学部専任講師. 博士 (工学). ネットワークセキュリティの研究に従事. 2002 年 IFIP/SEC2002 Best Student Paper Award 受賞. 電子情報通信学会 会員.





田胡 和哉 (正会員)

1986年筑波大学大学院工学研究科博士課程修了。工学博士。筑波大学電子情報工学系助手、東京大学工学部助手、日本IBM東京基礎研究所を経て2003年より東京工科大学コンピュータサイエンス学部助教授。オペレーティングシステムの構成方式に興味を持つ。1984年情報処理学会論文賞受賞。ACM会員。



松下 温 (フェロー)

1963年慶應義塾大学工学部電気工学科卒業。1968年イリノイ大学大学院コンピュータサイエンス専攻修了。1989年より2002年3月まで慶應義塾大学理工学部教授。2002年4月より東京工科大学コンピュータサイエンス学部長、教授および慶應義塾大学理工学部客員教授。マルチメディア通信、コンピュータネットワーク、グループウェア等の研究に従事。情報処理学会理事、同学会副会長、マルチメディア通信と分散処理研究会委員長、グループウェア研究会委員長、電子情報通信学会、情報ネットワーク研究会委員長、MIS研究会委員長、バーチャルリアリティ学会サイバースペースと仮想都市研究会委員長等を歴任、現在、情報処理学会ITS研究会委員長、郵政省、通産省、建設省、農水省、都市基盤整備公団、行政情報システム研究所等の委員長、座長、委員を多数歴任。現在、最高裁判所専門委員。『やさしいLANの知識』(オーム社)、『201x年の世界』(共立出版)等著書多数、1993年情報処理学会ベストオーサ賞、1995年および2000年情報処理学会論文賞、2000年10月20日情報処理学会40周年記念90年代学会誌論文賞、2000年10月2日電子情報通信学会フェロー、2000年10月VR学会サイバースペース研究賞、2001年5月情報処理学会功績賞、2002年3月情報処理学会フェロー。電子情報通信学会、人工知能学会、ファジイ学会、IEEE、ACM各会員。