

## 同じイベントを撮影した人たちが映像を共有できる動画編集 Web システム

○<sup>1</sup>市村哲; <sup>2</sup>中村亮太; <sup>1</sup>井上亮文; <sup>1</sup>松下温  
<sup>1</sup>東京工科大学; <sup>2</sup>慶応義塾大学

MediaBlocks は、同じイベントに参加した複数撮影者の映像をサーバに集め、自動時間同期処理を施して整理し、インターネット経由で映像編集できるようにする Web システムである。同じイベントを撮影した人たちが映像素材を交換しあうことが容易にできるようになり、自分が撮影できなかったシーンを取り込むことや、プロフェッショナルの映像技法にそった多彩なビデオ編集が可能となる。システムの設計、実装、評価について述べる。

Web-Based Video Editing System Synchronizing Multi-users' Video Clips Shot in the Same Event.

<sup>1</sup>Satoshi Ichimura; <sup>2</sup>Ryota Nakamura; <sup>1</sup>Akifumi Inoue; <sup>1</sup>Yutaka Matsushita  
<sup>1</sup>Tokyo University of Technology; <sup>2</sup>Keio University

We developed MediaBlocks, a web-based video editing system for sharing personal video clips. In the system, video sources are collected from multi-users' camcorders via the Internet, shared among the users, and edited through each user's web browser. The system allows users to use other users' video clips, each of which is spontaneously videotaped by different digital video camcorder, and also enables them to create home movies following the professionals' important editing rules.

### 1. はじめに

デジタルビデオデータを PC に取り込み、PC 上で動画編集を行うことが一般的となりつつある。平成 16 年の主要耐久消費財普及率調査によると、民生用デジタルビデオカメラ（以下、DV カメラ）の普及率は 42.0% に達している<sup>1</sup>。このように普及してきた動画編集環境であるが、これらの技術を用いていざ一般の人がビデオ編集をしてみてもなかなか満足できるような映像にならず、単調な映像の連続に感じられることが多い。

主たる原因として、一般撮影者の作成する映像は、単一角度から撮影した映像がほとんどであり、また、ビデオ編集に使う映像素材も自分自身が撮影した映像のみを用いていることが挙げられる。これに対し、普段見慣れているテレビの映像は、複数のカメラマンが撮影した複数映像を基に編集されており、それら複数の映像をテンポ良く切り替えることで視聴者を飽きさせないようにしている。

本論文では、同じイベントに参加した複数撮影者の映像をサーバに集め、インターネット経由で映像編集をできるようにするシステム「MediaBlocks」<sup>2</sup>を提案する。従来の一般撮影では、運動会のように

多数の人が同時に同じイベントを撮影している状況でも、撮影者同士が協力しあうことはほとんどなかった。一方、本システムを利用すれば、同じイベントを撮影した人たちがインターネットを介して映像素材を容易に交換しあえるようになる。例えば、クラスメートの親同士が各自撮影した運動会撮影映像をサーバに持ち寄れば、自分ではうまく撮影できなかった我が子の活躍の瞬間を、偶然に集められた他の人の映像の中から見つけて自分の映像に取り込むことが可能となる。

MediaBlocks は、複数の撮影者が撮影した映像をサーバに集め、自動時間同期処理を施して整理し、Web ブラウザを利用してインターネット経由で映像編集できるようにするシステムである。本論文では、システムの設計、実装、評価について述べる。

### 2. 背景

#### 2.1. 映像技法

一般撮影者がビデオ編集する際は、同じ角度から撮影した映像のみを利用する場合がほとんどである。このため、個人が撮影した映像は単視点となり、編集に用いることができる映像素材も、自分が撮影し

た映像のみとなる。少ない素材から映像編集を試みても素材選択の余地は無く、単視点映像の短縮にしかならないのが実情である。

一方、テレビ放映されるような映像は、複数のカメラによって撮り貯めた多くの映像素材の中から、適宜最も良い素材を選別して取り込むようにして編集される。プロフェッショナルの映像撮影においては、「寄りのカメラ」、「引きのカメラ」、「セカンドユニット(例えば、運動会ビデオであれば、校舎、アーチ、校庭全景等のコンテキスト映像)」を併用することが基本であると言われている<sup>3</sup>。飽きのこない映像にするためには、メインとなる映像だけではないことが基本であり、セカンドユニットを撮影しておいて、これをメイン映像に混ぜるという手法が用いられる。また、「同じ視点の似通ったサイズの映像をつながない」という映像技法上の鉄則がある<sup>4</sup>。似通った映像をつなげると、映像が一瞬とんだように見えてしまい、視聴者が時間の不連続性を感じてしまうためである。

映像技法に従って映像コンテンツを作ることが重要であることは、一般撮影者が作成した映像とプロフェッショナルが作成した映像とを見比べると明らかである。しかしながら、一台のカメラによる撮影では、セカンドユニット等の映像素材を用意することができず、一般撮影者が行うビデオ編集は、まさに、同じ視点の似通ったサイズの映像をつなぐ作業にならざるを得ないのが現状である。

MediaBlocks の狙いは、複数撮影者が撮影した様々な撮影位置からの映像を共有し、誰もがお互いの映像を利用して編集作業できるようにすることにある。著者らは実際の多数の運動会映像を観察し、素人の映像を寄せ集めた場合であっても、それぞれに角度、画角、構図、または、雰囲気の違いのある映像を集めることができ、それらをつなぐことで躍動感のある映像に仕上げることができる可能性が高いと判断した。

## 2.2. 多視点映像技術

前述の通り、複数ビデオカメラの映像を用いて映像編集することが重要であることは明らかであるが、一般撮影者がこれを行うためには非常な努力を必要としてきた。通常、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、独立した映像として扱うしか方法がないのが普通である。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。異なる 2 つの映像を同期する場合、両方の映像に共通に表れる物体や人物の動きを一致させるようにして映像を同期するが、実際は、一方の映像に現れた物体が期待したようにもう一方の映像に現れなかったり、一方の映像で聞き取れた音声がもう一方の映像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ることが普通である。

過去において著者らは、複数ビデオカメラで個別に撮影したビデオ映像から、時間同期した多視点映像コンテンツを自動作成するシステム「Multi Video Player」(以下、MVP)の開発を行った<sup>5</sup>。MVP は、機器操作説明のための多視点ビデオ教材を作る目的で設計されており、映像再生時には、時間同期した複数の映像がマルチ画面で再生されるようになっている。MVP では、撮影映像を時間同期させるために、民生用 DV カメラの規格(ISO/IEC 61834)<sup>6</sup>に規定された撮影メタ情報を利用している。この規格では、DV カメラは、DV テープ上のすべての動画フレームに、撮影日時情報を記録することが規定されている。さらに、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の全フレームにも、DV テープ上の撮影日時情報がコピーして保存される。MVP は、DV キャプチャによって作成された動画ファイルの中から動画フレームを抽出し、それぞれのフレームの撮影日時情報を取得する機能を備えている。このため、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、撮影日時を正確に同期させて複数映像を同期再生できるようになっている。

MediaBlocks の場合は、多視点映像を編集して、最終的に通常の 1 画面のビデオ映像に仕上げることが目的であり、MVP とはその目的や用い方が異なる。しかしながら、複数撮影者の映像の時間同期処理のために MVP の多視点映像同期機能を用いることで、付加装置を追加することなく、また、撮影準備として特別な操作をすることなく時間同期処理を実行することができるようになっている。MVP を核として用いた時の利点は以下のようなものである。

1. 各人が自由なタイミングで、撮影の開始・終了が可能

2. 民生用 DV カメラのみを使用し、外部制御装置が不要
3. 手持ちカメラを利用でき撮影位置を固定されない
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

### 3. 実装

MediaBlocks システムは、ビデオ収集、ビデオ共有、および、ビデオ編集の 3 つの部分から構成されている。以下に、それぞれについて詳細に述べる。

#### 3.1. ビデオ収集機能

ビデオ収集は、原則的に、各撮影者が自分自身の PC 上で実行する処理である。ビデオ収集機能には、デジタルビデオテープ（以下、DV テープ）上のビデオ映像をビデオキャプチャ（以下、DV キャプチャ）によって動画ファイル（以下、DV ファイル）に変換する機能、DV カメラまたは DV ファイルから撮影日時情報を検出する機能、および、高画質の DV キャプチャ映像（DV ファイル）を低画質ストリーミング映像に変換する機能が含まれている。

本システムで利用可能なビデオカメラは、ISO/IEC61834 規格に準拠した一般的な民生用 DV カメラである。著者らは、DV テープに記録されたビデオ映像を PC に DV キャプチャしながら撮影日時情報を取得するソフトウェア「DV ビデオキャプチャ」を開発した。このソフトウェアは、PC 画面上の操作パネルを操作することで、PC に IEEE1384 接続された DV カメラを制御できる Windows アプリケーションであるが、動画フレームを 1 フレームずつ抽出し、各フレームの撮影日時をフレーム単位の精度(約 33ms)で検出する機能を兼ね備えている。検出される撮影日時情報は、撮影カット（DV カメラの録画開始ボタンが押されてから次に録画停止ボタンが押されるまでの区間の映像）毎の撮影時刻インデックス情報であり、現在の実装では CSV ファイル形式で出力されるようになっている。

なお、後述するビデオ編集の過程では、編集途中映像のプレビューを何度も繰り返して行うことが必要であるが、インターネットを介して映像編集するという本システムの性質上、ネットワーク帯域を考慮する必要がある。DV キャプチャして作成される DV ファイルは、10 分映像で約 2GB 強のファイルサイズになり、高画質であるがファイルサイズが大きいという性質がある。そこで、本システムでは、

DV ビデオキャプチャツールが DV ファイルから RealMedia 形式<sup>8</sup>のストリーミングビデオを生成し、ビデオ編集の過程でこの低画質映像をプレビューできるようになっている。10 分映像の 2GB 強の DV ファイルを、300Kbps の RealMedia 圧縮形式（ブロードバンド用ストリーミングビデオ程度の画質）に変換した場合、ファイルサイズを約 20MB にまで削減できる。さらに、ストリーミングビデオ形式であるため、ダウンロードの完了を待たずに映像を再生開始することができるという特徴を有しており、この点からもビデオ編集過程でのプレビュー表示のために望ましいと考えた。

そして、DV ファイルと、撮影日時情報(CSV ファイル)と、RealMedia ファイルとは、標準的な FTP 転送によって MediaBlocks サーバ（MediaBlocks ソフトウェアがインストールされた Web サーバ）にアップロードされる。

#### 3.2. ビデオ共有機能とビデオ編集機能

ビデオ共有機能とビデオ編集機能は、MediaBlocks サーバ（本実装の場合、Tomcat サーバ上の Java サーブレット）が提供する機能である。ビデオ共有機能は、映像と撮影日時情報とを対応付けてデータベースで管理する機能である。一方、ビデオデータ編集機能には、低画質映像をプレビューしながらインターネットを介してビデオ編集操作できる機能と、ビデオ編集結果をスクリプト言語（以下、編集スクリプト）に出力し、最終的に作成された編集スクリプトに基づいてサーバ上の DV ファイルを自動編集する機能とが含まれる。

撮影者から DV ファイルと、撮影日時情報と、RealMedia ファイルとがアップロードされると、MediaBlocks サーバは、各撮影カットの撮影日時情報をデータベースに登録する。この時、先にデータベースに登録されている他の撮影カットの撮影日時情報との比較が行われ、同時刻に撮影されたカットが検出された場合にはその対応関係が記録される。

ユーザが Web ブラウザ上でビデオ編集操作を行うと、MediaBlocks サーバは、その編集操作を解釈して最も近い表現の SMIL（Synchronized Multimedia Integration Language）<sup>8</sup>スクリプトを生成するようになっている。SMIL は、Web 上でマルチメディアプレゼンテーションを実現するための仕様のひとつであり、ビデオ編集操作を XML 形式で記述すれば、その記述に沿ってビデオ

編集を施したような映像を再生できることが特徴である。実際にビデオデータを切り貼りすることなく、ビデオ編集結果をプレビューすることができる仕組みであり、MediaBlocks においては、ビデオ編集の過程でこの仕組みを利用している。SMIL2.0をサポートしているRealOne 動画プレーヤー<sup>8</sup>をHTML埋め込みすることによって、編集途中の映像をWebブラウザ上でプレビュー再生できるようにになっている。

編集途中のSMIL スクリプトは、映像素材クリップとしてMediaBlocksのデータベースに登録される。ひとつの映像素材クリップは、複数の撮影カットの組み合わせであることもあれば、単一撮影カットの部分映像である場合もある。いずれの場合にでも、ユーザは自分が編集して作成した映像クリップを他のユーザに公開することができ(他のユーザに公開したブロックを「公開ブロック」と呼ぶ)、他のユーザは自分の編集の映像素材としてこの公開ブロックを使用することができる。各クリップにはテキストメモを付与できるようになっており、このメモを手がかりにクリップ検索したり、一覧機能によってスナップショット画像と共に閲覧したりできる。

本システムには、MediaBlocksに特徴的な映像編集画面として「パタフライビュー」(図1)が備わっている。パタフライビューは、ユーザが現在編集中のメインの映像クリップに対して、同時刻に撮影された他の映像クリップ(他のビデオカメラの映像)が存在することをユーザに知らせる機能を提供している。ここで、ある映像クリップをメイン映像として編集作業を進めていると仮定する。メイン映像は、パタフライビューの中央部分に表示されるRealMedia映像であるが、ユーザは、メイン映像のプレビュー再生を繰り返しながら、このメイン映像の中から必要な映像区間を選択する。この時システムは、ユーザが選択した映像区間の撮影開始時刻をTs、撮影終了開始時刻をTeとすると、ユーザが映像区間を変更する度に、パタフライビュー左側には、時刻Tsに撮影された他の映像クリップの画像(ビデオ映像から動的生成されたTs時点の静止画スナップショット画像)一覧が表示され、パタフライビュー右側には、時刻Teに撮影された他の映像クリップの画像(ビデオ映像から動的生成されたTe時点の静止画スナップショット画像)一覧が表示される。この機能により、前後につながる他の撮影者の映像が即座にわかるため、他視点の映像の取り込み作業が効率化される。そしてユーザが、パタ

フライビューの左側または右側の映像クリップを選択すると、その選択された映像クリップが今度はメイン映像となった新たなパタフライビューが生成され、その上で同様な映像区間選択作業を継続できるようにになっている。

編集が完了し「ファイルに出力」ボタンを押すと、MediaBlocksサーバは、SMIL形式で記述された編集スクリプトに基づいて、オリジナル映像(DVファイル)を自動編集する。本処理はサーバ負荷が高い作業であるため、現在の実装では、Webサーバとは別に設置されたアプリケーションサーバ上で実行されるようになっており、自動編集機能は、DirectShow 編集サービス(DES)<sup>7</sup>を利用して実装されている。

#### 4. 評価実験

幼稚園の運動会を実際にビデオ撮影し、MediaBlocksを用いてビデオ編集する実験を行った。本実験では、クラスメートの親同士が各自撮影した撮影映像をサーバに収集して共有するという状況を想定しており、撮影対象園児の家族や親族が撮影した映像と、そのクラスメート園児の親が撮影した映像とを収集したいと考えた。そこで、家族や親族が撮影した映像を用意するために、撮影対象園児の母親(主婦)と同園児の叔父(大学生)に撮影を依頼した。両名には撮影方法等に関する指示は一切与えず、全く自由に撮影するように依頼した。

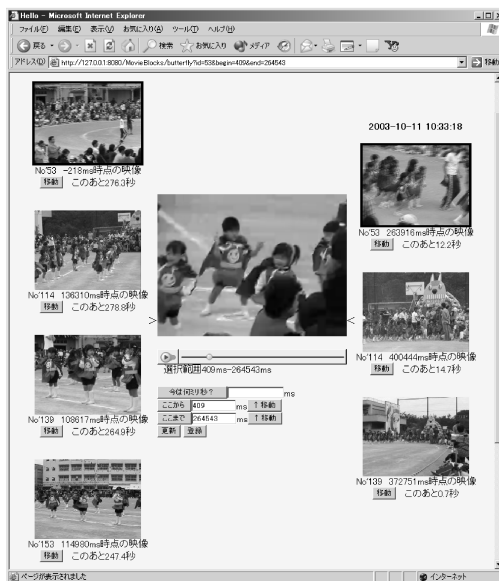


図1 パタフライビュー

一方、クラスメート園児の親の撮影映像を用意するために、本実験では、クラスメート園児の親を模擬する撮影者 A (大学生) と撮影者 B (大学院生) とを用意した。両撮影者には「クラスメート園児の親になったつもりで、クラスメート園児を中心に撮影するように」という指示だけを与え、撮影内容や撮影方法に関する細かい指示は与えないようにした。ただし、実験を効率的に進めるため、撮影者 A と B には、撮影対象園児の近くに存在する園児の親になったつもりで撮影するように指示した。このようにして、模擬的ではあるが、極力自由な条件に近い状態で撮影が行われるように実験条件を設定し、園児らが、徒競走、玉入れ、ダンス、入場行進の様子をハンディタイプの DV カメラを用いて収録した。

MediaBlocks を用いて作成した多視点映像の有効性について検証した。市販ビデオ編集ソフトを用いて撮影者 1 人の映像だけを編集した映像 A と、MediaBlocks を用いて複数人の映像を編集した映像 B とを被験者に見せ、作成されたコンテンツの質をアンケート調査によって評価した。編集対象とした映像は、園児の母親が撮影した前記撮影映像 (メインカメラ映像) である。園児のズームアップ映像が多く、一般撮影者が撮影することの多い構図を持ったビデオ映像である。まず、見たいシーンをつなぎあわせるという通常の編集方法により、このメインカメラ映像中の 6 分程度の映像部分を 2 分に短縮する編集を行い、映像 A を作成した。結果的に、映像 A は 10 個のシーンをつなぎ合わせたものとなった。続いて、MediaBlocks を用い、映像 A 中のシーンのいくつかを、他の撮影者が撮影したシーンに差し替える編集を行った。差し替えの対象となるシーンを映像 A から選択して削除した後、この削除したシーンの撮影開始時刻と撮影終了時刻に対応する映像区間を他の撮影者の映像から抜き出して、前記削除した部分に挿入するという編集を行い、別の映像 B を作成した。結果的に、映像 B は、母親が撮影した映像 (メインカメラ映像) が 5 シーン (計 1 分) と、他の撮影者の映像 (サブカメラ映像) 5 シーン (計 1 分) とを繋いだものとなった。シーンの開始点、終了点を自由に選べないという制約があることから、映像 B にとってはかなり

表 1 単視点映像 A と多視点映像 B の比較結果

番号	質問項目	映像 A の平均得点	映像 B の平均得点	P 値(両側)
Q1	面白い	2.42	3.92	**0.00195
Q2	退屈である	3.58	1.92	**0.00195
Q3	躍動感がある	2.25	3.92	**0.00586
Q4	全体の流れがわかる	2.25	4.25	**0.00391
Q5	場の雰囲気わかる	2.58	4.25	**0.00586
Q6	自然である	3.42	3.42	1.00000
Q7	見やすい	3.67	3.50	0.94531
Q8	楽しさがわかる	3.25	3.58	0.38281
Q9	また見たい	1.83	3.25	*0.01172
Q10	魅力がある	2.17	3.92	***0.00098
(n=12, ***: p<0.001, **: p<0.01, *: p<0.05)				

不利な条件ではあるが、比較実験を行い易くするためにこのような編集を行った。

実験では、第三者を被験者とし、映像 A と映像 B とを見せて、表 1 に示される項目についてアンケート調査を行った。被験者は 12 名である(被験者を 6 名ずつのグループに分け、一方のグループには映像 A を先に見せ、他方のグループには映像 B を先に見せるようにした)。各質問は「非常にあてはまる」、「あてはまる」、「どちらとも言えない」、「あてはまらない」、「全くあてはまらない」の 5 段階で評価され、これら 5 段階に 5 点～1 点をそれぞれ与え、映像 (A) と映像 (B) の別に各質問に対する平均得点を求めた。さらに、ノンパラメトリック検定法の 1 つである Wilcoxon の符号付順位検定を用いて P 値を算出した。

表 1 に示される結果から、有意水準 5% で両側検定を行った場合に、Q1-Q5 および Q9-Q10 の項目で、映像 A と映像 B の平均得点に有意な差があり、映像 B の方が好ましいと言えることがわかった。一方、Q6, Q7, Q8 に関しては、映像 A と映像 B の平均得点に有意な差は見られなかった。この原因として、映像 B は、映像 A から削除されたシーンの撮影開始時刻と撮影終了時刻を用いて機械的に他の撮影者の映像を抜き出して作られており、映像中の意味的なつながりが中断された箇所がいくつか存在し、これが「(Q6) 自然である」または「(Q7) 見やすい」という項目の評価を下げたと考えられた。

## 5. 改良

評価実験を実施した結果、ビデオ収集機能に関し、DV カメラの内部時計あわせに手間がかかるという問題がユーザから指摘された。MediaBlocks では、DV テープに記録されている撮影日時情報を手がかりとして映像の同期処理を行う仕組み上、撮影に用いる DV カメラの内部時計がすべて同期していることが必要である。しかしながら、DV カメラの内部時計が正確に標準時刻に合っているケースは極めて少ないと考えられ、実際、今回の実験においても、撮影に使った DV カメラ間で最大 10 分程度の内部時計の誤差があった。

著者らは、5 人の被験者（大学生）に対し、2 台の DV カメラの内部時計の日本標準時（JST）との差をストップウォッチを用いて測定し、手動補正するタスクを課した。5 人とも DV カメラの日時あわせ方法を数回練習した後に、計測作業を 3 回実施した。この実験の結果、DV カメラを起動してから、内部時計の差を計測するために、平均 122 秒（標準偏差=26.2）かかる結果となった。この時、時計あわせの精度としては、最大 5 コマ（165ms 程度）の誤差が発生した。

そこで著者らは、前述した DV ビデオキャプチャツールを改良し、日本標準時と DV カメラの内部時計の差を算出し、撮影日時情報を自動的に補正する機能を追加した。具体的には、本ソフトウェアに SNTP（Simple Network Time Protocol）を使ってネットワーク経由で正確な日本標準時を取得する機能を実装し、PC に IEEE1394 ケーブル経由で DV カメラを接続した際に、DV カメラの内部時計を取得して、日本標準時との差を算出するようにした。DV カメラを撮影モード（カメラモード）に設定することにより、ビデオカメラの内部時刻を IEEE1394 ケーブル経由で PC に取り込むことができることを利用して本機能は実装されている。DV キャプチャする際に PC と DV カメラとを IEEE1394 接続しなければならないが、この際、同時に、DV カメラの内部時計を取得するようになっている。

前記 5 人の被験者に対し、本ソフトウェアを用いて 2 台の DV カメラの内部時計の差を算出する実験を 3 回ずつ行ったところ、DV カメラを起動してから、内部時計の差を計測するまでの作業が、平均 33.6 秒（標準偏差=5.31）に短縮された。また、時刻あわせ誤差は 1 コマ（33ms）以下であった。この結果から、手作業で補正を行った場合の約 1/4 で作

業が完了することがわかり、時刻あわせ誤差も大幅に減少したことから、改善の効果があったと評価した。

## 6. おわりに

近年、デジタルビデオカメラは小型化・軽量化しており、運動会や結婚式の撮影、旅行先での撮影など、極めて幅広い分野で用いられており、MediaBlocks は、これらの幅広い用途に利用できると考えている。

なお、自分が撮影した未編集段階の映像を公開することは今まで存在しなかった行為であり、この行為に抵抗感を感じるユーザも多いと思われる。さらに、撮影者のプライバシーの問題に加え、被写体のプライバシーの問題があり、予め承認されたグループメンバー間で映像共有するというのが、さしあたり現実的であるように思われる。プライバシーの問題、映像権利の問題については今後の課題としたい。

## 参考文献

1. 内閣府：消費動向調査 <http://www.cao.go.jp/> (2004)
2. 谷，松下，市村，中村：遠隔共同動画編集のため Web システム，情報処理学会全国大会，6Q-4 第 4 分冊，pp.295-296 (2004).
3. 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 (2002).
4. 松本 剛著：図解 デジタルビデオ編集のしくみ ディー・アート社(2001).
5. 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).
6. <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).
7. <http://www.microsoft.com/japan/windows/directx/>.
8. <http://www.jp.realtalk.com/>.