

## 複数カメラを用いた 自動スイッチングによる自然な映像ストリームの生成手法

津村 弘輔<sup>†</sup> 加藤 淳也<sup>†</sup> 住谷 哲夫<sup>†</sup> 重野 寛<sup>‡</sup> 岡田 謙一<sup>‡</sup>

### 概要:

本研究では、会議中継や講義中継など常に映像が流れる中で演出した映像を自動生成することを目的とした、タイムシフトを用いたスイッチング手法を提案する。本手法では、複数台のカメラが撮影した映像・音声をメモリ上に微小時間である  $t$  秒蓄積する。その  $t$  秒の間に、センサーを用いて取得したイベント情報を基に蓄積した映像と音声のスイッチング方法を決定する。この決定された方法を基に、蓄積した  $t$  秒前の複数の映像・音声から一本の映像・音声を選択、切替、表示する。これにより、中継用途で準リアルタイムに演出された映像を提供可能となった。本手法を導入したプロトタイプを用いて会議中継を自動撮影した。評価実験を通じて、自然で退屈しない映像であること、ドラマや映画で用いられる演出用のスイッチングを実現できることを確認した。

## Generation technique of natural video image stream by automatic switching with multiple cameras

Kousuke Tsumura<sup>†</sup>, Junya Kato<sup>†</sup>, Tetsuo Sumiya<sup>†</sup>, Hiroshi Shigeno<sup>‡</sup>, Ken-ichi Okada<sup>‡</sup>

### Abstract:

The purpose of this research is to automatically switch video images shot by multiple cameras, to produce video images, which do not bore, nor discomfort the viewer due to natural switching. We proposed automatic camera switching using time shifting. First, a length of  $t$  of multiple data of video and audio shot by multiple cameras are stored. The decision of switching method of video image and audio is based on events acquired with sensors while the  $t$ . Finally, a set of video and audio data stored  $t$  beforehand is selected from a number of sets of data, and the shot is switched and displayed. I implemented a prototype system, which generates a single image stream using these techniques to automate switching. I evaluated experiments of the prototype system. As a result, I confirmed that proposal switching correspond with the switching of the drama and do not bore, nor discomfort the viewer due to natural switching.

### 1. はじめに

近年、映像制作の専門家でない人が、映像を撮影したり編集したりする機会が増えてきている。これに伴い、プロの編集者が作るような質の高い映像を自動的に生成する研究が行われている。プロのカメラマンが撮影するような映像を撮るために、カメラのズーム機能・首振り機能を用いて料理番組<sup>1)</sup>、プレゼンテーション<sup>2)</sup>、スポーツ<sup>3),4)</sup>、会議<sup>5)</sup>や講義<sup>6)</sup>における適した撮影領域や撮影対象を自動的に決定するカメラワークの研究がある。一方、このカメラワークに加え、プロのスイッチャーが複数のカメラ映像から現在のシーンを表現するのに適した一つの映像を選択するためのスイッチングも重要である。会議中継や講義中継など常に映像を視聴者側に送出する場合、複数カメラからの入力映像をリアルタイムにスイッチングを行っている。

従来の研究では現在までの発話・移動・板書などのイベント情報に応じたスイッチング<sup>2),4)</sup>を行っていた。しかし、一度映像と音声をメディアに蓄積した後にプロのディレクターが時間をかけて編集を行うドラマや映画などの蓄積編集番組と異なり、中継用途の映像では十分な編集時間がないため、臨場感が伝わらず、「雰囲気」「演出」といった観点から問題が生じる。そこで本研究では、タイムシフトを用いた複数カメラのスイッチング手法を提案する。本手法では、各カメラからの映像とマイクからの音声を送信前に一定時間蓄積する。その際、センサーから取得したリアルタイムなイベント情報を基に蓄積した映像と音声のスイッチングを決定する。結果的に、送出映像と音声は蓄積した分だけ遅延するが、編集者が数秒先の出来事を完全に予測できた時の編集と同等の編集が可能となり、従来より演出された映像にすることができる。実際に提案手法のプロトタイプシステムを実装し、評価実験により本手法の有効性を確認する。以下、2章ではリアルタイムスイッチングの課題について、3章では提案手法について、4章では実装について、5章では評価実験と結果と考察について、6章

<sup>†</sup> 慶應義塾大学大学院理工学研究科  
Graduate School of Science and Technology, Keio University

<sup>‡</sup> 慶應義塾大学理工学部  
Faculty of Science and Technology, Keio University

を本論文のまとめとする。

## 2. リアルタイムスイッチングの課題

図1に、A、B、Cの発話におけるリアルタイムスイッチングの撮影時刻と送出時刻の関係を示す。

$t_i$  は、カメラが撮影した時刻である撮影時刻、 $t_o$  は、スイッチングを行い視聴者に映像を送出した時刻である送出時刻と定義する。リアルタイムスイッチングでは、時刻が経過した時でも常に  $t_i = t_o$  の関係が成り立つ。中継用途の映像におけるシーンを演出するためには、常に適した時刻に適したカメラに切り替えることが重要である。しかし、予め専門知識を身に付けたプロのスイッチャーでさえ、リアルタイムにそのようなカメラに切り替えることは難しい。例えば、サッカー中継では、プレイ中に挿入した過去のハイライトシーンが、現在の選手のゴールシーンと重なり、慌ててゴールした選手にカメラを切り替える場面や、朝まで生テレビなどの討論番組では、参加者の発言に対して遅れてカメラが切り替わる場面が生じる。これは、スイッチャーが突発的な選手の行動や参加者の発言に対する認識の遅れや場面の予測だけでは完全に把握できないことが原因であり、人手によるスイッチングでは限界がある。

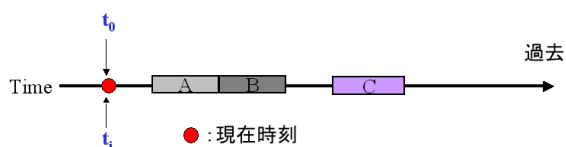


図1 リアルタイムスイッチングの撮影時刻と送出時刻の関係

これらの問題を解決するために、計算機が会議参加者の発話を認識してスイッチングを自動で行う研究<sup>4),7)</sup>があり、人間の認識の遅れを吸収できるが、次の話者を考慮したスイッチングはできない。また、遠隔講義において状態遷移を用いる方法<sup>8)</sup>もあるが、予測が外れる場合が生じる。

仮にスイッチャーや計算機が、数秒先の出来事を完全に把握できれば、サッカー中継におけるゴール時には、ゴールポストに設置されたカメラから選手のシュートシーンを捉えた映像をいつでも作り出すことができる。討論番組の生放送では、突然怒り出す参加者がいた際、怒り出す前の緊迫した様子を捉えたカメラの映像をスイッチングにより作り出すことができ、従来より魅力的な映像を作るための編集方法が増えるといえる。

## 3. 提案

### 3.1 タイムシフトによるスイッチング

撮影中の映像を、微小時間連続的に遅らせて送出することを、タイムシフトを用いたストリーミングとする。図2に、A、B、Cの発話におけるタイムシフトによるスイッチングの撮影時刻と送出時刻の関係を示す。

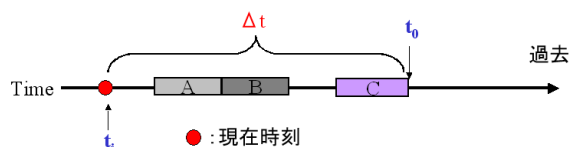


図2 タイムシフトによるスイッチングの撮影時刻と送出時刻の関係

タイムシフトによるスイッチングは、蓄積しながら送出するためにカメラが撮影した時刻  $t_i$  に対してスイッチングを行い視聴者に映像を送出する時刻  $t_o$  が蓄積時間  $t$  だけ遅れており、時刻が経過した時でも常に  $t_i = t_o + t$  の関係が成り立つ。この蓄積している間にセンサーを用いて取得したイベント情報は、時刻  $t_o$  に対して未来のイベント情報であり、この情報を用いて編集の一つであるスイッチングを決定する。図2の例では、編集者や計算機は最初の時刻  $t_o$  の時点でAが話した後にBが話すこと、そのあと間が空いた後にCが話すことをスイッチングと送出を行う  $t_o$  の時点でわかっている。一方、リアルタイムスイッチングでは常に  $t_i$  と  $t_o$  の時刻が同じであるため、時刻  $t_o$  に対する未来のイベント情報を用いることはできない。

人間は、一方向的な生放送の中継において数秒程度映像を遅れて見ても気にならないという感覚を持っている。事実、BS衛星放送は、電波が地上から放送衛星まで往復するために、地上波アナログ放送に比べて映像・音声共に0.2秒程度の遅延があり、地上波デジタル放送は、各中継局のA/D変換の繰り返しのために、地上波アナログ放送に比べて3~4秒程度の遅延が生じているが、我々は、生放送の国会中継、スポーツ中継、ニュースをそれらの遅延を意識することなくリアルタイムの番組として見ている。この許容される遅延時間を利用して、意図的に映像と音声を遅らせ送出することで、スイッチャーなどの編集者は、問題点の後半で述べたような数秒先の出来事を完全に把握した上での映像編集と同じことが可能となる。

タイムシフトの概念を用いたものとして、地震の発生時などその一瞬を収録することができるスキップバックレコーダーや交通事故が起きた際に、車載カメラで事故前後の映像を記録するドライブレコーダーがあるが、これは録画ボタンのON/OFFの単純な自動制御のみであり、中継用途で自動的に映像を編集する試みは成されていない。

### 3.2 実現方法

複数台のカメラが撮影した複数の映像・マイクの音声をメモリ上に  $t$  秒間蓄積する。次にその  $t$  秒の間に、センサーを用いて取得したイベント情報を基に蓄積した映像と音声のスイッチング方法を決定する。この決定された方法を基に、蓄積した  $t$  秒前の複数の映像・音声から一本の映像・音声を選択、切替、表示する。

今回は一方向的な生放送の中継の中から、撮影対象として会議中継を選択した。また、センサーにはマイクを用いて音声情報を基に音声の強弱、会話の長短、間の発話状況

と順序を認識し、この情報と各カメラがどの被写体を映しているのかという会議空間のレイアウトを基にスイッチング方法を決定している。本方法は利用するセンサーの種類と組み合わせを変えることによって様々な場面で拡張性のあるスイッチングを実現できる。例えば、講義中継などに応用する場合は、位置センサーを用いて取得した講義者の位置情報を基に、講義者が歩き始める方向のカメラから撮影した映像を生成し、受講者が飽きないような演出を加えることも可能である。

### 3.3 蓄積編集番組で用いられるスイッチング

会議中継を演出するために、映画やドラマなどの蓄積編集番組で用いられる代表的なスイッチングを以下の図3、4に紹介する。図3のずり上げ及びずり下げスイッチングは、主に人物同士の発話をベースに編集する際にショットの切り替えに対して音声をわざとずれこませる技法である。Aがまだ話し続けている時に、次の話者となるBの映像に切り替わるのが、ずり上げスイッチングであり、反対にAが話し終わり、Bが話し始めている時にまだAの映像が映っているのが、ずり下げスイッチングである。このスイッチングによって余韻を持たせたり、聞いている人物の態度を強調するというような効果がある。

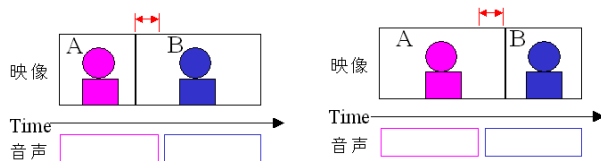


図3 映像のずり上げ(左)とずり下げ(右)スイッチング

図4のシーンカメラスイッチングは、シーンに登場する物の配置や人物の位置関係を認識してもらう技法である。A、B、Cが図4の左の様な位置に座っている時、カメラをテーブルから遠ざけた位置に置き右図の様な全員の姿を映すカメラに切り替えることで、その場の雰囲気や登場人物の位置関係、またその場の環境などが一目で分かるというような効果がある。このショットは、沈黙時間がある一定以上続く場面や数秒以内に任意の順番で複数の人が次々と話す場面で用いられている。このスイッチングは、リアルタイムに切り替える時でも利用されているが、沈黙間隔が短い場合でシーンカメラに切替えると、短時間に複数回のスイッチングとなり不快感を感じさせる問題があった。

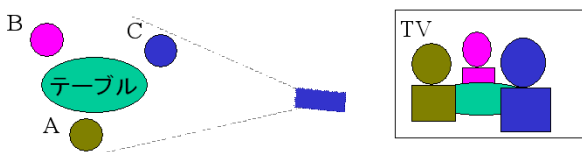


図4 シーンカメラスイッチング

これらのスイッチング技法は、見ている人を退屈させない上で有用であるが、次の話者は誰か、何秒後に話すか、沈黙時間は長いのか、など数秒後の会話の状況を把握しなければならぬため、常に複数の映像が流れる中でのスイッチングに用いることは難しかった。

### 3.4 スwitchingの分類と動作条件

本手法にて用いる5種類のスイッチングの種類とその動作条件を表1に示す。スイッチングの種類は、リアルタイムでのスイッチングでも実現可能な発話スイッチング、オーバーラップスイッチング、オーバーラップ戻しスイッチングの基本的なスイッチングに加え、蓄積編集番組で用いられている映像のずり上げスイッチング、シーンカメラスイッチングの2種類の計5つのスイッチングである。また、映像のずり下げスイッチングは、タイムシフトを用いなくてもリアルタイムに実現可能なので今回は導入していない。各スイッチングの動作条件を複数満たす場合は、条件が厳しい表中の下の方のスイッチングから優先的に動作する。

スイッチングの種類	概要(上)と動作条件(下)
発話	発話と同時に話者に切り替え (発話前に1秒以上の沈黙がない時)
映像のずり上げ	発話1秒前に話者に切り替え (発話前に1秒以上の沈黙がある時)
オーバーラップ	会話を被せた話者に切り替え (会話の重複発生時)
オーバーラップ戻し	会話を被せられた話者に切り替え (会話の重複終了時)
シーンカメラ	参加者全員を映すカメラに切り替え (5秒以上の沈黙がある時)

## 4. 実装

提案手法に基づいて、会議を自動撮影するプロトタイプシステムを構築した。

### 4.1 撮影システム

撮影システムの構成図を図5に示す。縦5m x 横10m程度のスペースに4つの固定カメラを配置した。カメラ1~3は、話者を映すカメラとして(カメラ1はA、カメラ2はB、カメラ3はCを撮影)、カメラ4は会議空間全体の撮影を行うシーンカメラとして、使用した。

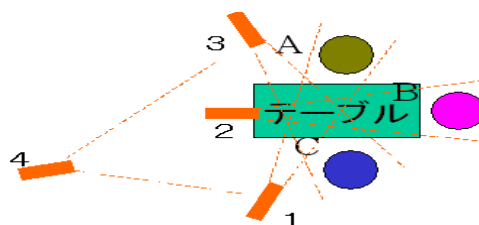


図5 会議空間のレイアウト

## 4.2 編集システム

撮影の流れを示した編集システムの構成図を図6に示す。

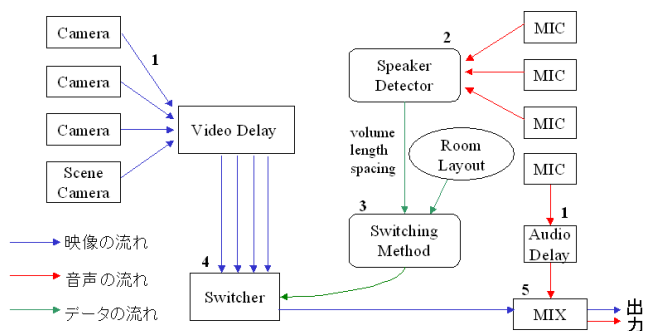


図6 プロトタイプによる撮影の流れ

- (1) それぞれのカメラで撮影されたアナログ映像の出力は、映像遅延装置に入力され、 $t$  秒間蓄積される。同様に参加者全体の音声を拾う集音マイクで取得したアナログ音声の出力もまた、音声遅延装置に入力され、 $t$  秒間蓄積される。
- (2) この  $t$  秒の間に話者の特定を行う。話者の特定は、参加者が自身に装着されたマイクに向かって話すことでシステムに認識させる。1 秒間に 4000 回のサンプリングを行い 8bit で量子化する。閾値以上の入力が連続した時点で話者と特定する。
- (3)  $t$  秒間で検出した音声の強弱 (volume)、会話の長短 (length)、間 (spacing) の発話状況と会議空間のレイアウトを基にスイッチング方法を決定する。
- (4) 遅延させた映像はスイッチャーへ入力される。決定したスイッチング方法を基にこのスイッチャーを制御することにより、1 本の映像を選択・切替・出力する。
- (5) この出力された映像は先ほどの遅延させた会議空間全体の音声とミックスされる。

### 4.3 プロトタイプの実装画面

図7に会議参加者が3人の場合の実装画面を示す。画面上部には、遅延させたすべての映像が表示される、そのすぐ下には、赤色と青色で塗られた矩形がある。

### 4.4 スwitching実行例

図8に3人で「卒業旅行の行先について」話し合う調整会議を、提案手法を用いて撮影・switchingした場合と既存手法を用いて撮影・switchingした場合のswitching時刻と種類を示す。この会議のレイアウトは図5の通りであり、表3に示す内容を47秒間話し合った。ここで使われているカメラは、会議室全体を映すシーンカメラ1台、各参加者を映すカメラ3台、計4台である。横軸に経過時間を示しており、タイムライン上の矢印が提案手法を用いて決定したswitchingポイントであり、下の矢印が発話自動切替システムを用いて決定したswitchingポイントである。蓄積編集番組で用いられているswitching



図7 実装画面

グが常に映像が流れる中で導入されていることがわかる。

## 5. 評価

本手法を取り入れたシステムを用いてswitchingして得られた映像が、ドラマでのずり上げswitchingに近づいているか確かめるため、そしてそれが閲覧者にどのような影響を与えるかを確かめるために評価を行った。

### 5.1 定量評価

#### 5.1.1 カバー率, 有効率

本システムを用いてswitchingして得られた映像が、ドラマでのずり上げswitchingに近づいているか確かめるために、ドラマの音声ファイルを基に本システムが映像のずり上げヶ所を決定し、実際のドラマの映像のずり上げヶ所と本システムの映像のずり上げヶ所の比較を行った。ドラマの音声ファイルの中でも、演出されたswitchingが多いドラマの一つである Trick の中から一対一の対話シーンを手動で抽出した音声ファイルを用いた。そのシーンの合計時間は23分27秒であり、ドラマでの映像のずり上げヶ所は57ヶ所あった。

本システムの映像のずり上げヶ所がドラマのそれに対していかに一致したかを評価するためにカバー率  $P$  を以下のように定義した。また、本システムの映像のずり上げヶ所がドラマのそれに対していかに無駄なく一致したかを評価するために有効率  $E$  を以下のように定義した。

- カバー率  $P$

$$P = N_m / N_d \times 100(\%)$$

( $N_m$ :本システムの映像のずり上げヶ所がドラマのそれと一致したヶ所の合計  $N_d$ :ドラマの映像のずり上げヶ所の合計)

- 有効率  $E$

$$E = N_m / N_s \times 100(\%)$$

( $N_m$ :本システムの映像のずり上げヶ所がドラマのそれと一致したヶ所の合計  $N_s$ :本システムの映像のずり上げヶ所の合計 =  $N_m$  + 本システムの余分なずり上げヶ所の合計)

カバー率が高くても、有効率が低ければ、一致したヶ所は多いが無駄な映像のずり上げヶ所が多く、逆にカバー率

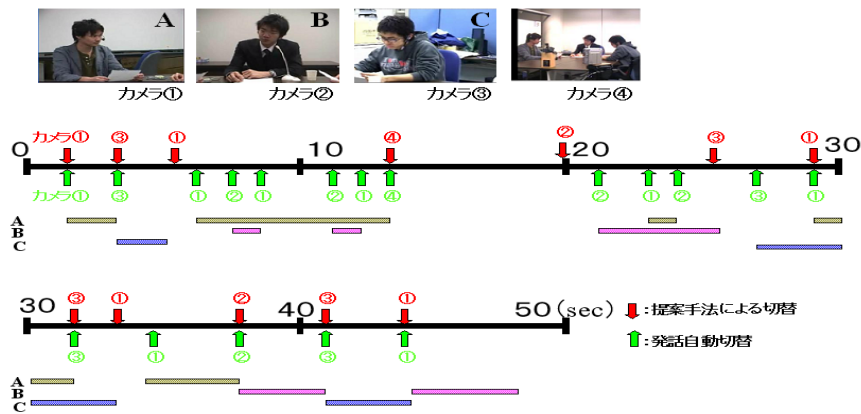


図 8 スイッチングポイント

が低く、有効率が高ければ、無駄な映像のずり上げヶ所は少ないが、一致したヶ所は多いことを意味する。つまり、カバー率、有効率ともに高い値の時ほど、無駄なく一致していることになり、より良い結果といえる。今回の実験において「一致した」とは、ドラマの映像のずり上げヶ所の時刻から前回の発話終了時刻と、ドラマのその時刻から次の発話開始時刻までの時間帯において、本システムでの映像のずり上げが起こった時のみを一致したとした。

### 5.1.2 比較システム

本システムの有用性を打ち出すために、比較システムを用意した。用意した比較システムは、過去の発話履歴から確率的に次の話者と発話時刻を予測してリアルタイムに映像のずり上げスイッチングを行う状態遷移自動切替システムである。次の話者の予測に関しては、発話終了後、計算機が図 9 の状態遷移確率から次の話者を決定する。

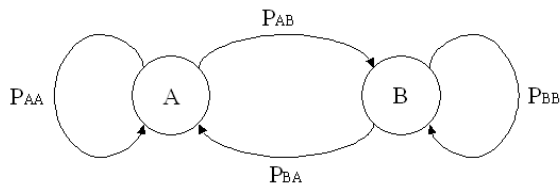


図 9 会話の状態遷移図

次の映像のずり上げ時刻の予測に関しては、計算機が過去の発話履歴から各参加者から各参加者への平均沈黙時間を計算する。これら 2 つの計算結果を基に、計算機が切替対象とずり上げ時刻を決定する。

### 5.1.3 実験結果

表 4 にカバー率、有効率の結果を示す。

表 2 カバー率、有効率の評価結果

評価項目/システムの種類	本システム	状態遷移システム
カバー率 P	72%(Nm=41)	29%(Nm=17)
有効率 E	64%(Ns=64)	61%(Ns=28)

本システムでは、状態遷移自動切替システムに比べて有効率 64 %、カバー率 72 %と無駄なくドラマでのずり上げヶ所に多く一致していることがわかる。一方、状態遷移自動切替システムでは、有効率 61 %と本システムの有効率とあまり変わらないが、カバー率が 29 %と低いことがわかる。つまり、無駄なく切替えているもの一致している映像のずり上げヶ所が少ないということがわかる。本システムの有効率及びカバー率が 100 %にならなかったのは、ドラマにおける映像のずり上げヶ所が発話時に必ずしも用いられていないからである。発話時に、その発話に対する聞き手の映像が映っていたケース、二人を映したケース、の 2 通りの発話スイッチングのケースが見られた。特に書類を渡しながら発話するなど動作と共に発話する場合は、二人を映したケースが多かった。対処方法としては、発話スイッチングとずり上げスイッチングの割合を基にドラマにおける度数分布を求め、その分布の出現確率に従ってスイッチングしたり、画像処理と並行するなどにより、カバー率と有効率を高めることができると考える。

### 5.2 アンケート評価

本システムを用いてスイッチングして得られた映像が、閲覧者にどのような影響を与えるかを確かめるために、大学生の被験者 13 人が図 7 の実装画面に表示された、本システムによって制作した 3 分の映像と発話時にその話者に切替える発話自動切替システムにより制作した 3 分の映像を用いて閲覧し、それぞれについてアンケートに 5 段階で評価した。アンケートの質問項目を表 3 に示す。位置関係に関する質問項目 (項目 5, 6) 参加者の映り具合に関する質問 (項目 4, 8, 9, 10, 12) 映像のスイッチングに関する質問 (項目 1, 2, 3, 7, 11) を用意した。

アンケートの結果を表 3 に示す。各質問は「まったくあてはまらない」、「あまりあてはまらない」、「どちらともいえない」、「ややあてはまる」、「かなりあてはまる」の 5 段階にそれぞれ 1 点から 5 点を与え、映像別に質問に対する平均得点を求めた。表中の各項目の値は評価値の平均得点である。さらに、評点に有意差があるか確認するため

表 3 アンケート結果

質問項目/システムの種類	本システムの平均得点	発話自動システムの平均得点	Wilcoxon 符号付順位検定 P 値
1. 切り替えのタイミングは適切だったか	4.54	2.31	***0.00024
2. 見やすい映像だったか	4.31	2.38	***0.00049
3. カメラの切り替えに違和感を感じなかったか	4.46	2.08	***0.00049
4. 話し手がよく分かった	4.54	2.92	***0.00098
5. 人物の位置関係がつかめた	3.92	2.54	**0.001563
6. その場の状況が分かりやすかった	4.08	2.53	**0.00195
7. 見たい映像に切り替わっていた	4.31	2.85	**0.00195
8. 画面上の人物の表情や身振りが分かった	4.15	3.15	**0.00195
9. 画面上の人物の存在感があった	4.23	3.08	**0.00781
10. 議論の流れがつかめた	4.31	3.54	*0.01563
11. 映像に退屈しなかった	3.92	2.85	*0.1563
12. 誰と誰が会話しているのかわかった	3.77	3.15	0.05469

Wilcoxon の符号付順位検定 p 値を求めた。

本システムによる映像はいくつかの項目で発話自動切替システムの映像を大きく上回ることができた。項目 1, 2, 3 の結果から、危険率 0.1 % 以下で違和感のない自然なスイッチングになっているという評価を得た。次に、項目 5, 6 から、危険率 1 % 以下で位置関係が明確である結果を得た。これは、会議参加者が 3 人と人数が少なかったこと、被験者に予め参加者の座席配置を伝えていたところによるところが大きい。次に、項目 8, 9 から、今回用いたショットは、シーンカメラ以外すべて単独の参加者を映す構図であり、個人の表情や存在感は必然的に出るといえる。複数の参加者を映す構図を導入する必要があり、その際は個人の表情や存在感が薄らいでしまう問題が生じると考えられる。最後に、項目 11 の結果から、危険率 5 % 以下で映像に退屈しないという評価を得た。項目 11 が、項目 1, 2, 3, 4 ほど良くなかった理由として、本システムの比較対象である発話自動切替システムでは、短時間におけるスイッチング数が本システムより多く、変化に富んだ映像になっていたからであると考えられる。しかし、項目 1, 3 から、発話自動切替システムがスイッチングした映像は、被験者にとって切替のタイミングが不適切であり、違和感を感じる映像である。単純に切替回数が多ければ、良いということではないことがわかる。

## 6. おわりに

リアルタイムスイッチングでは十分な編集時間がないため、演出に制約があるという問題があった。人間は、一方的な生放送の中継において数秒程度映像を遅れて見ても気にならないという感覚を持っていることを述べた。そこで本研究では、タイムシフトを用いた会議における複数カメラのスイッチング手法を提案した。まず始めに複数台のカメラが撮影した映像・音声をメモリ上に t 秒間蓄積し、映像を選択する。選択には t 秒の間に音声入力を基に音声の強弱、会話の長短、間の発話状況を用いている。最後に、蓄積した t 秒前の複数の映像・音声から一本の映像・音声を選択、切替、表示する。これにより準リアルタイムに映像を提供しつつ常に映像が流れる中でドラマや映

画で用いられている演出用のスイッチングを可能とした。評価実験を通じて、提案手法でのスイッチングがドラマでのスイッチングに近づけること、不快感を与えない自然なスイッチングを実現できることを確認した。また、スイッチングにより視聴者が退屈しない映像ができた。

謝辞

本研究の一部は、21 世紀 COE プログラム研究拠点形成費補助金のもとに行われた。ここに記して謝意を表す。

## 参 考 文 献

- 1) C.S.Pinhanes and A.F.Bobick. Approximate world models: Incorporating qualitative and linguistic information into vision systems. Proc. AAAI '96, pp. 1116-1123, Aug 1996.
- 2) 尾関基行, 伊藤雅嗣, 中村裕一, 大田友一. 複合コミュニティ空間における注目の共有 ~ 人物動作理解による物体への注釈付け ~. VRSJ 第 6 回大会論文集, Sep 2001.
- 3) 松本圭介, 須藤智, 斎藤英雄, 小沢慎治. サッカー放送における視点選択のための多視点画像の統合によるボール追跡. 電学論, Vol. 121-C, No. 10, pp. 1530-1539, Oct 2001.
- 4) 井口泰典, 土居元紀, 真鍋佳嗣, 千原國宏. スポーツ映像放送のための実時間映像解析によるマルチカメラの自動制御と自動スイティング. 映像学誌, Vol. 56, No. 2, pp. 271-279, Feb 2002.
- 5) 井上智雄, 岡田謙一, 松下温. テレビ番組のカメラワークの知識に基づいた tv 会議システム. 情報処理学会論文誌, Vol. 37, No. 11, pp. 2095-2104, Nov 1996.
- 6) M. Minoh and Y. Kameda. Image a 3d lecture room by interpreting its dynamic situation. Proc. 4th Int. Workshop on Cooperative Distributed Vision, pp. 371-412, March 2001.
- 7) 大西正輝, 影林岳彦, 福永邦雄. 視聴覚情報の統合による会議映像の自動撮影. 電子情報通信学会論文誌 D-II, Vol. J85-D-II, No. 3, pp. 537-542, 2002.
- 8) 先山卓朗, 大野直樹, 椋木雅之, 池田克夫. 遠隔講義における講義状況に応じた送信映像選択. 電子情報通信学会論文誌 D-II, Vol. J84-D-II, No. 2, pp. 248-257, 2001.