タイムシフトを用いた会議中継カメラの自動スイッチング手法

加藤淳也[†††] 住谷哲夫[†] 井上亮文[††]
重野寛[†] 岡田謙一[†]

本研究では，会議や講義の中継などつね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 Δt 秒間蓄積する．Δt 秒の間に，センサを用いて取得したイベント情報に基づいて蓄積した映像と音声のスイッチング方法を決定する．これにより，若干の遅延は出るが映像表現が改善された映像を提供可能である．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，比較対象のドラマと同等な演出が可能であること，アンケート結果から自然で退屈しない映像表現が可能であることを確認した．

Automatic Camera Switching Method for Meeting Scenes Using Time Shifting

Junya Kato,[†††] Tetsuo Sumiya,[†] Akifumi Inoue,[††]
Hiroshi Shigeno[†] and Ken-ichi Okada[†]

The purpose of this study is to produce smooth transition of video images by multiple cameras automatically. This paper proposes a video switching method for meeting scenes based on time shifting. First, the audio/video streams are captured by multiple video cameras. Those streams are not directed and transmitted in real time, but are buffered on a memory for Δt seconds. During the buffering, the conversational history of the scene is collected by various sensors simultaneously. Finally, the buffered streams are directed and merged into one stream based on the history. Even though there remains some delay, the stream can be well directed than that directed in real time. A prototype system was implemented and evaluated by experiments. The experimental results indicated that the prototype system could direct the meeting scene to be smooth as well as actual TV-drama program.

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．これにともない，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．カメラのズーム機能・首振り機能により，撮影領域や撮影対象を自動決定するカメラワークを実現する研究として，料理番組[1)]，プレゼンテーション[2)]，スポーツ[3),4)]，会議[5)]や講義[6)~8)]に用いたものがある．一方，このカメラワークに加え，プロのスイッチャ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1つの映像を選択するためのスイッチングも演出には重要である．

会議や講義中継などつね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する必要がある．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[2),4),6)]が主体であった．しかし，ドラマや映画などは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が時間をかけて編集を行う．これに対し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会議中継カメラの自動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会議の状況を判断し，バッファ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

† 慶應義塾大学大学院理工学研究科
Graduate School of Science and Technology, Keio University
†† 東京工科大学コンピュータサイエンス学部
Faculty of Computer Science, Tokyo University of Technology
††† エヌ・ティ・ティ・コミュニケーションズ株式会社
NTT Communications Corporation

予測できたとき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映像の時刻の関係を示す．この図の横軸は時間の経過を示しており，軸上に表示された矩形が各参加者の発話時間を示している．A が撮影開始から 2 秒後に発話を開始し，その 1.5 秒後に発話をやめている．A が発話をやめるのと同時に B が発話を開始し，その 1.5 秒後に発話をやめている．そして，1 秒間の沈黙時間の後 C が発話を開始し，その 3 秒後に発話をやめている．ここで t_i はカメラがあるシーンを撮影した時刻である撮影時刻，t_o は視聴者に送出される映像の映像中の時刻と定義する．リアルタイムスイッチングではつねに $t_i = t_o$ の関係が成り立つ．よって，このような状況でシーンを演出するためには，つね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，専門知識を身に付けたプロのスイッチャでさえ，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．たとえば，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中に選手がゴールしてしまい，慌ててゴールした選手にカメラを切り替える場面がある．また，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ある．前者は，スイッチャが突発的な選手の行動に対する認識の遅れ，後者は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る．

これらの問題に対して，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[5),9)] がある．この手法では，人間の認識の遅れを吸収できるが，場面の予測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て次の行動を予測する方法[10)] もあるが，予測が外れる場合が生じる．

ここで仮にスイッチャや計算機が数秒先の出来事を完全に把握できれば，従来より魅力的な映像を作るための編集方法が増える．サッカー中継におけるゴール時には，ゴールポストのカメラに事前に切り替え最も迫力がある映像を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ときには，怒り出す前の緊迫した表情をとらえたカメラの映像に事前に切り替えておくことができる．

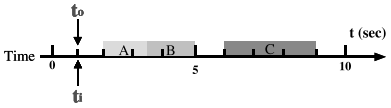


図 1 リアルタイムスイッチングの撮影時刻と送出映像の時刻の関係
Fig. 1 The relationship between shooting time and video time of realtime switching.

3. 提　　案

3.1 タイムシフトによるスイッチング

本論文では，撮影中の映像を故意に一定時間遅らせて送出したものを，タイムシフトを用いたストリーミングと定義する．**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映像の時刻の関係を示す．図の横軸，各参加者が発話するタイミングは図 1 と同様である．タイムシフトによるスイッチングでは，映像を送出前に蓄積するためシーンを撮影した時刻 t_i に対して，視聴者に送出される映像の映像中の時刻 t_o が蓄積時間 Δt（図の例では $\Delta t = 6\,(\mathrm{sec})$ とした）だけ遅れており，つねに $t_i = t_o + \Delta t$ の関係が成り立つ．この Δt の間にセンサを用いて取得したイベント情報は，時刻 t_o から見れば未来のイベント情報であり，この情報を用いてスイッチングを行う．図 2 の例では，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，カメラをスイッチングし送出する映像の時刻 t_o の時点で分かっている．一方，リアルタイムスイッチングでは，つねに 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．衛星放送は，電波が地上から放送衛星まで長距離を往復するために，地上波アナログ放送に比べて映像・音声ともに 0.2 秒程度の遅延がある．また，地上波デジタル放送は各中継局の A/D 変換の繰返しのために地上波アナログ放送に比べて 3〜4 秒遅程度の遅延が生じている．しかし，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 3〜4 秒程度の遅延時間を利用して，意図的に映像と音声を遅らせて送出することで，スイッチャなどの編集者は，数秒先の出来事を完全に把握したのと同等の映像編集が可能となる．

タイムシフトの概念を用いたものとして，地震の

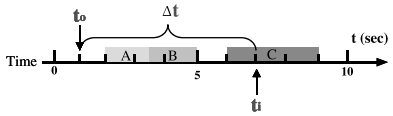


図 2 タイムシフトによるスイッチングの撮影時刻と送出映像の時刻の関係
Fig. 2 The relationship between shooting time and video time of switching using time shifting.

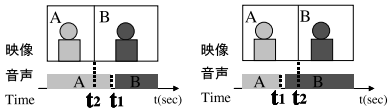


図 3 映像のずり上げ（左）と映像のずり下げ（右）スイッチング
Fig. 3 Vision shift switching.

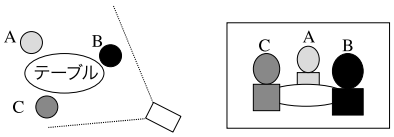


図 4 シーンカメラスイッチング
Fig. 4 Scene camera switching.

発生の瞬間を収録することができるスキップバックレコーダ[11] 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[12] がある．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メモリに一定時間蓄積する点では同じだが，映像の編集は想定していない．

3.2 実 現 方 法

本手法ではまず，複数台のカメラが t_i において撮影した複数の映像・音声をメモリ上に Δt 秒間蓄積する．次に，その Δt 秒の間にセンサを用いて発話などのイベント情報を蓄積しておく．最後に，蓄積されたイベント情報を基にシーンの状況を判断し，1 本の映像・音声に編集する．

本研究では，中継映像の例として会議シーンを撮影対象とした．イベント情報を取得するセンサとして，マイクを用いて音声の強弱，会話の長短，発話順序を認識し，この情報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している．

本手法は，利用するセンサの種類と組合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．たとえば，講義中継などに応用する場合は，位置センサを用い講師の位置情報を取得することで，講師の顔を正面から映したカメラに事前に切り替え，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 録画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録画番組で用いられる代表的なスイッチングを**図 3**，**図 4** に紹介する．

図 3 の「ずり上げ」および「ずり下げ」スイッチングは，ショットの切替えと発話の切替えのタイミングをずらす技法である．A が t_1 まで話し続けているときに，t_2 の時点で次の話者となる B の映像に切り替わるのが，ずり上げスイッチングである．反対に，A が t_1 で話し終わり，B が話し始めているときにまだ A の映像が映っていて，t_2 の時点で B の映像に切り替わるのが，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，聞き手，話し手の態度を強調したりする効果がある．

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 左のような位置に座っているとき，カメラをテーブルから遠ざけた位置に置き図 4 右のような全員の姿を映すカメラに切り替えることで，その場の状況が一目で分かるという効果がある．このショットは，実際の放送において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．この技法は，リアルタイムに切り替えるとき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り替えると，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．

これらのスイッチング技法は，視聴者を退屈させないうえで有用である．しかし，この手法を適切に利用させるには，次の話者はだれか，何秒後に話すか，沈黙時間はどの程度か，など数秒後の会話の状況を把握しなければならない．そのため，リアルタイムでの切替えに用いることは難しかった．

3.4 スイッチングの分類と動作条件

本手法で用いるスイッチングの種類と動作条件を**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，オーバラップスイッチング，オーバラップ戻しスイッチングの基本的なものに加え，録画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．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．

スイッチングの動作条件における，発話何秒前に次の

話者に切り替わるかという映像のずり上げ間隔，シーンカメラに切り替える沈黙間隔は，録画番組の 1 つであるドラマ番組の 1 対 1 の対話シーンを分析することによって決定した．映像のずり上げスイッチングの度数分布と相対累積度数分布を**図 5** に示す．横軸に映像のずり上げ間隔を秒で，左の縦軸に度数分布に対する頻度を回数で，右の縦軸に相対度数分布に対する割合を%で表している．映像のずり上げスイッチングのうち 67%が 1 秒以内に集中した．よって，提案手法でのずり上げ間隔を 1 秒と決定した．また，沈黙時におけるシーンカメラへのスイッチングの度数分布と相対累積度数分布を**図 6** に示す．横軸に沈黙間隔を秒で，左の縦軸に度数分布に対する頻度を回数で，右の縦軸に相対度数分布に対する割合を%で表している．沈黙時のシーンカメラスイッチングのうち 97%のスイッチングが 4 秒以上の沈黙がある場合に集中した．よって，提案手法での沈黙間隔を 4 秒と決定した．以上の結果を考慮し，システムが沈黙時間 4 秒を認識できるように映像・音声を蓄積する時間 Δt を 4 秒と決定した．

表 1 スイッチングの分類と動作条件
Table 1 The switching classification and operating condition.

スイッチングの種類	説明（上） 動作条件（下）
発話	発話と同時に話者に切替え
	発話前に 1 秒以上の沈黙がないとき連続的に会話が切り替わる
映像のずり上げ	発話 1 秒前に話者に切替え
	次の話者の発話前に 1 秒以上 4 秒未満の沈黙があるとき
オーバラップ	会話を被せた話者に切替え
	会話の重複発生時
オーバラップ戻し	会話を被せられた話者に切替え
	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替え
	4 秒以上の沈黙時間あるとき

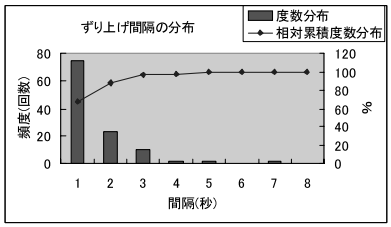


図 5 映像のずり上げスイッチングの度数分布と相対累積度数分布
Fig. 5 Frequency distribution and cumulative frequency distribution of vision shift switching.

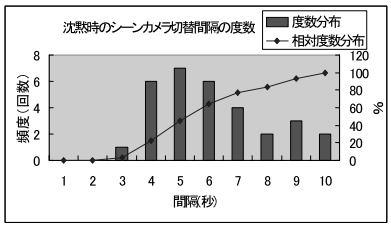


図 6 沈黙時のシーンカメラスイッチングの度数分布と相対累積度数分布
Fig. 6 Frequency distribution and cumulative frequency distribution of scene camera switching.

4. 実　　装

提案手法に基づいて，会議を自動撮影するプロ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 7** に示す．縦 5 m × 横 10 m 程度のスペースに 4 台の固定カメラを配置した．各カメラの役割を**表 2** に示す．カメラ 1〜3 は話者を映すカメラとして，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．

4.2 編集システム

編集システムの構成図を**図 8** に示す．表 2 の各カ

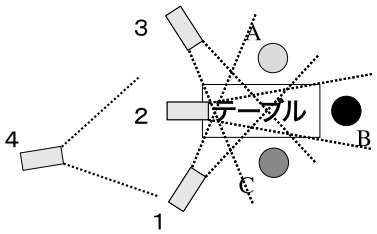


図 7 会議空間のレイアウト
Fig. 7 Meeting space layout.

表 2 各カメラのショット
Table 2 Each camera's shot.

Camera	shot
1	A
2	B
3	C
4	ABC

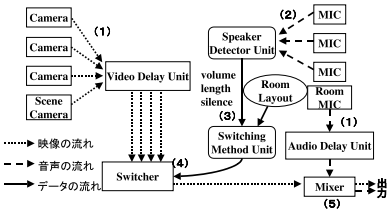


図 8 プロトタイプの構成図
Fig. 8 Structure of prototype system.

メラは，編集システムによってスイッチングが行われ，選択されたカメラの映像が選出される．

(1) それぞれのカメラ（Camera，Scene Camera）で撮影されたアナログ映像の出力は，映像遅延装置（Video Delay Unit）に入力され，Δt 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（Room MIC）で取得したアナログ音声の出力もまた，音声遅延装置（Audio Delay Unit）に入力され，Δt 秒間蓄積される．

(2) この Δt 秒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（MIC）に向かって話すことでシステムに認識させる（Speaker Detector Unit）．1 秒間に 4,000 回のサンプリングを行い 8 bit で量子化し，閾値以上の入力が連続した時点で話者と特定する．

(3) Δt 秒間で検出した音声の強弱（volume），会話の長短（length），沈黙時間（silence）の発話状況と会議空間のレイアウト（Room Layout）を基にどのカメラを用いるかを決定する（Switching Method Unit）．

(4) 遅延させた映像はスイッチャ（Switcher）へ入力される．このスイッチャを制御し，(3) で決定したカメラの映像を選択し出力する．

(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（Mixer）．

4.3 実装環境

ソフトウェア環境に関しては，映像遅延部，音声遅延部，音声認識部，会議空間のレイアウト情報，スイッチング方法決定部，スイッチャ，ミキサの各モジュールは J2SDK1.4 と JMF2.1.1e API の Java 言語で実装した．ハードウェア環境に関しては，カメラは Canon 社製 VC-CI，Sony 社製 DCR-VX2000 を，マイクは Elecom 社製 Multimedia Earphone with Microphone（MS-HS59SC）を使用した．各ソフトウェアを実行する計算機として DELL 社製 DIMENSION8300（CPU: Pentium4 1.5 GHz，OS: WindowsXP Professional）の PC を使用した．

4.4 プロトタイプの実装画面

図 9 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 2 種類のマーカが表示される．上のマーカは，提案手法によって選択された現時刻のカメラの映像であることを示す．一方，下のマーカは，発話と同時に話者へ映像をスイッチングする発話自動切替え（以下，既存手法と呼ぶ）によって選択されたカメラの映像であることを意味する．また，そのマーカの下には，左は既存手法によって，右は提案手法によって 1 本に編集された映像が表示される．その映像の下にはそれぞれスイッチングごとに切替えの種類が表示される．既存手法のスイッチングの種類は，発話スイッチング，オーバラップスイッチング，オーバラップ戻しスイッチングの 3 つのスイッチングである．一方，提案手法のスイッチングの種類は，この 3 つに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 つを加えたものである．提案手法では 2 秒以内の発話は，あいづちや瞬間的なノイズなど会話における意味の含有率が低いものと見なしフィルタリングされる．

図 9 実装画面
Fig. 9 Display of the implementation.

スイッチャが読み込む設定ファイルの内容を変更することで，参加人数，カメラの台数，スイッチングの種類，ずり上げ時間・沈黙時間，意図的な遅延時間 Δt，フィルタリングの閾値を調整できる．

4.5 スイッチング実行例

図 10 に，3 人での会議における提案手法と既存手法のスイッチングの様子を示す．会議の形態は調整会議であり，議題は卒業旅行の行先についてである．この会議のレイアウトは図 7 のとおりであり，会議の一部（47 秒間）の内容を**表 3** に示す．横軸に経過時間を示しており，タイムラインの上部の矢印が提案手法におけるスイッチングポイント，下部の矢印が既存手法におけるスイッチングポイントである．

5. 評価

本手法を用いてスイッチングした映像が，実際の録画番組でのずり上げスイッチングに近づいているか，視聴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．前者は定量評価を，後者は定性評価を行った．

5.1 定量評価

5.1.1 再現率，適合率

実際の録画番組の例として，演出されたスイッチングが多いドラマの 1 つである TRICK☆を取り上げた．そ

☆ テレビ朝日系列

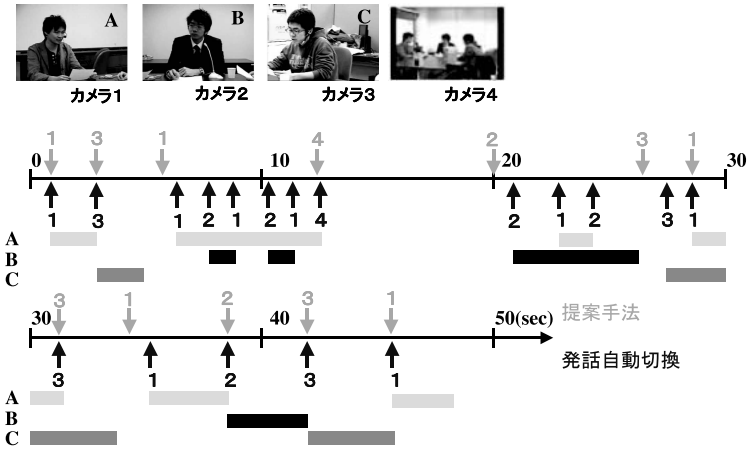


図 10 スイッチングポイント
Fig. 10 Switching points.

表 3 会話の内容
Table 3 The content of the conversations.

話し始めの話者	内容
A	そろそろ卒業旅行の先を決めないとね.（1〜2 秒）
C	イタリアとかどう思います？（3〜4 秒）
A	一年前にイタリアいってきたよ，歴史的建造物がたくさんあって（B：はい，はい（8 秒））面白かったけども，治安は悪かったし，（B：はい，はい（11 秒））2 回もいきたくはないな.（6〜12 秒）
	（沈黙）
B	俺は，昔ベルサイユの薔薇とかよく読んでいたから（A：うん，うん（23 秒）），一度はフランスには行きたいなと思うけど.（21〜25 秒）
C	そうだね，フランスなら，水に浮かんだ中世ヨーロッパの城（A：モンシャンミッシェルだね.（29〜31 秒））もあるし，なんせ凱旋門通ってパリの町を歩けるからね.（27〜33 秒）
A	そうしたら，フランス行った後にスイス行くのはどう？（35〜37 秒）
B	アルプスの少女ハイジの世界だしね.（38〜40 秒）
C	俺も，大自然に触れられるような旅行をしてみたかったので，賛成ですね.（41〜43 秒）
A	そうしたら，フランス行ってからスイス行くことにしよっか.（44〜47 秒）

の中から 1 対 1 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箇所は 57 カ所あった．この音声ファイルを基に提案手法で決定したずり上げ箇所と，実際のドラマの映像の各ずり上げ箇所の比較を行った．この一致具合を評価するために再現率 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箇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適合率 E を以下のように定義した．

- 再現率 P

$$P = \frac{N_m}{N_d} \times 100\,(\%)$$

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

- 適合率 E

$$E = \frac{N_m}{N_s} \times 100\,(\%)$$

（N_m：本システムの映像のずり上げ箇所がドラマのそれと一致した箇所の合計，
N_s：本システムの映像のずり上げ箇所の合計）

再現率が高くても，適合率が低ければ，一致した箇所は多いが無駄な映像のずり上げ箇所が多いことを意味する．逆に再現率が低く，適合率が高ければ，無駄な映像のずり上げ箇所は少なく，一致した箇所も少ないことを意味する．つまり，再現率，適合率ともに高い値のときほど，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．**図 11** に，今回の実験において実際のドラマ映像のと本システムのずり上げ箇所が「一致した」と判断する範囲を示す．図上部は映像の，下部は音声の移り変わりを示しており，ドラマでのずり上げ時刻を（t_d），人物 A の発話終了時刻を（t_1），人物 B の発話開始時刻を（t_2）と定義する．そして図 11 左のように，ドラマ映像において発話途中でスイッチングが発生した場合は，区間 $((t_d), (t_2))$ で本システムでもスイッチングが発生すればずり上げが「一

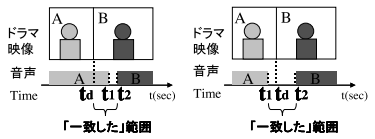


図 11 ドラマ映像と一致したずり上げ箇所の範囲
Fig. 11 The range of coincidence.

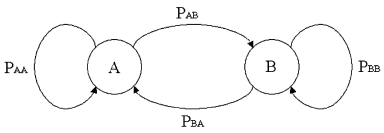


図 12 会話の状態遷移図
Fig. 12 State transition diagram of a conversation.

致した」と判断する．同様に図 11 右のように，ドラマ映像において二者の沈黙中にスイッチングが発生した場合は，区間 (t_1, t_2) でプロトタイプでもスイッチングが発生すれば「一致した」と判断する．

5.1.2 比較システム

本システムの有用性を検証するために，比較システムを用意した．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えシステムである．次の話者は，発話終了後に，**図 12** の状態遷移確率を用いて決定する．次の映像のずり上げ時刻の予測に関しては，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，計算機が切替え対象とずり上げ時刻を決定する．

5.1.3 実 験 結 果

表 4 に再現率，適合率の結果を示す．本システムでは適合率 64%，再現率 72%であった．一方，比較システムでは適合率 61%と本システムとあまり変わらないが，再現率が 29%と低くなった．つまり，本システムは比較システムよりも，多くの箇所でずり上げが一致した．本システムの適合率および再現率が 100%にならなかったのは，ドラマで 2 人の対話を撮影する場合では発話時に聞き手の映像が映っていたケース，2 人を映したケースがあり，ずり上げが必ずしも用いられていないからである．特に書類を渡しながら発話するなど動作とともに発話する場合は，2 人を映したケースが多かった．この場合，画像処理を用いシーンの状況を推測し，再現率と適合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た映像が，視聴者にどのような影響を与えるかを確かめるために映像の主観評価を行った．大学生の被験者 13 人に図 9 の実装画面に表示された映像と，比較システムを用いて制作した映像（ともに 3 分）を見てもらい，それぞれについてアンケートに 5 段階で評価してもらった．アンケートの質問項目を**表 5** に示す．映像の見映えを確認するため，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．

表 4 再現率と適合率の評価結果
Table 4 Results of the cover rate and effective rate.

評価項目＼システムの種類	本システム	比較システム
再現率 P	72% ($N_m = 41$)	29% ($N_m = 17$)
適合率 E	64% ($N_s = 64$)	61% ($N_s = 28$)

アンケートの結果を表 5 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 Wilcoxon の符号付き順位検定 p 値を求めた．

映像のスイッチングに関する質問の結果から，項目 1，2，3 に関しては危険率 0.1%以下，項目 7 に関しては危険率 1%以下，項目 11 に関しては危険率 5%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．タイムシフトによって，これまではできなかったずり上げスイッチング，シーンカメラスイッチングが可能になり，スムーズな映像の切替えが可能になったためと考えられる．一方，項目 11 が項目 1，2，3 ほど高くなかった理由として，比較システムでは短時間のスイッチング数が多く，変化に富んだ映像になっていたからである．しかし，項目 1，3 から，比較システムの映像は，被験者にとって切替えのタイミングが不適切であり，違和感を感じる映像であった．単純に切替え回数が多ければ良いということではないことが分かる．位置関係に関する質問の結果から，項目 5，6 に関しては危険率 1%以下で位置関係が明確であるという評価を得た．タイムシフトによって適切なタイミングでのシーンカメラスイッチングが可能となり，全体の位置関係の把握をしやすかったためと考えられる．一方で，項目 1，2，3，4 ほどは比較システムとの差が大きくなかった理由としては，会議参加者が 3 人と少なかったこと，被験者にあらかじめ参加者の座席配置を伝えていたことも影響している．参加者の映り具合に関する質問の結果から，項目 4 に関しては危険率 0.1%以下，項目 8，9 に関しては危険率 1%以

表 5 アンケートの結果
Table 5 Results of the questionnaire.

質問項目＼システムの種類	本システムの平均得点	比較システムの平均得点	Wilcoxon 符号付順位検定 P 値
1. 切替えのタイミングは適切だった	4.54	2.31	***0.00024
2. 見やすい映像だった	4.31	2.38	***0.00049
3. カメラの切替えに違和感を感じなかった	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	2.54	**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	*0.01563
12. だれとだれが会話しているかが分かった	3.77	3.15	0.05469

(n = 13, ***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$)

下，項目 10 に関しては危険率 5%以下で参加者の見映えが良かったという評価を得た．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．しかし，タイムシフトによって可能となったずり上げスイッチングの効果により，既存の手法よりも参加者に存在感を与えることができたため，良い評価を得ることができたと考えられる．以上の結果から，タイムシフトにより可能となった演出を用いることによって，既存の手法による演出よりも映像の見映えを良くできることを確認した．

また，アンケート項目以外のコメントとして「会議をドラマのように編集してよいものか」という意見があった．この指摘に関しては，会議の自動撮影における基本的な必要用件を満たしていればよいと考える．この用件は，以下の 2 つに分類される．

- だれが発言しているかを特定できること．
- ショットサイズが視聴者にとって見やすい構図になっていること．

今回導入した録画番組で用いられているスイッチングが，話者の特定を妨げるものはない．また，適切なショットサイズになるように，パニング，チルティングを利用しない固定カメラを用いて，つねに自然なショットサイズを用いている．よって，本システムはこの 2 つの用件を満たしているため，会議をドラマのように編集してもよいと考える．

6. おわりに

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．まず初めに複数台のカメラが撮影した映像・音声をメモリ上に Δt 秒間蓄積し，映像を選択する．選択には Δt 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 Δt 秒前の複数の映像・音声から 1 本の映像・音声を選択，切替え，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つね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が導入可能になった．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るという結果も得た．

今回の論文においては，映像・音声を蓄積することにより初めて可能となるずり上げスイッチング，シーンカメラスイッチングに重点を置き，リアルタイムでも可能であるずり下げショットに関しては考慮していない．しかし，映像の良い演出のため，システムとしての完成度を高めるためにはずり下げショットの導入も重要となる．今後は，このずり下げショットの自動判定を可能にするため，音声認識や高度な文脈の自動判断の技術を取り入れる必要がある．また，ずり上げスイッチング，シーンカメラスイッチング以外のタイムシフトによって可能となる演出手法（参加者中の 2 人の間でしばらく会話が続く場合，その 2 人だけを映すショットを選択するなど）も検討していく．

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) Pinhanez, C.S. and Bobick, A.F.: 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ッチング，映情学誌，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 (1996).
6) 村上昌史，大西正輝，福永邦雄：状況理解と映像評価を考慮した講義の知的自動撮影，情報処理学会研究報告，CVIM-125-5 (Jan. 2001).
7) 錦織修一郎，菅沼　明，谷口倫一郎：黒板講義を対象とした講義自動撮影システムの構築，信学技報，PRMU2000-212 (Mar. 2001).
8) Minoh, M. and Kameda, Y.: Image a 3d lecture room by interpreting its dynamic situation, *Proc. 4th Int. Workshop on Cooperative Distributed Vision*, pp.371–412 (Mar. 2001).
9) 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).
10) 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).
11) HV スキップバックディスクレコーダー. http://www.nhk.or.jp/pr/marukaji/m-giju109.html（2005 年 3 月 16 日現在）.
12) ドライブレコーダー. http://www.witness-jp.com/（2005 年 3 月 16 日現在）.

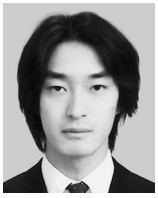
(平成 17 年 4 月 4 日受付)
(平成 17 年 12 月 2 日採録)

加藤　淳也

2003 年慶應義塾大学理工学部情報工学科卒業．2005 年同大学院前期博士課程修了．現在，NTT コミュニケーションズ株式会社に勤務．在学中，自動撮影技術の研究に従事．

住谷　哲夫（学生会員）

2004 年慶應義塾大学理工学部情報工学科卒業．現在，同大学院前期博士課程在学中．自動撮影技術，ウェアラブルコンピューティングの研究に従事.



井上　亮文（正会員）

1999 年慶應義塾大学理工学部計測工学科卒業．2001 年同大学院理工学研究科前期博士課程修了．2005 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助手．マルチメディアコンテンツ処理，ネットワークセキュリティの研究に従事.

重野　　寛（正会員）

1990 年慶應義塾大学理工学部計測工学科卒業．1997 年同大学院理工学研究科博士課程修了．1998 年同大学理工学部情報工学科助手（有期）．現在，同大学理工学部情報工学科助教授．工学博士．計算機ネットワーク・プロトコル，モバイル・コンピューティング，マルチメディア・アプリケーション等の研究に従事．著書『〜ネットワーク・ユーザのための〜無線 LAN 技術講座』（ソフト・リサーチ・センター），『コンピュータネットワーク』（オーム社）等．電子情報通信学会，IEEE，ACM 各会員.

岡田　謙一（フェロー）

慶應義塾大学理工学部情報工学科教授，工学博士．専門は，CSCW，グループウェア，HCI．情報処理学会誌編集主査，論文誌編集主査，GW 研究会主査等を歴任．現在，MBL 研究会運営委員，BCC 研究グループ幹事，日本 VR 学会 CS 研究会委員長．情報処理学会論文賞（1996 年，2001 年），情報処理学会 40 周年記念論文賞，日本 VR 学会サイバースペース研究賞，IEEE SAINT'04 最優秀論文賞を受賞．情報処理学会フェロー，IEEE，ACM，電子情報通信学会，人工知能学会各会員.