社団法人 情報処理学会　研究報告
IPSJ SIG Technical Report

2009－HCI－132（18）
2009／3／13

講義映像を対象としたルールベース編集手法の検討

吉高淳夫†，平嶋宗‡

†北陸先端科学技術大学院大学情報科学研究科　‡広島大学大学院工学研究科

講義を映像アーカイブとして整備することが大学等で行われているが，大学講義では映像の撮影，編集プロセスに十分な人的資源を配分することが困難であるため，固定カメラで連続的に撮影した映像をそのままアーカイブ化することがしばしば行われている．しかしながら，固定カメラによる講義映像撮影では注目対象物に関する十分な視覚情報を得にくい場合がり，また，視聴者の興味をひきつけることが難しいという問題がある．また，講師を撮影するのみならず，講義中に質問がある場合は講義映像に関する臨場感を高め，またノンバーバル情報を伝えるために質問者にカメラを向けてその様子を撮影することが望ましい．このような点を考慮し，本稿ではタグ付きマイクロホンの使用を前提としたルールベースの映像編集手法について検討する．

A Framework on Rule Based Video Editing for Lecture Video Archiving

Atsuo Yoshitaka†, Tsukasa Hirashima‡

†School of Information Science, Japan Advanced Institute of Science and Technology
‡Graduate School of Engineering, Hiroshima University

Lecture archiving is common especially in universities. It is not reasonable to assign persons for camera control and camera switching to all the lectures, since the number of lecture is too many to manage by university staff for lecture recording and editing manually. Currently, most popular way of lecture archiving is to place fixed cameras in a lecture room, and one of the cameras is used for recording throughout a lecture without active camera control and editing. A lecture video taken this way is neither enough to get visual information with reasonable resolution nor attractive to the viewer. In addition, we need to shoot a student when he/she is making a question so as to convey nonverbal information as well as improve presence. Considering these requirements, we propose a framework of rule-based video editing method for lecture video under the assumption of applying tagged microphones.

１．はじめに

大学等における視聴覚機器の整備が進み，また映像配信をするに耐える帯域を持つコンピュータネットワークの整備などにより，映像情報を蓄積し，配信することが可能な段階にきている．近年では講義受講者の利便性向上のため，また大学において多数開講される講義を大学の知的資産と捉えてそれをアーカイブ化する動きがある．しかしながら，講義を映像として記録するために必要なビデオカメラ等が全ての講義室に備えられている状況ではないことがまだ一般

的である．講義映像記録のためにビデオカメラやマイクロホンが備え付けられている講義室であっても，講義の記録に際しては講師の存在すると考えられる教壇周辺の領域を撮影できるようにカメラの向きや画角を事前に調整したカメラで撮影することがまだ一般的である．

図１　天井設置のカメラの例

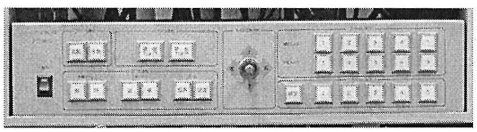
図２　カメラ操作インタフェースの例

図１は天井に設置されたビデオカメラの一例であり，このようなカメラはパン，チルト，ズーム制御することが可能なものもあるが，その操作は図２に示すようなインタフェースによりマニュアルで操作するのが一般的であり，講師等の状態に応じてカメラが自動的に制御されるものではない．講義撮影，編集のための要員を配置すれば講義中にパン，ズームあるいはカメラのスイッチングなどを適切に行い質の高い講義映像を制作することが可能である．しかしながら，大学等で実施される講義数は非常に多いため全ての講義に対して必要な人員を配置し，そのようなことを行うのは現実的ではない．

このような背景から，固定した単一のカメラで撮影する講義映像よりも質の高い映像の撮影，編集，生成を計算機処理により実現するための研究が行われている．講義における主な撮影対象は講師であることから，講師の検出・追跡や板書の検出等の撮影制御が多く研究されている．講義中は講師が内容の説明等を行うのみならず，受講学生が質問することもあるが，教室内の複数受講学生のうち，質問者を検出し撮影する課題に着目したものはほとんどない．それに対して，本稿では講義中に受講学生が質問することも前提とし，講義中の各時点において撮影されるべき対象を撮影ルールによって決定する手法について検討する．本稿で述べる撮影ルールによる映像編集手法では，これまでに提案したタグ付きマイクロホンシステムの使用を前提とする．タグ付きマイクロホンシステムは赤外線タグを取り付けたマイクロホンをパンチルトカメラで追跡撮影するものであり，話者検出のために事前に背景画像を登録したりする必要が無く，ポータビリティの高いシステムである．

以降２章では講義映像の撮影や生成についての関連研究について述べ，３章では本稿で前提とする装置，講義形態と撮影計画について述べる．４章では撮影計画を実現するためのルールベース撮影方式について述べ，最後に５章で本稿についてまとめる．

２．関連研究

映像コンテンツ生成は主に映像撮影と映像編集（映像のスイッチング）の２つのプロセスから成り立つ．映像撮影に関しては，

講師の撮影を対象にした研究が多く[1-4]，通常講師は教壇周辺に 1 人だけ存在するという前提をおいて差し支えないため，黒板等の背景領域を事前に獲得し，画像の差分を求めることにより講師や板書された文字の領域を抽出するものが多い[1-2]．講師以外の発話者も撮影の対象とする手法[5][6]は研究例が少なく，比較的少人数でのセミナー形式講義を想定し映像の差分により話者を検出するもの[5]や，マイクロホンアレイを使用して講義室中の発話者（質問者）の位置を特定しカメラ制御を行うもの[6]がある．前者は話者の特定を行ってはいないため，通常の 1 対多の講義形式には適応できず，また，マイクロホンアレイを使用する方式は設置や調整に手間がかかりポータビリティが低いと考えられ，さらには，拡声装置を使用できないという問題がある．

講師等の発話者の撮影に関しては，視聴者の講義映像への臨場感を向上させ，話者のノンバーバル情報を容易に得るためにパン，ズームなどのカメラ操作が適宜適用されることが好ましいといえる．外部からパン，チルト，ズーム制御が可能なカメラを用いる手法[1]やハイビジョン映像を使用して仮想的にカメラワーク映像を生成する手法[3]が提案されている．講師や黒板領域の撮影のみを対象とした場合はハイビジョンで撮影した映像から仮想的なカメラワーク映像を生成して提示する手法も有効だと考えられるが，質疑中の受講者をも対象とする場合，講義室の形態によっては上記手法の適用が困難な場合も考えられる．

これらの手法では，講師などの発話者は映像中に現れるオブジェクト（人物）であると見なし，映像の差分により検出しているものが多く，複数受講者の中から発話者（質問者）を検出することに適用することは難しい．また，マイクロホンアレイにより受講者中の発話者を検出する手法では，ポータビリティや拡声装置との併用に関して課題がある．

映像編集に関しては，視聴者に不必要なストレスを与えず，また講義に対する臨場感をより高めるために専門家の知見を編集上の制約として適用した手法が提案されているが，講義受講者の質疑における映像切り替えに関して検討している研究例は少ない．

提案手法では，タグ付きマイクロホンシステムの使用を前提とし，ポータビリティや拡声装置併用に関してより自由度の高いシステムでの受講者も含めた編集制御手法に関して検討を行う．

なお，映像操作を視聴者側に委ね，映像操作に関するミスマッチを解消する手法[7]や，映像，音声特徴によりインデクシング[8]することも考えられるが，本稿では撮影，編集操作がなされた映像を生成する視点から講義映像アーカイビングを捕らえる．

3．講義形態および撮影・編集目標

本稿では，タグ付きマイクロホンシステム[9]の使用を前提とし，ルール記述による映像編集制御を考える．講義の形態には様々なものが考えられるのため，まず前提とする講義形態，タグマイクロホンシステムについて述べ，目標とする撮影・編集操作について述べる．

3.1 講義形態

講義における教授内容の伝達は板書，OHP シートによる投影，プレゼンテーションソフトウェアにより編集した教材のビデオプ

ロジェクタによる投影等で与えられる視覚情報と，講師が発話により伝達する音声情報によってなされる．従来の視覚的な情報伝達手段は板書がほとんどであったが，近年大学の講義においてはプレゼンテーションソフトウェアを使用し，ビデオプロジェクタで教材を投影する形態が増加している．また，このようにして作成された教材はネットワーク経由で受講者に配布され，効果的な利用がなされていることが多い．将来的にもこの形態の講義資料提示・配布の割合が増加していくと考えられる．そのため，本稿では視覚的教材はプレゼンテーションソフトウェアによって作成され，それがビデオプロジェクタにより投影される形態を想定する．この想定下では視覚的に提示された教材は電子データとして別途取得可能であるので，講義中にそれを撮影対象に含めることは必須ではないといえる．

講義の進行に関しては，

(1)講師が一方向的に教授する形態

(2)講義中に受講者からの質問を適宜受けつつ進行させる形態

(3)割り当てられた受講者の発表や他の受講者の質疑をおりまぜて進行させる形態

などが考えられる．本稿では講師と受講生が説明や質疑を行うインタラクティブな講義進行を想定する．

講義中に使用するデバイスとしては，視覚的教材を提示するビデオプロジェクタのほか，講師，受講者共にワイヤレスマイクの使用を想定する．受講者が多く広い講義室で講義が行われる場合は拡声のためにワイヤレスマイクが使用されることが一般的であり，また講義記録の観点より確実に音声情報を獲得するためにもその使用は望ましい．

3.2 タグ付きマイクロホンシステム

タグ付きマイクロホンシステムは下部に赤外線タグを付加したワイヤレスマイクロホンをパン・チルトカメラ（光学的ズーム操作も含む）で検出，追跡撮影するシステムである[9]．図3にプロトタイプを示す．

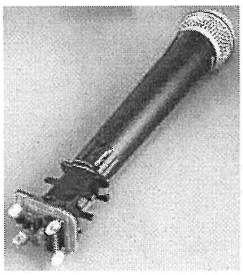

図3　タグ付きマイクロホン

赤外線タグの点滅パターンは PIC のプログラムにより制御され，その点滅パターンをパンチルトカメラで検出することによりマイクロホン，すなわち話者の位置を検出し，マイクロホンの位置が変化した際にはパン，チルト操作を行い追跡する．タグを検出して撮影するカメラモジュールを図4に示す．

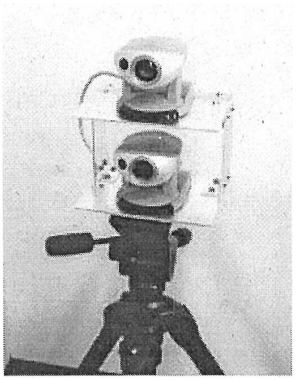

図4　カメラモジュール

カメラモジュールは光学ズーム操作が可能な２台のオートフォーカスパンチルトカメラから成り立ち，上が話者撮影用カメラ，下がタグ検出用カメラである．可視光線や

タグが発する波長以外の領域の赤外光の影響を抑制するために，タグ検出用カメラのレンズ前部には特定波長の赤外光のみを透過させるフィルタを装着している．

タグ検出用カメラでは差分処理と点滅パターン検出処理によりタグを検出し，タグが中央付近に配置されるようにカメラを制御する．その際２台のカメラのパン，チルトパラメタは同期制御されるためタグを映像中央付近に配置することにより話者撮影用カメラで話者をとらえた撮影が可能となる．また，オートフォーカスパラメタから話者までの距離を算出し，その距離に基づいて話者撮影用カメラのズームパラメタを設定し，講義室内の奥行き方向の距離に依存せず常にミドルショット（上半身が映像フレームに入るような撮影）での話者撮影を可能としている．

本稿では 1 本，あるいは 2 本のタグ付きマイクの使用を想定し，また，1 本のタグ付きマイクには１組以上のカメラモジュールが割り当てられることとする．また，タグ付きマイクの撮影用とは別に講義室全景撮影用のカメラの設置を想定する．

3.3 目標とする撮影・編集操作

インタラクティブな講義映像の撮影ならびに映像のスイッチングを実現するにあたって，話者の情報伝達，映像としての視聴しやすさ，臨場感の向上のためには以下の撮影，編集条件を満たす必要があると考えられる．

(a)発話中の講師，受講者が存在する場合はその人物を撮影すること

(b)アーカイブされる映像は不要なカメラ操作が施されていない映像であること

(c)映像のスイッチングを適切に行うこと

（d）話者撮影時ののショットサイズが適切であること

(a)は発話中の人物を常に撮影対象とすることを意味する．板書や OHP により視覚教材が例示される場合はこれらを撮影することが必須となるが，本稿では視覚教材は電子データの形式で受講者に配布されることを想定しているため，話者の撮影を主とした映像をアーカイブのための出力とすればよいということになる．

(b)は講義映像の視聴時に視聴者へ不要なストレスを与えないという目的がある．例えば静止して発話している講師をパン，チルト，あるいはシェイク（カメラを上下左右に振りながら撮影する）操作によって撮影したすると，視聴者が不要なストレスを受けることになるため避けられるべきである．これは言いかえれば，本稿で前提とするタグ付きマイクロホンシステムで発話中でないマイクロホン移動中のタグ追跡カメラの映像は出力とせず講義室全景などのスチル映像にスイッチすべきであるということを意味する．また，１つの発話区間が一定時間以上になる場合は，視聴者に冗長性を感じさせる一因になると考えられるため，他のスチル映像に適宜スイッチさせ，メリハリをつけることも必要であるといえる．

(d)は発話中の講師や受講者のノンバーバル情報をより把握しやすく伝達し，講義映像視聴者の講義に対する臨場感を高めるために必要であるといえる．撮影対象が講師のみである場合は，その移動範囲は限定的であるためズームパラメタの変化を伴わずに単に講師を追跡するのみでこの条件を満たせる場合が多い．しかし，撮影対象として講義室内の任意の場所に位置する発話受講者を含めた場合，カメラからの距離が受講

者毎に大きく異なることが一般的であると考えられるので，ズームパラメタを動的に設定することが必要となる．本稿ではタグ付きマイクロホンシステムを想定しているため，この条件が満たされた映像を得ることが可能である．

4．イベント - ルールに基づく編集

本章ではイベントと状態制約のルール記述による映像編集手法を検討する．ルールの評価によりアーカイブに出力する映像が決定される．

4.1 イベントと制約表現

ここで，ソース映像を出力するとなる映像オブジェクトを V_1, V_2, ..,V_n とする．ソース映像オブジェクト V_i (i=1,,.n)は，イベント e_i と映像データ v_i の組

$V_i=(e_i, v_i)$

で表現される．ここでイベントは講義中映像 v_i に関連した発話の有無とする．すなわち，

e_i={utterance, silent}

となる．

ソースとなる映像オブジェクトに関する制約は C_i=(source, cond, pri)で表現される．source は制約評価の対象となる映像オブジェクトであり，source={ V_1, V_2, ..,V_n}である．また cond はその映像オブジェクトに関する制約であり，cond=(cm, cs, du}すなわち対象映像オブジェクトを撮影するカメラのパン，チルト，ズーム等の操作量 cm，現在選択されている映像か否か cs={selected, standby}，選択（あるいは非選択）状態になってからの経過時間 du で表す．また，pri は制約評価時の優先度(0≤ pri ≤1)とし，その値が大きいほど優先度が高く，また各々は重複した値をとることはないものとする．

4.2 ルールに基づく映像編集

映像オブジェクトに関するイベントとそれに伴う制約によりソース映像の編集，すなわちスイッチング処理を表現する．ここでは映像オブジェクトを V_{tag1}, V_{tag2}, V_{still} とし，tag1，tag2 の付いたマイクを追跡撮影する映像オブジェクトをそれぞれ V_{tag1}，V_{tag2}，講義室の全景撮影を行う映像オブジェクトを V_{still} とする．

ここで，制約評価とそれに伴う動作はイベント発生によって起動されるものとする．イベントの生起により評価される制約と，制約が充足された場合の映像スイッチング動作を

if(C_{tag1})e_{tag1}=utterance then select V_{tag1}

と表現する．上の例は tag1 のマイクが発話を検出し，かつ tag1 のマイクに関連付けられる映像オブジェクトの状態 C_{tag1} が成立する場合に出力映像を V_{tag1} に変更することを表す．3.3 で述べた制約のうち，例えば(a)～(c)の目標を満たす最も簡単なルール表現は，

if(V_1, (still,selected,<t_d), p1)e_{tag1}=utterance then select V_{tag1},

if(V_2, (still,selected,<t_d), p2)e_{tag2}=utterance then select V_{tag2}

if(V_3, (*,*,*), p3)e_{still}=* then select V_{still}

(p1>p2>p3 または p2>p1>p3)

となる．ここで“*”は対応部分の条件評価を真とする記述である．

5． まとめ

本稿では，講義映像を自動撮影，編集してアーカイブ化する際に必要となるルールベースの編集手法について検討した．今後提案手法により生成される映像が従来の固定カメラによる映像，あるいはカメラマン，編集者によって生成される映像と比較

してどの程度の品質が得られるかを実験・評価する必要がある．ここでは提示教材は電子データの形式で配布，共有されることを前提とし，それを撮影する必要はないという前提を置いた．しかしながら，プロジェクタに投影した資料が指示された場合にはスクリーンを撮影することが好ましい場合もあると考えられる．これに関してはスクリーンへの指示を発話と同様なイベント発生ととらえることにより対応可能だと考えられる．また，各々の時点で１つの映像ソースを選択してアーカイブへの出力とするのではなく，イベントや制約充足状況をインデクスとして映像に付与するにとどめ，それを参照しながら視聴者に映像を選択してもらう形式との比較評価についても今後検討していきたい．

謝辞

本研究の一部は科学研究費補助金（基盤（C））による助成を受けた．ここに記して謝意を表す．

参考文献

[1] 大西正輝，松本昌紀，福永邦雄，“動画像処理による遠隔講義映像の自動生成とその評価，” 電子情報通信学会技術研究報告（ET），Vol.98，No.310，pp. 9-16，1998.

[2] 大西正輝，村上昌史，福永邦雄，“状況理解と映像評価に基づく講義の知的自動撮影，” 電子情報通信学会論文誌，Vol.J85-D-II，No.4，pp. 594-603 ，2002.

[3] 横井隆雄，藤吉弘亘，“高解像度映像からの自動講義ビデオ生成 - 仮想カメラワークの実現 - ，”画像センシングシンポジウム論文集，2005.

[4] 丸谷宜史，杉本吉隆，角所考，美濃導彦，“講師行動の統計的性質に基づいた講義撮影のための講義状況の認識，” 電子情報通信学会論文誌，Vol.J90-D，No.10，pp. 2775-2786，2007.

[5] 宮下剛，品川高廣，吉澤康文，“アクティブカメラ間の協調による知的自動撮影システム，” 第 7 回プログラミングおよび応用のシステムに関するワークショップ論文集，pp. 144-151，2004.

[6] 西口敏司，村上正行，亀田能成，角所考，美濃導彦，“受講者撮影機能を持つ双方向コミュニケーション記録型講義自動アーカイブシステム，日本知能情報ファジィ学会誌，Vol.17，No. 5，pp. 587-598，2005..

[7] 田代直之，島田敬士，菅沼明，“遠隔講義受講者のためのアクティブな講義映像生成システムの開発，” 第 47 回プログラミング・シンポジウム プログラム，pp. 203-208，2006.

[8] 石塚健太郎，亀田能成，美濃導彦，“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（NLC），Vol.99，No.707，pp. 91-98，2000.

[9] 川野晃寛，吉高淳夫，平嶋宗，”赤外線タグ付きマイクロホンを用いた講義における発話者の追跡撮影”，情報処理学会研究報告，2008-HCI-127，pp. 105-112，2008.