

Title	講義映像を対象とした発話受講者の自動撮影と編集制御に関する研究
Author(s)	吉高, 淳夫
Citation	科学研究費補助金研究成果報告書: 1-4
Issue Date	2009-06-10
Type	Research Paper
Text version	publisher
URL	http://hdl.handle.net/10119/8462
Rights	
Description	研究種目: 基盤研究 (C), 研究期間: 2007 ~ 2008, 課題番号: 19500804, 研究者番号: 60263729, 研究分野: 総合領域, 科研費の分科・細目: 科学教育・教育工学

平成 21 年 6 月 10 日現在

研究種目：基盤研究（C）
 研究期間：2007～2008
 課題番号：19500804
 研究課題名（和文） 講義映像を対象とした発話受講者の自動撮影と編集制御に関する研究
 研究課題名（英文） Research of speaker tracking and editing control for lecture video authoring
 研究代表者
 吉高 淳夫（YOSHITAKA ATSUO）
 北陸先端科学技術大学院大学・情報科学研究科・准教授
 研究者番号：60263729

研究成果の概要：

本研究ではマイクロホンに赤外線タグを装着することにより発話者の位置を検出し映像の適時切り換えを実現する事を目標とした。提案手法ではマイクロホンの位置変化特性やフォーカスパラメタから発話時と移動時を識別することにより早期に発話状態を検出し、一定のショットサイズで被写体を撮影する手法を検討・実装した。編集制御に関しては、イベントー制約記述による編集ルール駆動のカメラ制御により、個々の講義環境やインタラクション特性に適応可能な手法を提案した。

交付額

(金額単位：円)

	直接経費	間接経費	合計
2007年度	1,600,000	480,000	2,080,000
2008年度	1,300,000	390,000	1,690,000
年度			
年度			
年度			
総計	2,900,000	870,000	3,770,000

研究分野：総合領域

科研費の分科・細目：科学教育・教育工学

キーワード：マルチメディアと教育，自動撮影，対象物追跡

1. 研究開始当初の背景

大学などにおいて講義の内容を撮影し、Web上に公開する講義映像のアーカイブ化が盛んになってきている。講義を記録する際に必要となる情報としては「講師が発する情報」、「受講者が発する情報」、また電子スライドや板書内容などの「教材」の3つが挙げられる。電子スライドは元々電子データとし

て記録されており、また、板書内容は電子ホワイトボードなど板書内容を電子データとして記録するものがある。このように、電子スライドや板書内容などの教材に関しては撮影して記録する必要は必ずしもないと考えられる。ここでは、講師と受講者が発する情報を映像として記録することに着目する。講義を撮影するための方法の一つとして、固

定カメラを用いた方法が挙げられる。しかし、固定カメラでは撮影する範囲が固定されており、広範囲を撮影するためには人物の写る大きさであるショットサイズが小さくなる。そのため、講師や受講者の表情などのノンバーバル情報が伝わらず、講義の様子を捉えるには不十分な映像となってしまう。講師や受講者を表情が伝わる程度のショットサイズであるミドルショット（人物の上半身が写っている程度の大きさ）で撮影することで、講師や受講者の様子を詳細に捉えることのできる質の高い映像を撮影することができる。そのような映像を記録するために大学などで毎日行われている多くの講義に撮影者を配置するのは、多大な人的コストがかかるため現実的でない。そこで、人的コストを抑えながらも、講義の様子が十分に伝わるような質の高い映像を撮影するため、講義の撮影を自動化する手法の確立が求められている。

2. 研究の目的

講師の撮影を対象とした研究では、複数カメラを用い、画像処理により講師と板書内容を検出してそれらの自動撮影を行っているもの、講師に超音波ビーコンを装着することで、講師の位置を検出しパンチルトカメラによって追跡撮影を行っている手法等が既に報告されている。それに対し、受講者の撮影は講師の撮影に比べて困難であることが言える。なぜなら、受講者は講義室に複数存在するため、受講者全員にビーコン等を取り付けるといったことは現実的でない。また、固定カメラで受講者全体を撮影した場合、撮影範囲が広いミドルショットで撮影することができない。テレビ会議における映像の印象評価を行った調査では、人物の拡大機能を付けた方が発言者の表情が伝わり、映像としての評価が高いという結果が得られている。このことから、受講者の撮影を行う際には、複数の受講者の中から撮影対象を特定し、適切なズーム操作を行うことで表情などのノンバーバル情報も記録することが重要であると考えられ、それを十分に捉えることができるミドルショットで撮影する必要があると言える。

受講者の撮影を対象とした既存手法には、講義室の天井に複数のマイクロホンを設置し、マイクロホンに到達する音声の位相差により発言中の受講者の位置を検出し、撮影制御する手法が提案されている。この手法では、システムのポータビリティやキャリブレーションの手間、さらにはPAシステムとの併用が原理的にできないといった課題がある。それに対して本研究では、不特定多数の受講者の中から発言中の受講者の位置を検出し、従来手法の問題点である設置コストを抑えつつ、発言者の表情を捉えることのできるミ

ドルショットでの質の高い映像を撮影し、複数カメラでの撮影環境を想定した場合の映像切り換え、あるいは編集手法を検討、実装することを目的とする。

3. 研究の方法

提案手法では、講義中において受講者はマイクロホンを持って発言するという状況を想定する。マイクロホンにはマイクロプロセッサ制御により一定パターンでの点滅をする赤外線タグを取り付ける。タグの検出には、赤外線タグの波長領域周辺のみ透過するフィルタを装着したパンチルトカメラと話者撮影用の可視光カメラを1組とし、パン、チルトパラメータはシリアルデータとして送信し、両者を同期制御することとした。上記カメラモジュールにより赤外線タグの位置を検出し、発言中の受講者の位置の特定・撮影を行う。この手法により、発話人物の追跡、撮影という問題を他との区別が容易な赤外線タグの追跡、撮影という問題に置き換えることが可能なため、外乱に強く、安定性を高められることが期待できる。また、複数存在する受講者の中から発言者を特定することは通常困難であるが、提案手法では、発言時には赤外線タグの付いたマイクロホンを用いるという制約を設けることによって、複数存在する受講者の中から発言者を確実に特定することが可能となる。

ノンバーバル情報を的確に伝達するため、また、講義映像得の臨場感を高めるためには発話者をミドルショットで撮影することが必要である。これに対しては、カメラのフォーカスパラメータにより被写体の測距とそれに応じたズーム操作を行い、ミドルショットで発言者を撮影する手法を考察した。フォーカスパラメータとは、AFカメラが合焦した際の焦点距離に応じた値であり、システム構築時に採用したパンチルトカメラでは、被写体との距離が短いと大きい値となり、距離が長いと小さい値を返す。このフォーカス値によりカメラから被写体までの距離を換算し、距離に応じたズーム制御を行い、ミドルショットで発言者を撮影する。また、マイクロホン移動中とマイクロホン使用中の違い、すなわち発話中のタグ移動特性の違いからこれらの状態を区別し、発話の検出よりも早期に発話状態への移行を検出して適切なズーム制御を行うこととし、映像切り換え時の余裕時間をより多く確保することを考慮した。

講師と受講者の間にインタラクションがある講義における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とすると、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。これは言い換えれば、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の意味する。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、この条件を満たされた映像を得ることが可能である。

これらの要件を満たし、かつ個々の講義運営に依存したインタラクティブ性や講義形態に映像切り換えを適応できるように、カメラやタグ付きマイクロホンからのイベントによって駆動される編集ルールを記述し、それにより映像切り換えに関する柔軟性を持ち合わせた手法を考察した。

4. 研究成果

本研究において構築した赤外線タグ付きマイクロホンならびにパンチルトカメラモ

ジュールをそれぞれ図 1、図 2 に示す。



図 1 赤外線タグ付きマイクロホン



図 2 パンチルトカメラモジュール

マイクロホン底部には表面処理した 4 個の赤外 LED と定電流回路、点滅制御回路からなる赤外線タグが取り付けられている。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し、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。

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た。

複数カメラを使用し、各々の時点で適切な映像に切り換えるために各々の映像オブジェクトに関連付けられるイベントを発話の有無とし、その映像オブジェクトに関する制約をPTZ（パン、チルト、ズーム）パラメタ、映像の選択／非選択、選択／非選択状態になってからの経過時間ならびに制約評価優先度からなる記述により映像切り換えルールを記述し、映像切り換え処理を固定化するのではなく、先に挙げた要求を満たしつつ、講義のインタラクション特性や実施形態に応じた映像切り換えを実現する手法を提案した。

実装したタグマイクロホンシステムの追跡特性やショットサイズ維持性能を検証したところ、タグマイクロホンの追跡性能に関しては、カメラからおおむね4 m以上の距離があれば真横に横切る歩行動作であっても追跡可能であり、それより遠方であれば走る動作中でも追跡可能であった。ショットサイズ調整に要するレスポンス時間に関しては、フォーカスパラメタによる測距精度を上げるために発話状態移行直後にズームインした上でフォーカスパラメタを求めるアルゴリズムとしたために、ズームパラメタ調整プロセスの完了が発話開始以降になってしまうケースも見られた。これはプロトタイプ構築時に使用したパンチルトカメラの動作性能に依るものである。この問題の改善のためにはズーム機構の改善やフォーカス制御アルゴリズムの改良、あるいは顔領域抽出等の画像処理に基づくズームパラメタ制御の検討などが考えられ、これらの性能比較評価によりより適切な制御方式を明らかにすることなどが今後の課題として挙げられる。

5. 主な発表論文等

〔雑誌論文〕（計3件）

（1）

吉高淳夫, 平嶋宗, “講義映像を対象としたルールベース編集手法の検討,” 情報処理学会研究報告, 2009-HCI-132, pp. 123-129, 2009, 査読無

（2）

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. 105-111, 2009, 査読無

（3）

Atsuo Yoshitaka, Akihiro Kawano, and Tsukasa Hirashima, “Speaker Tracking for Automated Lecture Archiving using Tagged Microphone,” Proc. Tenth IEEE International Symposium on Multimedia, pp. 45-52, 2008, 査読有

〔学会発表〕（計3件）

（1）

吉高淳夫, 平嶋宗, “講義映像を対象としたルールベース編集手法の検討,” 情報処理学会 HCI 研究会
2009年3月13日, 神奈川県

（2）

Atsuo Yoshitaka, Akihiro Kawano, and Tsukasa Hirashima, “Speaker Tracking for Automated Lecture Archiving using Tagged Microphone,” Tenth IEEE International Symposium on Multimedia, 2008年12月17日, バークレー（アメリカ合衆国）

（3）

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 HCI 研究会
2008年2月1日, 広島県

6. 研究組織

(1) 研究代表者

吉高 淳夫 (YOSHITAKA ATSUO)
北陸先端科学技術大学院大学・
情報科学研究科・准教授
研究者番号：60263729

(2) 研究分担者

平嶋 宗 (HIRASHIMA TSUKASA)
広島大学・大学院工学研究科・教授
研究者番号：10238355

(3) 連携研究者