

教師の動作推定を利用した講義自動撮影システムの構築と評価

島田, 敬士
九州大学システム情報科学研究院知能システム学部門

菅沼, 明
九州大学システム情報科学研究院知能システム学部門

谷口, 倫一郎
九州大学システム情報科学研究院知能システム学部門

<https://hdl.handle.net/2324/5943>

出版情報 : 火の国情報シンポジウム, 2004-03
バージョン :
権利関係 :

教師の動作推定を利用した 講義自動撮影システムの構築と評価

島田 敬士, 菅沼 明, 谷口 倫一郎

九州大学大学院システム情報科学府

〒816-8580 春日市春日公園 6-1

E-mail: {atsushi, suga, rin}@limu.is.kyushu-u.ac.jp

あらまし 近年, 大学などの教育機関において遠隔講義が行われている. 我々の研究室では, 黒板とスクリーンを使用する講義を自動的に効率よく撮影するシステム ACE を構築している. 従来の ACE では, 教師が新しく板書した領域を板書終了後一定時間ズームして撮影するという手法を採っていた. しかし, その撮影方法は教師が過去に書いた板書を参照して説明する場合などに対処することができなかった. そこで, 本研究では, 画像処理により教師の動作を推定し, その推定結果に基づいて教師の説明対象を撮影する方法を考案した. 本稿では, 実際の講義を対象にして ACE で撮影した映像とカメラマンが撮影した映像との比較を, 学生のアンケートにより評価した. キーワード 遠隔講義支援, 画像処理, 動作推定, 講義撮影法

Development and Validation of an Automatic Camera Control System Based on an Estimation of Teacher's Behavior

Atsushi Shimada, Akira Suganuma and Rin-ichiro Taniguchi

Department of Intelligent Systems, Kyushu University

6-1 Kasuga-koen, Kasuga, 816-8580 Japan

E-mail: {atsushi, suga, rin}@limu.is.kyushu-u.ac.jp

Abstract The growth of communication network technology enables people to take part in a distant lecture. We are developing a supporting system for a distant lecture named ACE (Automatic Camera control system for Education). The previous version of ACE captures the lecture focusing on the latest object written on a blackboard because a teacher frequently explains it. However, a teacher does not always explain the latest object. We have designed, therefore, ACE to take a more suitable shot according to a teacher's behavior. This paper describes our system, our camera control strategy, the algorithm to estimate a teacher's behavior and our experimental result.

Key words Distant lecture, Image processing, Estimation of teacher's behavior, Camera control strategy

1. はじめに

現在, 様々な教育機関において遠隔講義が行われている. その講義の撮影には, カメラを固定したまま行うか, あるいは人間がカメラを操作して行うかのいずれかの方法が採られている. しかし, 前者は固定した領域だけを撮影するために十分な情報が得られず, 後者はコストがかかるという問題がある.

我々の研究室では, 講義を自動的に効率よく撮影す

るシステム ACE (Automatic Camera control system for Education) ^[1, 2] の開発を行っている. ACE は, 講義中において重要な場所を優先して撮影する. 講義中において重要な情報としては, 教師の身振り, 教師が書いた板書などの教師が学生に説明を行っている対象 (以下, 説明対象) が考えられる. 教師は, 説明をする際に, 内容の要点を板書し, それを指しながら, または身振りを加えながら説明する. そのため, これらは講義の映像において重要な情報である.

これまでに開発してきた ACE では、教師の説明対象を教師が新しく書いた板書領域と仮定して、最新の板書領域に重点を置いて撮影を行ってきた。そのため、講義中に教師が以前に板書した内容を説明するなど、仮定にそぐわない説明行動を採ると説明対象を撮影することができないなどの問題があった。これに対処するためには、教師の動作を検出し、動作意図の推定を行う必要があると考えられる。本研究では画像処理により教師の動作推定を行い、教師の説明対象を抽出する方法を考案した。

2. 講義自動撮影システム ACE

2.1 ACE の概要

ACE が想定している遠隔講義の形態は、一つの講義室で教師が講義を行い、カメラでその風景を撮影し、その講義室から空間的に離れた場所に存在する複数の講義室で学生が映像と音声によって講義を受講する一対多の形式 (図 1) である。ACE は、黒板のみを用いた講義、あるいは黒板とスクリーンの両方を用いた講義を対象としている。

ACE は講義室内に設置した固定カメラから得られる映像を動画画像処理し、教師の動作や教師が説明を行っている対象を解析する。その後、講義の状況に適した映像を撮影するために、適宜首振りカメラを制御して、講義を自動的に撮影する。

2.2 ACE の構成

図 2 に ACE の構成を示す。ACE は 2 台のカメラを必要とする。1 台は画像処理用の画像を撮影するための固定カメラで、もう 1 台は遠隔地に送信する講義映像を撮影するための首振りカメラである。教室の前方の黒板とスクリーン全体が収まるように固定カメラで撮影し、画像処理を行う計算機 (以下、画像処理 PC) は IEEE 1394 経由でその画像を取得する。画像処理 PC では、画像の解析を行い、講義状況を認識する。その結果に応じて、首振りカメラを制御する。首振りカメラにより撮影された講義映像は DVTS (Digital Video Transport System)^[3]を用いて遠隔地へ送信される。

2.3 ACE の撮影対象

講義の映像を見る学生が複数存在する場合、それぞれの学生が見たいと思う箇所は様々である。そのため ACE では、多くの学生が注目していると考えられる箇所を撮影する。

実際の講義において、学生は教師が説明している箇所に注目していることが多い。教師が黒板上に文

字や図形などを書いているときは、学生は板書に注目し、ノートなどに書き写す。

教師が板書を終え、学生に板書内容について説明をするときは、学生は教師の説明対象に注目している。教師の説明対象になる可能性が高いのは、教師が最も新しく書き加えた板書 (最新の板書) であるが、教師が長い間続けて板書したときや、過去に書いた板書を参照したときなどは、必ずしも最新の板書が説明対象になるとは限らない。

また、黒板とスクリーンの両方を用いた講義形式の場合、教師が説明を行っている対象は黒板上に書かれた板書に加えて、スクリーン上に投影されたスライドなどの文字や図形である可能性もある。

従って、学生の注目箇所は、教師の動作に応じて変化する。そこで、ACE では教師の動作に応じた教師の説明対象を撮影の対象とする。

2.4 撮影対象の撮影方法

以前の ACE では、最新の板書に重点を置き、教師の板書終了後、最新の板書領域を一定時間ズームするというカメラワークを繰り返していた。しかし、教師が一度に書く板書の量が多いときなどは頻繁にズームイン、ズームアウトが起こって、長時間映像を見続けると目が疲れてしまうという問題があった。

カメラマンが講義を撮影する場合、教師の動作を把握して、状況に応じた撮影対象を判断し撮影を行う。さらに、映像を見る学生の目にできるだけ負担をかけないように、滑らかにカメラをパン、チルトしたり、ズームイン、ズームアウトの頻度を調整したりしている。そこで、ACE はカメラマンの撮影方法を模倣して、教師の説明対象を撮影する。

3. 教師の動作モデルの作成および動作の識別手法の考案

3.1 講義中における教師の動作

ACE では、講義中における教師の動作の推定を行う。そのため、教師が講義中にどのような動作を行っているかを調べる必要がある。そこで、実際の講義をデジタルビデオカメラ (以下、DV カメラ) で撮影し、教師の動作を観察した。表 1 は実際に観察した講義映像である。

表 1 に示した講義映像から、講義中における教師の動作は、「板書中」、「説明中」、「移動中」の 3 種類に大きく分類できることがわかった。教師が黒板上に文字や図形などを書いているときの動作を「板書中」、板書した内容や、スクリーン上に投影されたス

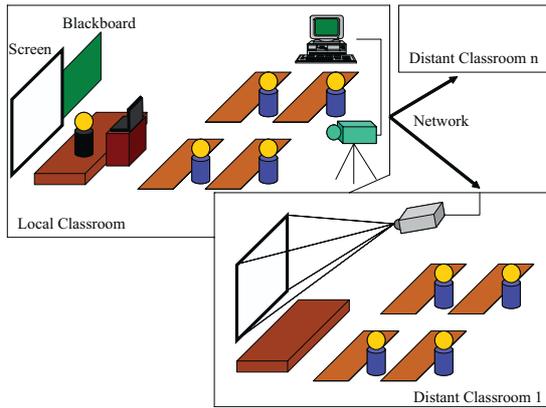


図 1: 本研究で想定する遠隔講義の形態

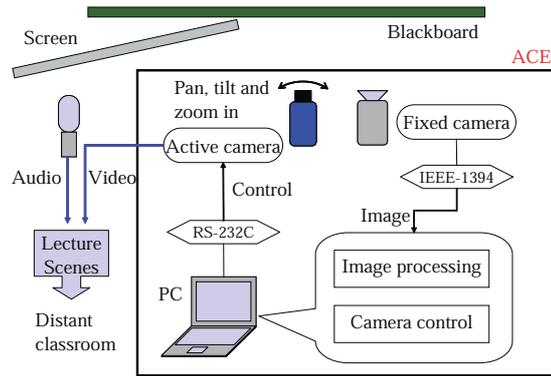


図 2: ACE の構成図

表 1: 観察した講義映像

	講義時間	講義環境
教師 A	40 分	黒板のみ
教師 B	60 分	黒板のみ
教師 C	25 分	黒板とスクリーン
教師 D	60 分	黒板とスクリーン

表 2: 講義中における各動作の割合

	板書中	説明中	移動中
教師 A	47.9%	44.6%	7.5%
教師 B	36.8%	54.3%	8.9%
教師 C	10.8%	85.9%	3.3%
教師 D	13.2%	81.7%	5.1%

ライドの内容を説明しているときの動作を「説明中」とした。これらの 2 種類の動作に当てはまらない動作（例えば、黒板の前を単に移動している動作やスライドを切り替えている動作など）を「移動中」とした。表 2 に講義中における各動作の割合を示す。

表からわかるように、黒板のみを用いた講義形式では「板書中」と「説明中」の動作の割合がほぼ同程度であるのに対して、黒板とスクリーンの両方を用いた講義形式では「説明中」の動作が講義の大半を占めている。これは、教師はスクリーン上に投影されたスライドの内容について主に説明を行っており、黒板は補助的に使っている場合が多いからであると考えられる。また、「移動中」の動作は他の動作に比べて極端に割合が低い。

3.2 教師の動作モデルの作成

ACE が教師の動作推定を行うためには、教師の動作モデルを作成する必要がある。そこで、先に述べた講義中における教師の「板書中」、「説明中」、「移動中」の 3 種類の動作モデルを作成することにした。

3.2.1 モデルの特徴点取得

我々は、教師の動作モデルを作成するために、教師の体の重心の位置、顔の位置、(利き)手の位置を教師 A の講義映像から $2fps$ の間隔で手作業により取得した。ここで、ある画像フレーム t における体の重

心の位置を $g(t) = (g_x(t), g_y(t))$ 、顔の位置を $f(t) = (f_x(t), f_y(t))$ 、手の位置を $h(t) = (h_x(t), h_y(t))$ とすると、フレーム t での特徴点は

$$I(t) = (g_x(t), g_y(t), f_x(t), f_y(t), h_x(t), h_y(t)) \quad (1)$$

で表すことができる。これらの特徴点を、板書動作について 1,463 点、説明動作について 1,248 点、移動動作について 857 点のサンプルを取得した。

3.2.2 特徴ベクトルの作成

得られた $I(t)$ を用いて 5 次元の特徴ベクトル、 $v = (v_1(t), v_2(t), v_3(t), v_4(t), v_5(t))^T$ を作成する。ここで、 v^T は v の転置を表す。以下に、各要素を示す。

$$v_1(t) = |f_x(t) - f_x(t-1)| \quad (2)$$

$$v_2(t) = |f_y(t) - h_y(t)| \quad (3)$$

$$v_3(t) = \sqrt{(f_x(t) - h_x(t))^2 + (f_y(t) - h_y(t))^2} \quad (4)$$

$$v_4(t) = |g_x(t) - h_x(t)| \quad (5)$$

$$v_5(t) = |g_y(t) - h_y(t)| \quad (6)$$

$v_1(t)$ は前フレームからの教師の顔の横方向への移動量、 $v_2(t)$ は教師の顔と手の高さの差を表している。一般的に、教師は板書をするとき顔の位置はそれほど動かさず、さらに顔と手の高さの差は、ほぼ一定であると考えられる。ゆえに、これらの特徴から $v_1(t)$ と

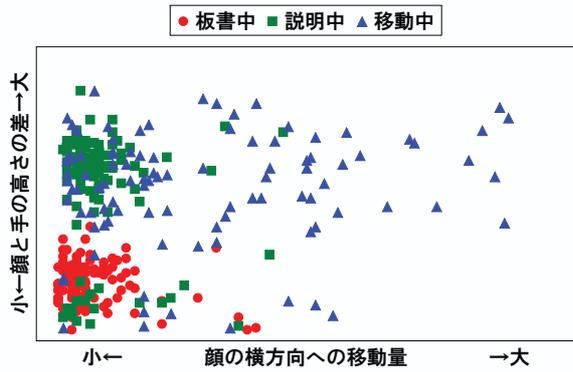


図 3: 部分空間 (v_1, v_2) における特徴ベクトル

$v_2(t)$ を用いることにした。一方、教師は学生に説明を行うとき、身振り手振りで内容を伝えようとする。その際、1 フレーム中の教師の腕は伸びていたり、曲がっていたりしている。従って、板書時に比べ説明時の教師の手、顔、体の重心の位置関係は、様々に変化すると考えられる。そこで、顔と手の距離 $v_3(t)$ 、体の重心と手の横方向の距離 $v_4(t)$ 、体の重心と手の縦方向の距離 $v_5(t)$ を用いた。

3.3 本研究で用いる識別器

図 3 は各動作の特徴ベクトル 300 個を部分空間 (v_1, v_2) に投影したものである。同図中の円形で表された点が板書中、矩形が説明中、三角形が移動中のときのベクトルをそれぞれ表している。

我々は、三つの動作を分類する方法として確率論的手法を用いることにした。図 3 のような分布は、一種の確率密度分布と見なすことができる。我々は、今回用意した各動作の特徴ベクトルの確率密度分布をガウス混合モデルを用いて近似することにした。ガウス混合モデルは、特徴ベクトルを $x = (x_1, \dots, x_d)^T$ (d は特徴量の個数) を用いて、

$$p(x) = \sum_{i=1}^n w_i g(x; \mu_i, \Sigma_i) \quad (7)$$

で表される。ここで n はガウスモデルの要素数である。また、 $g(x; \mu, \Sigma)$ はベクトル空間 x における多変量ガウス密度関数で、平均値ベクトル μ および共分散行列 Σ を用いて、

$$g(x; \mu, \Sigma) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \times \exp \left\{ -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right\}, \quad x \in \mathbf{R}^d \quad (8)$$

で表される^[4]。従って、モデルを生成するためにはパラメータ $\theta = \{w_i, \mu_i, \Sigma_i | i = 1, \dots, n\}$ を推定する必要がある。我々は、EM アルゴリズムにより混合分布モデルのパラメータ推定を行った。

ここでは、 W, E, M をそれぞれ「板書中」、「説明中」、「移動中」という動作に属する特徴ベクトルの集合としたときに、 $v \in W, v \in E, v \in M$ を満たす各特徴ベクトル v を用いて、3 種類の動作の混合モデルパラメータ $\theta_W, \theta_E, \theta_M$ を推定した。

4. ACE における画像処理

2 節で述べた撮影方法を実現するために、ACE は、画像から次の情報を取得する。

教師領域 教師を含む矩形領域

教師の動作 教師が行っている動作内容

板書領域 教師が黒板に板書した領域

説明対象領域 教師が説明をしている領域

これらの情報を取得するために ACE では、教師の動作推定および教師の説明対象の抽出を行う。以下、各処理について詳しく述べる。

4.1 教師領域の抽出

教師領域の抽出は以下の手順で行う。

(1) 背景差分による前景の抽出

講義開始前に教師が映っていない背景画像を撮影しておき、現在の講義映像と背景画像の差分を取り、前景を抽出する。

(2) ノイズを除去

前景画像には、教師以外に板書や光の影響によるノイズが含まれているため、収縮処理により、これらのノイズを除去する。

(3) 前景画素のヒストグラム作成

前景として残っている画素の分布を調べる。

(4) 教師領域抽出

ヒストグラムから閾値処理により教師領域を抽出する。

4.2 教師の動作推定

教師領域内から 3 節で述べた特徴点、すなわち教師の体の重心の位置、顔の位置、手の位置を取得する。

教師の体の重心は、教師領域内の前景画素の重心とする。

教師の顔、手の位置を抽出するために、まず教師領域内の肌色画素を抽出する。抽出された肌色画素は、顔、右手、左手の 3 つの部分から構成されると考え

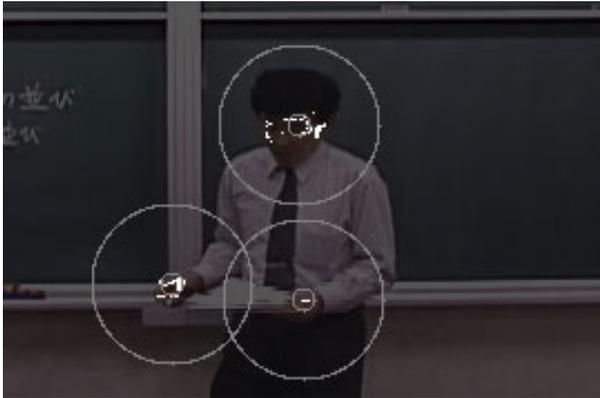


図 4: 肌色領域のクラスタリング

られる。しかし、教師が黒板側を向いている場合は顔を表している肌色画素の抽出ができなかったり、片方あるいは両方の手が体に隠れて見えないことがある。そこで、我々は抽出した肌色画素を最大で3つのクラスタに分類するようにした。図4に教師領域から抽出した肌色領域のクラスタリング結果を示す。図中の白円がクラスタを表している。これらのクラスタを教師の顔部分と手部分に分類しなければならない。そこで我々はまず教師の頭を表すクラスタを選び出すことにした。人の顔は丸い形状をしているという性質を用いて、教師領域内において円検出 Hough 変換を適用し、教師の顔の中心を抽出した。この中心点に最も近いクラスタを教師の顔部分とし、残りのクラスタを手の部分とする。

得られた特徴点から特徴ベクトル v を計算し、3種類の動作の混合モデルパラメータ θ_W (θ_E, θ_M) を用いて、式(7)により「板書中」「説明中」「移動中」である確率を求める。求めた確率の中で、最も高い値を示した動作を教師の動作推定の結果とする。

4.3 板書領域の抽出

黒板の文字は前景であるので、背景差分を行うことにより、板書領域を抽出できる。しかし、背景差分の手法で得られる前景には教師も含まれてしまう。そのため、黒板に書かれた文字だけを抽出するためには、教師の領域をマスクする必要がある。ACE では、以下の手順で板書領域を抽出している。

- (1) 背景差分を行い画像を二値化する
- (2) ノイズを除去する
- (3) 教師領域をマスクする
- (4) 残った前景の外接矩形を板書領域とする

ACE では抽出した板書領域に以下の二つを付加情報として記録しておく。

表 3: 撮影対象を決定するためのパラメータ

- 教師の動作
- 教師の位置
- 教師の説明対象
- 最後にズーム処理をした時刻
- 現在撮影している領域

- 板書が書かれた黒板上の位置
- 板書領域として抽出された時刻

4.4 教師の説明対象の抽出

(ア) 最新の板書領域とその周辺の領域

教師の動作が「板書中」または「説明中」と推定されたときは、板書が書かれた時刻情報を参照し、 t 秒前から現在までに書かれた板書領域を含む外接矩形領域を説明対象領域とする。

(イ) 過去の板書領域

教師の動作が「説明中」と推定され、教師の指差し動作が検出されたときは、教師の体の重心から手へのベクトル d が指す方向に、それまでに書かれた板書があるかどうか調べ、ベクトル d 方向の板書領域全てを囲む矩形を説明対象領域とする。また、ベクトルの方向に板書が存在しない場合は、最新の板書領域とその周辺領域を説明対象領域とする。

(ウ) スクリーン上の領域

教師がスクリーンの側に来て説明しているときは、スクリーン全体を教師の説明対象領域とする。

4.5 撮影対象の決定

以上で得た情報を基に撮影対象を決定する。撮影対象を決定するために用いるパラメータを表3に示す。

ACE では、基本的に教師の動作推定結果に基づいてズーム率を決定する。教師が「板書中」のときは、教師と説明対象(ア)がカメラに収まる程度のズーム率 Z_W 、教師の動作が「説明中」と推定され、黒板上を説明をしているときは、説明対象(イ)がカメラに収まる程度のズーム率 Z_E 、教師がスクリーン上を説明しているときは、説明対象(ウ)がカメラに収まる程度のズーム率 Z_S で撮影を行う。また、教師の動作が「移動中」と推定されたときは、ズーム率 $Z_M = 1.0$ として撮影を行う。ACE では $Z_W > Z_E$ とすることで、教師が板書しているときは、撮影対象を大きめに撮影するようにして、学生がノートを取りやすいようにしている。

しかし、教師の動作推定結果が頻繁に変化する場合、ズーム率も頻繁に変化するため、目が疲れやすい映像となってしまう。そこで、最後にズーム率を変化させた時刻 T_{zoom} を記録しておき、次にズーム率を変化させようとしたときに、 T_{zoom} から一定時間以上時間が経過している場合にズーム率を変化させる。ただし、次の撮影対象がスクリーンの場合は、無条件にズーム率を変化させる。上記の二つの場合以外ではズーム率を変化させない。

以上の条件から、ズーム率を変化させる場合は、そのズーム率で撮影対象を撮影する。一方、ズーム率を変化させない場合は、次に撮影の対象となっている範囲が、現在撮影している範囲内に含まれているかどうか判定する。もし、範囲内に含まれていない場合は、パン、チルトのみで、撮影対象にカメラを向ける。

5. 実験

本手法の有効性を確認するため、実際の講義を ACE で撮影し評価実験を行った。

5.1 実験環境

実験で使用したシステム構成は、図 2 に示した通りである。画像処理用の入力映像を撮影する固定カメラは、教室前方の黒板やスクリーンが画面に収まる程度の距離に配置した。講義撮影用のカメラは、固定カメラの横に配置した。固定カメラからの映像の 1 フレームあたりの画像サイズは 640×480 画素で、 $2fps$ で処理を行った。講義撮影用のカメラとして、首振り台に装着した DV カメラを用い、画像処理 PC から首振り台と DV カメラを制御した。

実際に九州大学工学部 1 年生を対象に行われているプログラミング演習の講義を、講義日以前に教師に講義を行ってもらい、ACE で撮影した映像 (ACE 映像) と、カメラマンが撮影した映像 (カメラマン映像) をそれぞれ DV テープに録画した。講義は、黒板とスクリーンの両方を用いて行う形式であり、1 台の DV カメラのみで講義映像の撮影を行う。また、カメラマンは、プロではなく学生に依頼した。普通の講義では、講義前にスクリーンに投影するスライド内容を資料として学生に配布していたが、本実験では、資料を配布せずに純粋にビデオ映像と音声のみで講義を受けてもらった。撮影した二つの映像を講義室においてプロジェクタでスクリーンに投影し、学生 58 人にビデオ講義として見てもらい評価を行った。また、以前の ACE の評価実験^[5]を行ったときに固定カメラで撮影した講義映像 (固定カメラ映像) を見てもらった。

そのときのアンケート結果も踏まえて評価を行った。
5.2 評価方法

ACE が撮影した講義映像およびカメラマンが撮影した講義映像それぞれを学生が見終わった後に、アンケートを行った。アンケートでは次の質問に答えてもらった。

- (1) 教師の様子がよく分かったか?
- (2) 黒板の文字や図形は見やすかったか?
- (3) 自分の見たい所がよく見えたか?
- (4) 講義の臨場感は得られたか?
- (5) 講義映像の総合的な評価はどうか?
- (6) 通常の講義と比べて理解できたか?
- (7) スクリーンの文字は見やすかったか?
- (8) カメラの動きは目にやさしかったか?
- (9) 撮影者 (ACE, カメラマン) の意図する撮影対象が分かったか?

アンケートの質問 (1)~(6) は以前の評価実験で答えてもらった質問と全く同じものである。質問 (7)~(9) は、今回の実験の設定に特化したものである。アンケートはそれぞれの質問に対して評価の尺度となる選択肢を五つ用意し、最高値 5、最低値 1 の 5 段階で評価してもらった。

5.3 固定カメラ映像、カメラマン映像、ACE 映像の比較

アンケートの質問 (1)~(6) に関して、三つの講義映像に対するアンケート評価値の平均値に差があるかどうか分散分析を行った。「3 標本は同一の母集団から得られた標本である」と帰無仮説を立て、仮説を棄却する危険率 (有意水準) を 0.05 (5%) とした。

表 4 に各質問に対する平均値、表 5 に分散分析結果を示す。表内では、固定カメラ映像の場合を A、カメラマン映像の場合を B、ACE 映像の場合を C としている。

質問 (1) に関して、分散分析により標本間の平均値に差があるとは言えないことが分かった。これは、データの平均値としては、カメラマンあるいは ACE が撮影した場合のほうが固定カメラで撮影した場合よりも数値上は良くなっているが、統計的には差があるとは言いきれないことを意味している。実際に、固定カメラで撮影した場合、教師は常にカメラに映っているため、教師を常に観察できる長所があるが、ズーム率が低いため教師の様子までは確認するのが困難であるという短所もある。一方、カメラマンや ACE が撮影した場合は、教師の様子は確認できるが、常に

表 4: 質問 (1)~(6) のアンケートの平均値

	質問 (1)	質問 (2)	質問 (3)	質問 (4)	質問 (5)	質問 (6)
A. 固定カメラ	2.60	1.45	1.53	2.47	1.81	1.85
B. カメラマン	2.90	2.19	2.66	2.71	2.22	2.55
C. ACE	2.93	2.09	2.41	2.34	1.97	2.17

表 5: 分散分析結果

	質問 (1)	質問 (2)	質問 (3)	質問 (4)	質問 (5)	質問 (6)
帰無仮説採否判定	保留	棄却	棄却	保留	棄却	棄却
A-B 間	——	有意差有	有意差有	——	有意差有	有意差有
B-C 間	——	有意差無	有意差無	——	有意差有	有意差無
C-A 間	——	有意差有	有意差有	——	有意差有	有意差無

教師がカメラに映っているわけではないことが問題となっていると思われる。

質問 (2) に関して、固定カメラ映像-カメラマン映像、固定カメラ映像-ACE 映像間の平均値間に有意差が認められるが、カメラマン映像-ACE 映像間の平均値には差があるとは言えないことが分かった。これは、カメラマンや ACE が撮影した映像は、教師が板書した部分をズームインして撮影しているため、固定カメラ映像よりも板書の文字や図形が大きく映り読み取りやすかったためと考えられる。

質問 (3) に関して、固定カメラ映像-カメラマン映像、固定カメラ映像-ACE 映像間の平均値間に有意差が認められるが、カメラマン映像-ACE 映像間の平均値には差があるとは言えないことが分かった。固定カメラ映像は黒板全体が映っているが、文字が小さいため教師が板書している箇所や説明している対象を把握することが難しい。そのため、評価が低かったと考えられる。一方、カメラマン映像や ACE 映像は、教師の説明対象を中心に撮影しているため固定カメラ映像に比べて評価が高かったと考えられる。しかし、今回の実験の場合、黒板側を撮影しているときはスクリーン側を撮影することができなく、その逆の場合もあるため自分の見たいところが見れなかったという意見もあった。また、配布資料等があればよかったという意見もあった。

質問 (4) に関して、分散分析により標本間の平均値に差があるとは言えないことが分かった。アンケートで評価値が 4 (大体よく得られた) 以上と答えた人の割合は、固定カメラ映像よりもカメラマン映像や ACE 映像のほうが高かった。しかし、臨場感があまり得られなかったと答えた人の割合も高く、大半はど

ちらとも言えないという答えであった。これは、ビデオ講義ということが一因と考えられる。さらに学生がいない教室で講義の撮影を行ったため、教師が一方的に講義を行う形となってしまったことも要因と考えられる。

質問 (5) に関して、全ての映像間の平均値間に有意差があることが分かった。固定カメラ映像の場合、常に同じアングルからの撮影であるため、不満という意見が多かった。ACE 映像の場合は、固定カメラの場合とは異なり、ズーム率を変化させたり、パン、チルトによりカメラを動かしたりしているため固定カメラ映像に比べて評価は高かったと考えられる。しかし、機材的な問題で、ズーム速度を制御できなかったり、パン、チルトの際に首振り台の振動でカメラが揺れて、その結果講義映像が多少ぶれて撮影されていたりしていたため、講義映像として不満が残るものであった。

質問 (6) に関して、固定カメラ映像-カメラマン映像間の平均値間に有意差があることが分かった。しかし、同じプログラミング演習の講義であっても講義内容によって難易度が異なるため一概にはどちらが良いとは言えない。また、今回の実験ではスクリーンに投影するスライドの内容を講義前に学生に配布しなかったため理解ができなかったという意見が多かった。

5.4 カメラマン映像、ACE 映像の比較

次に、質問 (7)~(9) までのアンケート結果を示す。各質問に対する評価値の平均値間に差があるかどうか調べるために、t 検定を行った。t 検定を行う際、仮説を棄却する危険率 (有意水準) を $\alpha = 0.05(5\%)$ とし、「平均値間に差はない」という帰無仮説を立てた。表 6 に各質問に対する平均値、表 7 に t 検定の結

表 6: 質問 (7)~(9) のアンケートの平均値

	質問 (7)	質問 (8)	質問 (9)
B	2.71	3.28	3.48
C	2.62	2.53	2.91

表 7: t 検定結果

	質問 (7)	質問 (8)	質問 (9)
判定	有意差無	有意差有	有意差有

果を示す。

質問 (7) に関して、t 検定により標本間の平均値に差があるとは言えないことが分かった。今回の実験では、スクリーンもカメラで撮影を行ったが、カメラマンがスクリーン上を撮影するときも ACE がスクリーン上を撮影するときもカメラにスクリーン全体が収まる程度のズーム率で撮影したためアンケート結果にはさほど差異はなかった。しかし、実験ではスクリーン上に投影されたスライドを撮影した映像をさらにスクリーンに投影しているため、スクリーン上の文字が普段より見づらかったという意見が多かった。

質問 (8) に関して、t 検定により標本間の平均値に差があることが認められた。これは、質問 (5) の考察でも述べたように、ACE が撮影した映像は、パン、チルトの際に首振り台の振動がカメラに伝わるため、パン、チルトの度に映像が小刻みに揺れてしまうという問題が生じた。また、PC から DV カメラのズーム率を制御する際も、所定のズーム率までにかかる時間を調整することができないため、ズーム速度が速すぎて目が疲れてしまうという問題も生じた。

質問 (9) に関して、t 検定により標本間の平均値に差があることが認められた。ACE は、画像処理により、教師の動作推定を行い、その結果に基づいて説明対象を絞り込み撮影対象を決定している。一方、カメラマンは教師の動作を含め、今教師がどこを説明しているかを、視覚と聴覚を使って判断している。さらに、カメラマンは次に撮影すべき場所をある程度、経験により把握することができる。そのため、カメラマンは講義状況に応じて (例えば、この部分はもうしばらく映したほうが良いなど) 撮影対象を柔軟かつ的確に捕らえているが、ACE は、撮影対象を的確に捕らえたとしても、その対象を適切な時間映し続けることができなかつたため、評価結果に差がでたと考えられる。また、ACE が撮影対象を誤ってしまったときなどは、学生に撮影の意図が伝わらなかつたと考えられる。

6. おわりに

本稿では、講義中に教師が説明をしている対象を撮影するという撮影戦略を採った。画像処理により教師の動作を推定し、推定結果に基づいて教師の説明対象の抽出を行い撮影対象を決定した。ACE を実際の講義に適用した結果、総合的にカメラマンが撮影した映像と大差ない映像が撮影できていることが分かった。

今後の課題としては、

- 教師の発話内容から説明対象の抽出
- 学生と教師のリアルタイムでの応答
- 講義状況を理解した、より柔軟な撮影戦略

などが挙げられる。

謝辞

本研究の一部は、21 世紀 COE プログラム「システム情報科学での社会基盤システム形成」および科学研究費基盤研究 (C)(2) 課題番号 14580224 の補助を受けた。

参考文献

- [1] 錦織修一郎, 菅沼明, 谷口倫一郎, “黑板講義を対象とした講義自動撮影システム,” 電子情報通信学会, 信学技法, Vol.100, No.701, pp.79-86, 2001.
- [2] A.Suganuma and S.Nishigori, “Automatic Camera Control System for a Distant Lecture with Videoing a Normal Classroom,” Proc. World Conf. on Educational Multimedia, Hypermedia and Telecommunications, pp.1892-1897, 2002.
- [3] WIDE project, “DV Stream on IEEE1394 Encapsulated into IP,” <http://www.sfc.wide.ad.jp/DVTS/>
- [4] N.Johnson and D.Hogg, “Representation and synthesis of behavior using Gaussian mixtures,” Proc. Image and Vision Computing 20, pp.889-894, (2002)
- [5] 芦川 平, 菅沼明, 谷口倫一郎, “黑板講義におけるチョーク音検出を利用した講義自動撮影システムの構築と評価,” 火の国情報シンポジウム 2003, pp.60-67, 2003.