

## 講義中の教師の動作に基づく説明対象の抽出

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/2324/5945>

---

出版情報 : 画像の認識・理解シンポジウム, pp. 353-358, 2004-07  
バージョン :  
権利関係 :

## 講義中の教師の動作に基づく説明対象の抽出

島田 敬士<sup>†</sup> 菅沼 明<sup>††</sup> 谷口倫一郎<sup>††</sup>

<sup>†</sup> 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

<sup>††</sup> 九州大学大学院システム情報科学研究科 〒 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, 黒板のみを用いた講義を対象としており, 教師が新しく板書した領域を中心に撮影するという手法を採っていた. しかし, 近年では黒板とスクリーンの両方を用いた講義が増えてきており, 以前の撮影方法では教師の説明対象をうまく撮影することができなくなってきた. 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構築した.

キーワード 遠隔講義支援, 動作推定, 講義撮影法, 確率的手法

## Extraction of Explained Object

### Based on Estimation of Teacher's Behavior in a Lecture

Atsushi SHIMADA<sup>†</sup>, Akira SUGANUMA<sup>††</sup>, and Rin-ichoro TANIGUCHI<sup>††</sup>

<sup>†</sup> Department of Intelligent Systems, Kyushu University 6-1 Kasuga-koen, Kasuga, 816-8580 Japan

<sup>††</sup> 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 supports a lecture in which a teacher teaches his/her students by using only a blackboard. It captures the lecture focusing on the latest object written on a blackboard. However, a lecture in which a teacher uses both a blackboard and a screen has been increasing in recent years. The previous version of ACE cannot take a suitable shot in such a lecture style. 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 results.

**Key words** Distant lecture, Estimation of teacher's behavior, Camera control strategy, Stochastic approach

### 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. そこで, 講義を自動的に効率よく撮影するシステムが望まれる.

これまでに, 講義の自動撮影を目的として, 様々な研究が行われている. 例えば, 遠隔地へ配信する映像を自動的に切り替える手法 [1], [2] が提案されている. 配信するのに適した映像

をうまく選択できれば, 効果的な映像を学生に提供できるが, 固定カメラでは高画質の映像を提供することができなかったり, 複数台のカメラが必要なため, 遠隔講義用の機材が大掛かりなものになってしまう問題がある. 一方で, 1台のカメラに対し計算機でパン, チルト, ズーム制御を行い, 講義を自動撮影する手法 [3] の提案がなされている. 固定カメラ画像を解析し, 教師の動作を認識させ, その結果をカメラ制御に反映しているが, 教師が新しく書いた板書を中心に撮影するため, 教師が以前に書いた板書を説明している場合は, その説明対象をうまく撮影できていない. 我々の研究室でも, これまでに講義自動撮影システム ACE (Automatic Camera control system for

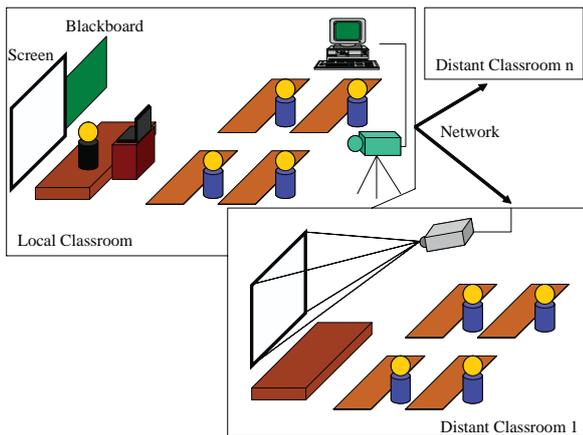


図1 本研究で想定する遠隔講義の形態

Education) [4], [5] の開発を行ってきた。これまでに開発してきた ACE は、黒板のみを用いた講義を対象としていた。黒板のみを用いた講義であれば、教師は板書した内容を説明するため、黒板上に新しく書かれた板書を中心に撮影する撮影戦略を採り、教師の説明対象を撮影していた。しかしながら、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。また、近年では黒板とスクリーンの両方を用いた講義が増えてきている。このような講義を想定した場合、以前のように最新の板書領域を撮影する撮影方法では教師の説明対象をうまく撮影することができなくなってきた。

カメラマンが講義を撮影する場合を考えると、カメラマンは、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、教師の身振りや、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。カメラマンは、講義状況に応じて、これらの撮影対象をうまく切り替えて撮影する。カメラマンが撮影対象を切り替えるタイミングは、教師の動作が切り替わったときであると考えられる。例えば、教師が板書をしているときは、教師が書いている板書と教師を中心に撮影する。教師が学生に説明を行っているときは、教師の説明対象を中心に撮影する。

そこで、カメラマンが撮影すると考えられる場所を撮影する方針で ACE の設計を行った。ACE が、カメラマンと同様の撮影対象を撮影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、画像処理により講義中における教師の動作を推定し、カメラマンが撮影する講義映像と同程度の映像を撮影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

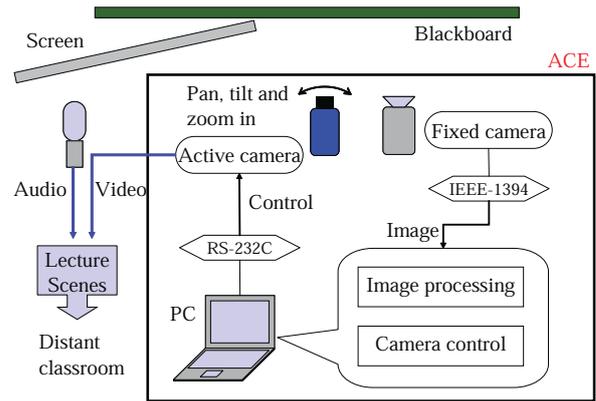


図2 ACE の構成図

固定カメラから得られる映像を動画像処理し、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) [6] 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。そのためには、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いて、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、どちらの講義形式においても「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

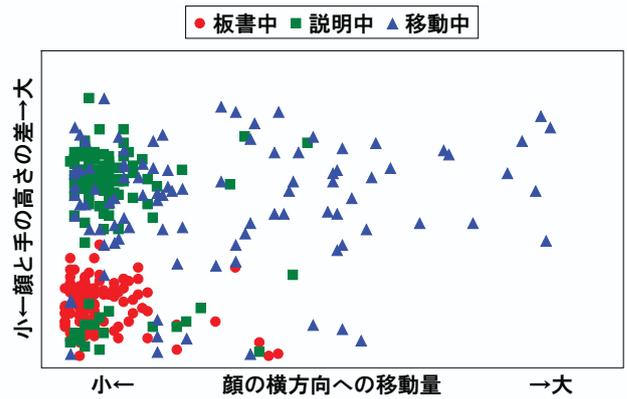


図3 部分空間 $(v_1, v_2)$ における特徴ベクトル

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 $v_2(t)$ 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 $v_5(t)$ を特徴ベクトルの要素とした。

#### 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、

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

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

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， $g(\mathbf{x}; \mu, \Sigma)$  はベクトル空間  $\mathbf{x}$  における多変量ガウス密度関数で，平均値ベクトル  $\mu$  および共分散行列  $\Sigma$  を用いて，

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

$$\mathbf{x} \in \mathbf{R}^d \quad (8)$$

で表される [7]．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．

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%

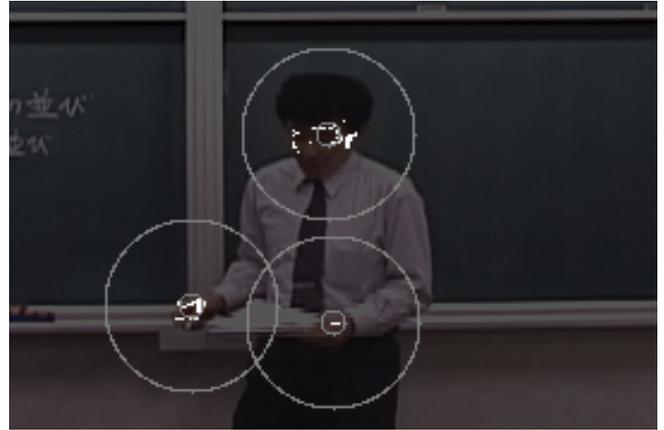


図 4 肌色領域のクラスタリング

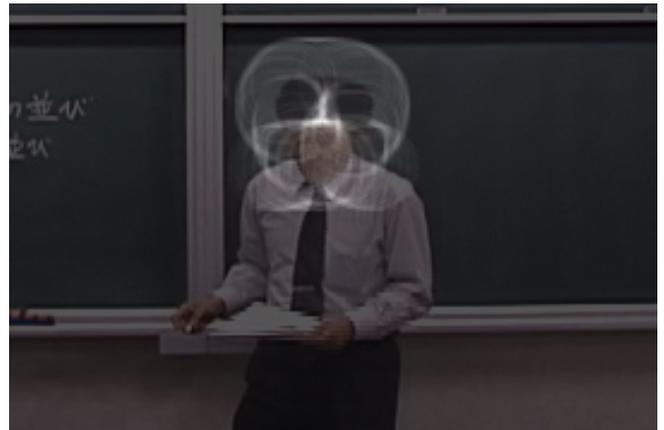


図 5 顔の中心点候補投票後の画像

#### 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 変換を適用する．

教師の体の輪郭をエッジ検出により抽出し、エッジ点からある半径  $r$  の距離にある画素をエッジ点を円周上に持つ、円の中心の候補として投票を行う。図 5 は、教師領域の上側半分の領域内でエッジ点から、ある半径  $r$  の円の中心の候補を画像平面上に投票した結果である。教師の顔領域の中心部に近づくにつれて多く投票が行われているのが確認できる。投票空間で最も投票数が多かった画素を円の中心すなわち教師の顔の中心として抽出した。この中心点に最も近いクラスを教師の顔部分とし、残りのクラスを手の部分とする。

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  $\theta_W$  ( $\theta_E, \theta_M$ ) 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。

- 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

教師領域、教師動作、板書領域から教師の説明対象を抽出する。

##### ・説明対象 1

教師の動作が「板書中」と推定されたときは、板書が書かれた時刻情報を参照し、 $t_W$  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。

##### ・説明対象 2

教師の動作が「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、 $t_E$  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。ここで、 $t_W < t_E$  とすることで、教師が説明を行っているときは板書時よりも広く説明対象領域を抽出している。

##### ・説明対象 3

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。指差し動作は、教師の体の重心と手の距離が閾値以上の場合としている。

##### ・説明対象 4

表 3 撮影対象を決定するためのパラメータ

<ul style="list-style-type: none"> <li>• 教師の動作</li> <li>• 教師の位置</li> <li>• 教師の説明対象</li> <li>• 最後にズーム処理をした時刻</li> <li>• 現在撮影している領域</li> </ul>
--

表 4 実験に用いた講義映像

	教師	講義時間	講義環境
講義映像 1	教師 A	20 分	黒板のみを使用
講義映像 2	教師 C	19 分	黒板のみを使用
講義映像 3	教師 C	25 分	黒板とスクリーンの両方を使用

教師の動作が「説明中」と推定され、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 4.4 節で述べた説明対象を撮影するためのズーム率を決定する。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

### 5.1 実験環境

あらかじめ DV カメラで撮影しておいた講義映像に対してオフラインで画像処理を施すことで実験を行う。表 4 に実験に用いた講義映像を示す。表中の教師は表 1 に示した教師と対応している。講義映像 1 に登場する教師は、教師の動作モデルを作成するときのモデルとなった教師である。映像の 1 フレームの画像サイズは  $720 \times 480$  画素で、画像処理は動作モデルの特徴点を取得したときと同様に  $2fps$  で行った。

### 5.2 実験結果

#### 5.2.1 特徴点抽出結果

表 5 は、ACE が、どれだけ正確に教師の特徴点すなわち、顔と手を抽出できているかを適合率と再現率により評価したものである。ここで、適合率、再現率はそれぞれ以下の式 (9), (10)

表 5 特徴点抽出結果

		適合率 (%)	再現率 (%)
講義映像 1	顔	94.3	96.0
	手	86.2	90.6
講義映像 2	顔	92.5	98.2
	手	89.7	93.1
講義映像 3	顔	93.9	97.5
	手	87.1	90.5

表 6 教師の動作推定結果 (単位:%)

	板書中		説明中		移動中	
	適合率	再現率	適合率	再現率	適合率	再現率
講義映像 1	81.5	79.2	78.3	76.8	20.5	30.1
講義映像 2	76.8	79.9	75.7	77.8	17.9	11.6
講義映像 3	77.9	72.4	80.4	81.7	16.5	13.2

表 7 撮影対象抽出結果

	成功率 (%)
講義映像 1	90.7
講義映像 2	94.3
講義映像 3	92.1

で計算されるものである。

$$\text{適合率} = \frac{\text{抽出した特徴点のうち正しい特徴点の総数}}{\text{システムが抽出した特徴点の総数}} \quad (9)$$

$$\text{再現率} = \frac{\text{抽出した特徴点のうち正しい特徴点の総数}}{\text{システムに抽出して欲しい特徴点の総数}} \quad (10)$$

第 4.2 節で述べたように、教師の手は体に隠れて片方しか見えない場合や両手ともに見えない場合がある。表 5 に示した結果では、教師の手が両方見えているのに、ACE が両手の位置を抽出できない場合や、片方の手しか見えていないのに、手の位置を二つ抽出した場合は、抽出に誤りがあったと見なしている。抽出精度を厳しく評価しても、全体的に特徴点が抽出できていることが確認できた。

### 5.2.2 動作推定結果

表 6 に教師の動作推定結果を示す。講義映像 1 は教師の動作モデルを作成したときのモデルとなった教師の別の講義映像であるため講義映像 2、講義映像 3 に比べて全体的に推定結果が良くなっている。教師が異なる講義映像 2 と講義映像 3 の場合でも、「板書中」と「説明中」の動作に関しては適合率、再現率ともに 70% を超えている。一方、教師の「移動中」という動作については、いずれの映像でも推定結果が悪い。これは、教師の「移動中」という動作は、他の二つの動作に比べ特徴が捕らえにくいことが原因と考えられる。しかし、表 2 から分かるように、教師の「移動中」という動作は講義内で起こる割合が非常に低いため、「移動中」の動作推定結果が悪くても、教師の説明対象を抽出するのにあまり影響はないと考えられる。また、人が講義映像を観察しても、音声情報がなければ、教師の動作が「説明中」であるのか「移動中」であるのか判断が難しかったため、画像処理のみで「移動中」と推定することには困難があると思われる。この問題は、教師が発話中であるかどうかを

音声処理をすることで解決できると考えられる。

### 5.2.3 撮影対象抽出結果

ACE がどれだけ正しく撮影対象を捕らえていたかを測る実験を行った。その結果を表 7 に示す。実験は、著者が撮影対象として抽出してほしいと考える範囲を、ACE が適切に抽出することができているかを検証するという設定で行った。いずれの映像に対しても、90%以上の精度で撮影対象が抽出できていることが分かる。5.2.2 節の動作推定結果にもかかわらず、撮影対象をうまく捕らえることができているのは、ACE は教師の動作推定結果のみを用いて撮影対象を決定しているのではなく、教師の動作情報の他に、教師が板書した板書履歴なども考慮して撮影対象を決定しているため結果的には撮影対象をうまく捕らえることができたと思われる。また、撮影対象範囲は厳密にその範囲が限定されるのではなく、撮影対象に収まってほしい黒板上の板書や図形が撮影対象内に含まれていればよいということも表 7 のような結果が得られた要因であると考えられる。

## 6. おわりに

本稿では、講義自動撮影システムについて述べた。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。実際に講義映像に画像処理を適用した結果、ACE は総合的に適切な撮影対象を抽出できることが確認できた。しかし、教師の動作推定については教師の音声情報などを取り入れることで、さらに推定精度を向上させる必要があると考えられる。

### 文 献

- [1] 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, Vol. J84, No. 2, pp. 248–257, 2001.
- [2] 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”. 情処研報, 2001-CVIM-125-5, 2001.
- [3] 山口達, 吉川大弘, 篠木剛, 鶴岡信治. “状講師の動作認識に基づいた遠隔授業映像の自動撮影”. 信学技報, PRMU2000-181, 2001.
- [4] 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”. 電子情報通信学会, 信学技法, Vol. 100, No. 701, pp. 79–86, 2001.
- [5] 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.
- [6] WIDE project. “DV Stream on IEEE1394 Encapsulated into IP”. <http://www.sfc.wide.ad.jp/DVTS/>.
- [7] N.Johnson and D.Hogg. “Representation and synthesis of behavior using Gaussian mixtures”. *Proc. Image and Vision Computing 20*, pp. 889–894, 2002.