

黒板講義を対象とした遠隔講義のための

講義自動撮影システムの構築

錦織 修一郎, 菅沼 明, 谷口 倫一郎

九州大学大学院システム情報科学研究科

〒 816-8580 福岡県春日市春日公園 6-1

TEL: 092-583-7618

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

あらまし 近年、大学およびその他の教育機関において遠隔講義が行われている。これらの遠隔講義においてはカメラを固定して撮影をしているか、カメラマンがカメラを操作して講義を撮影している。しかし、カメラを固定した場合には板書が見づらく、変化のない退屈した映像になってしまう。また、カメラマンを用意する方法では、毎回の講義にカメラマンを雇うのは難しい。本研究では、黒板を使用して進める講義を対象としてカメラで撮影した映像を計算機で処理し、カメラの向きやズームを計算機で制御することで講義を自動撮影するシステム ACE(Automatic Camera control system for Education) の構築を目指している。本稿では ACE を模擬講義に適用した結果と考察を述べる。

キーワード 遠隔講義, 画像処理, カメラ制御

Development of An Automatic Camera Control System for Distance Lecture with Blackboard

Shu'ichiro Nishigori, Akira Suganuma and Rin'ichiro Taniguchi

Graduate School of Information Science and Electrical Engineering, Kyushu University

6-1, Kasuga-kouen, Kasuga, Fukuoka, 816-8580 Japan

TEL : 092-583-7618

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

Abstract The growth of a communication network technology enable us to attend a distance lecture. When the lecture scenes are taken for the distance lecture, a camera-person usually controls the camera in order to capture suitable scenes; otherwise, the camera sets up to capture the same position at all times. It is not easy, however, for every occasion to employ a camera-person. And the scene captured by a fixed camera is unable to give us the feeling at the live lecture. It is necessary, consequently, to control a camera automatically. We have developed ACE(Automatic Camera control system for Education) system which control a camera with computer vision techniques. This paper presents ACE system and an experiment with it.

key words distance lecture, image processing, camera control

1 はじめに

現在、ネットワークの高速化に伴い、遠隔講義を取り入れる大学および各種学校が増えている。遠隔講義とは、空間的な制約を超越して受講する講義のことである。現在の遠隔講義の形態は二つに大別される。一つは、実際の講義の映像と音声を送信し、遠隔地でこれをプロジェクタ等により投影して受講する形態である。これは予備校および放送大学等で現在取り入れられている形態である。もう一つは、提供された電子教材を受講者が計算機上のブラウザ等のツールを用いて参照する形態である。インターネットの発達に伴い、計算機上のツールを使用する形態の遠隔講義の研究は近年盛んに行われている。前者は実現が比較的事業であること、講義の雰囲気が伝わりやすいことが利点であるが、講義映像を見るだけの受身の講義になるという問題がある。それに対して後者は学生個人に合わせた学習が可能である反面、実現するにあたってコストの問題、教材作成の手間といった点が問題である。

現在、大学の講義は以下の形態に大別できる。

- 教師が適宜板書を行いながら講義を行う
- OHPもしくはPowerPoint等のプレゼンテーションツールを用いてスクリーンに投影した資料を基に講義を行う
- 学生各人に計算機が割り当てられ、計算機上に存在する電子教材を用いて講義を行う

実際には、上記の講義形態を組み合わせた講義が存在する。近年では情報処理など一部の講義では学生各人が計算機を使用し、教材も計算機上の電子教材を用いる形態が増加してきた。しかし、現在も大学の講義で黒板を主に使用する講義が多数を占めている。さらに、将来においてもOHP等と組み合わせた形の講義になることはあっても、消滅することは無いと思われる。

本研究では、遠隔講義を支援するために、黒板を使用する講義を自動撮影するシステムACE(Automatic Camera control system for Education)の構築を目指している。本稿では、講義を自動撮影するための画像処理ならびにカメラワークの方法を提案する。

以下2節で遠隔講義の撮影を自動化する意義と概要について説明し、3節で遠隔講義映像において撮

影すべき対象とその撮影方法とを説明する。その後4節で画像内の教師の位置を推定する手法について説明し、5節で画像内の最新の板書が書かれている部分を抽出する手法を説明する。6節で本手法の有効性を確認するための実験を行い、その結果および考察を述べる。最後に本研究の結論と今後の課題を述べる。

2 講義撮影の自動化

本研究で想定する遠隔講義は、講義をカメラで撮影した映像ならびにマイクで集音した音声を遠隔地で再生する形態である(図1)。その際、講義の撮影法が問題になる。カメラを固定したまま講義を撮影した映像は受講者にとって飽きやすい映像になる。また、使用する黒板が広い場合、黒板全体を画面内に捉えるためにカメラと黒板を離さなければならない。そのため、撮影した講義映像中の板書の文字が読みづらくなる。そのため、講義毎にカメラマンを用意して講義の状況に応じた映像を撮影することが望ましい。しかし、毎回の講義にカメラマンを雇うことは難しい。本研究では、計算機によりカメラを自動的に制御することで講義の撮影を自動化する。本研究で提案する手法では、固定カメラ1台と、パン・チルトおよびズームの機能を持つ首振りカメラを1台使用する。固定カメラで撮影した映像を画像処理することで講義撮影に必要な情報を獲得し、獲得した情報に基づいて首振りカメラが適切なカメラワークを行う。

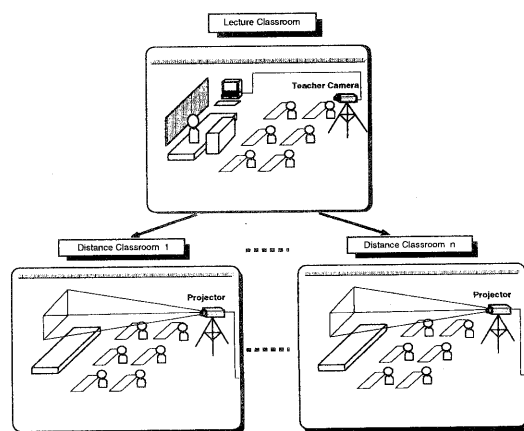


図1: 本研究で想定する遠隔講義の形態

3 カメラワーク

講義を自動撮影する際には、カメラの撮影対象と撮影方法が重要である。以下に、講義中においてカメラが撮影する対象の選択方針と、学生に対して効果的な撮影方法を説明する。

3.1 撮影対象の選択方針

講義の映像を見る学生が複数存在する場合、講義室内で各人が見たいと思う箇所は様々である。そのため、撮影する講義映像の基準を学生側に合わせることは困難である。そこで、本手法では撮影する講義映像の基準を教師側に合わせることにする。つまり、講義中に教師側の立場で重要であると考えられる部分を選択して撮影する。

本手法においては、現在教師が説明している対象を教師側の立場において最も重要であると考えられる。したがって、講義中に教師が説明している対象を常にカメラが捉えるように講義映像を撮影する。

講義中における教師の説明対象は黒板に書かれた文字や図形(板書)であると考えられる。板書は映像を見る学生側にとっても非常に重要な情報であり、最も注目度が高い。そこで、講義中に板書された部分に注目して撮影する。講義中は常に板書が書き加えられたり消されたりする。その中で、最も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の内容である。そこで、本手法では最新の板書を撮影の対象とする。

3.2 対象の撮影方法

講義映像を撮影する際に、激しく動作する映像や過剰な演出を与える映像は目に負荷がかかるため、自然な動作で目に負荷をかけない撮影方法が望ましい。このことを踏まえ、本手法では最新の板書を撮影する際に、板書を画面の中央に寄せ(センタリング)、板書が画面からはみ出さない程度にまでズームした映像を一定時間撮影する(図2)。このように撮影することで、映像を見る学生側に対して読みやすい板書の映像を提供することができ、学習効果が期待できる。

現在の大学で使用されている黒板は、上下移動式の小黒板が複数枚合わさって構成されるものが多い。また、最新の板書が書き加えられた小黒板の内容がしばらくの間、教師の説明の対象になる可能性が高い。そこで本手法では、通常時には最新の板書を含む小黒板をセンタリングおよびズー

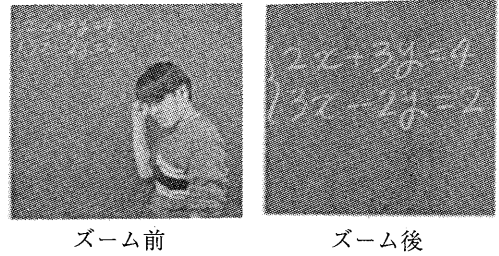


図2: 板書のセンタリングおよびズーム

ムして撮影しておき、新たに板書が書き加えられた際に、その板書を一定時間センタリングおよびズームして撮影する。小黒板に分割されていない場合には、黒板を小黒板に相当する大きさに分割し、最新の板書を含んでいる領域をセンタリングおよびズームして撮影する。

4 教師位置の推定

講義中において、教師の位置の情報は重要である。ACEは講義の映像から最新の板書の位置を探索する。その際、最新の板書は教師の近傍に存在することを利用する。そこで、講義映像中において教師の位置を把握する必要がある。ACEでは、講義の映像の中で常に移動する物体は教師のみであると仮定し、フレーム間差分を用いて教師位置を推定する。フレーム間差分は、移動物体を抽出する際に効果がある手法である [1][2][3] ため、講義中の映像から教師を抽出することが可能である。本手法では、色の変化に強い濃淡画像を用いてフレーム間差分を計算する。フレーム間差分を計算した後、ある閾値で二値化を行い、ノイズを除去するために各画素に対して式(1)のフィルタをかける。

$$X = \begin{cases} \bar{X} & \text{if } |X - \frac{(P_1 + P_2 + \dots + P_8)}{8}| > \frac{1}{2} \\ X & \text{otherwise} \end{cases} \quad (1)$$

ここで、 X および $P_1 \dots P_8$ は二値の画素値であり、 $P_1 \dots P_8$ は画素 X の隣接8近傍の画素値である。フィルタをかけた後に残った二値画像内の白色部分の画素が教師を表す画素の候補となる。

図3において、(a)と(b)の連続した濃淡画像のフレーム間差分を閾値で二値化した画像が(c)である。(c)に式(1)のフィルタをかけた画像が(d)で

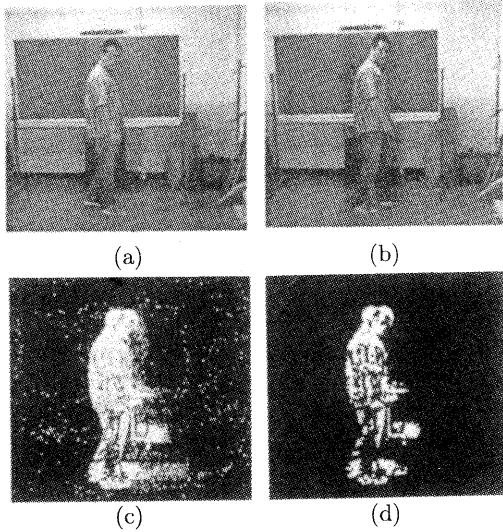


図 3: フレーム間差分による移動物体抽出

ある。(d)の画像において、白色部分の画素が教師の存在する領域として抽出されている。

その後、二値画像内の白色部分の画素の重心 (x_g, y_g) を式 (2) によって求める。

$$(x_g, y_g) = \left(\frac{\sum_{i=1}^n x_i}{n}, \frac{\sum_{i=1}^n y_i}{n} \right) \quad (2)$$

ただし、 x_i, y_i はそれぞれ n 個の白色部分の画素のうち i 番目の画素の座標を表す。この重心はおおよそ教師の重心とみなせるため、この重心の近傍における白色部分の画素の外接矩形を教師の存在する領域(教師領域)とする。本手法ではより精度の高い推定を行うため、一度求めた重心の近傍で教師領域の再計算を行う。

5 最新の板書領域の抽出

最新の板書をズームして撮影する際に、最新の板書の始点と終点の情報が必要である。板書の始点と終点を対角線を持つ矩形を板書領域と呼ぶことにする。この節では画像中から最新の板書領域を抽出する手法について説明する。

5.1 前景の抽出

画像中から最新の板書領域を抽出するためには、毎フレームの画像処理の初期段階で画像内の物体を背景と前景に分ける必要がある。前景を抽出す

る手法として頻繁に使用されるものに背景差分が挙げられる。背景差分とは、事前に撮影しておいた背景画像と、現在の画像との差分を計算するものである。背景差分は単純で高速な計算が可能であるが、背景に変化がない場合にのみ有効であり、照明の変化に影響されやすい。通常、講義室内の照明は蛍光灯であり、この照明条件下ではフレーム毎に画像の輝度が微小ながら変化する。そこで、照明の変化に比較的ロバストである背景モデル [4] を導入する。黑板上の変化には以下のように 2 種類ある。黑板の上にチョークで文字が書かれた場合と、黑板上の文字が消された場合である。前者は前景として抽出してほしいが、後者は前景として抽出してほしくない。この条件を満たすように背景モデルを以下のように変更した。ACE で使用する背景モデルは画像内の各画素 P に対して以下の二つの値を保持する。

- 画素 P における輝度の最大値 $Max(P)$
- 画素 P における隣接フレーム間の輝度の差分最大値 $D(P)$

講義撮影を開始する前に初期の背景モデルを計算(背景の学習)する。その際、教師が画面内に存在せず、黑板には何も書かれていない状態しておく。ある一定のフレーム数の間に上記の二つの値をフレーム毎に更新する。

講義撮影中に背景モデルを用いて画像内から前景を抽出する際には、画像 I において以下の式 (3) を満たす画素 P を前景の候補とする。ここで、 $I(P)$ は画像 I における画素 P の輝度である。

$$I(P) - Max(P) > D(P) \quad (3)$$

背景モデルを使用した前景の抽出例を図 4 に示す。(a) は処理の対象となる画像である。(a) の画像において上述の処理を行い、ノイズ除去を経て抽出されたものが (b) である。この図では前景が白色の部分で表されている。画像からも分かる通り、板書の部分のみが抽出されている。

5.2 板書領域と教師領域の分離

5.1 節の方法で前景を抽出するが、講義中の映像に対して前景抽出を行った場合、板書だけではなく、教師も前景として抽出される。そのため、板書と教師を分離するために 4 節で説明した教師領域を

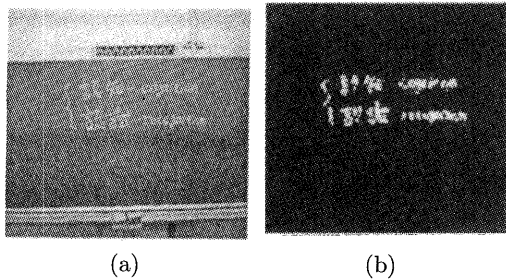


図 4: 背景モデルを使用した前景抽出

利用する。前景を抽出した後に、教師領域の部分の画素値を 0 にすることでマスクを行う。その後マスクされていない前景の画素を板書であるとみなし、その外接矩形を板書領域とする。また、講義中において最新の板書が書かれている場所は比較的教師の近くに存在することが多い。そこで、抽出された前景の中で教師の重心の近傍に存在するものを最新の板書領域とする。以下に処理の流れを示す。

1. フレーム A を取得する
2. 一定時間後にフレーム B を取得する
3. A と B のフレーム間差分により教師領域を計算する
4. フレーム B に対して背景モデルを使用した前景抽出を行う
5. 前景抽出後の二値画像において教師領域の部分をマスクする
6. マスク後に残った前景において教師近傍の外接矩形を板書領域とする

5.3 板書のズーム開始の判定

各フレームごとの板書領域は 5.2 節の方法で求めるが、板書をズームするタイミングを考慮する必要がある。教師が板書を書いている最中にズームを行うと、教師の後ろ姿で板書が隠れた映像を撮影してしまう。したがって、教師が板書し終えたことを判断して、その時書かれた分の板書のみをズームする。

教師の板書が一段落したかどうかを判定する方法として、板書が書き終った時点で教師側に特別なスイッチを押してもらうといった方法が考えら

れる。しかし、これでは教師に負担をかけることになる。また、教師が講義に集中すると、スイッチのことを忘れてしまうであろう。そのため、教師に特別なアクションをとってもらうことは適当ではない。そこで動画像中から板書の書き終りのタイミングを判定することを考える。ここで、教師の姿勢および動作を推定してタイミングを判定することが考えられるが、黒板に文字を書くといった細かい動作を判断するような動作認識は非常に困難であり、精度も期待できない。

そこで、ACE では各フレームごとに抽出される板書領域の変化に着目する。教師が板書を行っている間に抽出された板書領域は常に変化している。これは以下の 2 点に起因する。

- 教師が各フレームごとに板書を書き加えるもしくは消すことにより板書領域が増減すること
- 教師の移動に伴って教師領域のマスクの位置も移動するため、板書領域のマスクされる部分が変化すること

教師にもよるが、大半の教師は板書を一旦書き終えると、学生にその内容を見せるために図 5 のように、板書が書かれている場所から少し離れる。教師が板書から少し離れている間は、新しく板書が加わることがなく、さらに教師領域のマスクによって板書領域がマスクされることが無いため、板書領域は変化しない。従って、板書領域が変化しない状態が一定時間続いた場合には、教師の板書が一段落したとみなして、その時点で抽出されている板書領域をズームする。

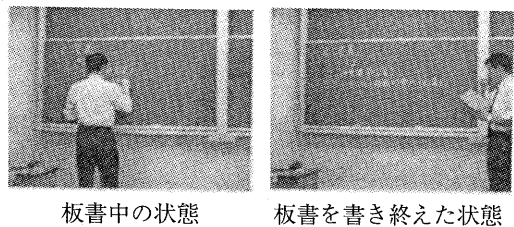


図 5: 板書の書き終りの判断

ACE では最新の板書をズームして撮影する。そのため、一度最新の板書と判断された板書は、次回は最新の板書領域として抽出されてはほしくない。最新の板書領域を抽出した後に、抽出された板書領域の部分のみを対象として背景モデルを更新することで常に最新の板書領域のみを抽出する。

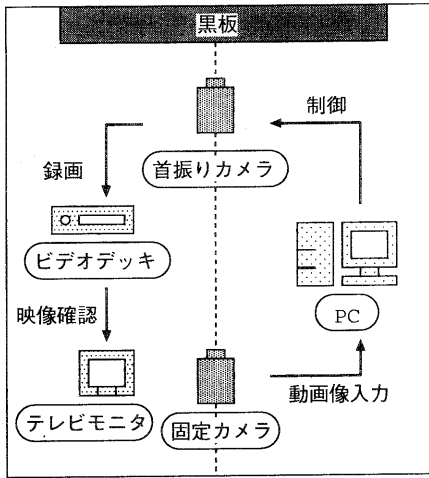


図 6: 実験環境

6 実験

本研究で提案する手法の有効性を確認するため、講義の自動撮影実験を行った。

6.1 実験の環境ならびに設定

本手法で使用した実験環境を図 6 に示す。画像処理用の映像を撮影する固定カメラは、図 7 のように黒板が画面に収まる程度の距離で中央に設置し、板書がどの位置に書かれても画面内に収まるようにする。遠隔講義映像を撮影する首振りカメラは、ズームした際に板書がはっきりと読める程度の距離で中央に設置する。その際に、2 台のカメラの座標を一致させるため、事前にキャリブレーションを行っておく。

首振りカメラを制御する PC は、PentiumII 300MHz の CPU および 128MB のメモリを搭載している。さらに、画像処理を高速化するために、超並列高速画像処理ボード (IMAP-VISION ボード) を PC に装着して使用している。この画像処理ボードは、256 個のプロセッサを用いて、1 回の命令で 256 個分のデータを処理する SIMD 型の演算が可能である。画像処理用に入力する画像のサイズは 256 × 240 ピクセルである。

講義撮影を開始する前に、黒板に何も書かれておらず、教師も存在しない状態の映像で背景の学習 (100 フレーム約 4 秒間) を行う。その後、教師に

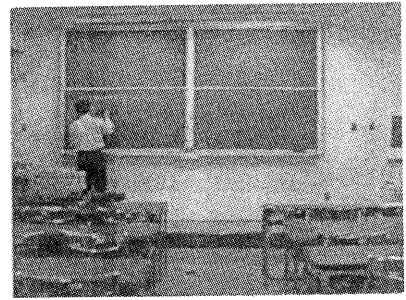


図 7: 固定カメラからの映像

数学の講義を約 10 分間行ってもらった。その際に講義では黒板のみを使用し、OHP 等は使用していない。生成される遠隔地の講義映像をテレビモニタで確認しながらビデオテープに録画した。また、その時にマイクを使用して音声も録音した。

6.2 実験結果ならびに評価

実験による本手法の評価は二つに分けて行った。一つは、本手法で撮影した映像がシステムとして意図した映像であるかを評価する。もう一つは、本手法で撮影した映像が学生側にとって望ましい映像であるかを評価する。前者をシステム内部評価、後者を学生評価と呼ぶ。システム内部評価は、生成された講義ビデオ映像を著者が見て評価を行い、学生評価は何人かの学生に講義ビデオ映像を見てもらい、アンケートにより評価を行った。

6.2.1 システム内部評価

教師位置の推定に関しては良好な結果が得られ、本手法の有効性を確かめられた。これは、フレーム間差分画像から教師の重心と教師領域を求めた後に、重心の近傍で教師領域を再計算し、教師領域の絞り込みを行ったためと考えられる。ただし、講義中に教師が立ち止まった際に推定が出来ない状態も見られた。これは、フレーム間差分の手法が移動物体の抽出を前提としているため、動きが少ない場合には精度が悪くなるからである。しかし、本手法では教師領域が推定できないフレームが少々存在しても最新の板書領域抽出には影響がなく、大きな問題にはならない。

最新の板書領域の抽出処理については、全体的には教師によって書かれる板書のうち、最新の板

表 1: アンケート結果

アンケート質問	最高値	最低値	平均値
板書は見やすかったか?	5	3	4.29
教師の様子がよく分かったか?	5	4	4.57
自分の見たい所がよく見えたか?	5	2	3.29
講義の臨場感は得られたか?	5	2	4.00
講義映像として学習効果があったか?	5	3	4.14

書のみを正確に判断することができ、本手法の有効性を確かめることができた。しかし、板書領域が途切れて抽出されることがあった。これは、教師が書く板書に濃淡があることに起因すると考えられる。背景モデルとの差分で前景を抽出する際に、薄く書かれた板書の画素値は閾値を越えず、前景として判別されない。そのため、板書の始点もしくは終点の付近で薄く書かれた板書が存在した場合に、途切れた板書領域が抽出される結果となる。

6.2.2 学生評価

本手法で自動撮影した映像をビデオテープに録画した約 10 分間の音声付き講義ビデオ映像を 7 人の学生に見てもらい、アンケートに回答してもらった。アンケートは各質問に対して 5 段階評価で行い、5 が最も良く 1 が最も悪い。アンケート結果を表 1 に示す。表中において、最高値は各質問に対しての最も高い評価の値であり、最低値は最も低い評価の値である。また、平均値は単純に評価値の平均を計算し、小数第 2 位まで求めた値である。評価の最高値ならびに評価の最低値を参照するのは、講義映像に対しての学生の評価が個人の感性によって左右されるため、平均値だけでなく、最高値と最低値によって個人差を見るためである。

板書の見やすさの評価については、平均値が 4.29 と高く、本手法の有効性を示す結果となった。評価の最低値も 3 と極端に悪い評価ではなく、板書のズームの効果がどの学生にとっても高いことを示している。

教師の様子がよく分かったかという質問に対する評価は非常に高いものになっている。これは、3.2 節で説明した本手法の撮影方法が功を奏していると考えられる。本手法の撮影方法は最新の板書が抽出された際には一定時間の間ズームを行うが、その後は黒板全体を撮影するのではなく、最新の板

書が含まれる小黒板をある程度ズームして撮影する。そのため、結果として教師の動作が分かるくらいまでズームした映像が提供できる。評価の最低値も 4 となっており、学生の個人差がほとんどなく満足している。

自分の見たい所が見えたかという質問に対しては予想通り個人の差が大きく影響した。3.1 節で説明したように、撮影対象の基準を教師側に合わせたため、学生によっては提供された映像が自分の望んでいる場所を撮影したものではない場合が存在する。数字にも現れているように、本手法の撮影基準に対して同調した学生は最高の評価値 5 を与えているのに対し、同調できなかった学生は低い評価値 2 を与えている。平均では 3.29 と中間の評価となっている。

講義の臨場感に関する評価も個人差があり、ビデオ映像を見るだけの受身の講義では臨場感に欠けるという学生が存在する一方、固定カメラで撮影しただけの映像よりは本手法で撮影したビデオ映像に臨場感を感じるという学生も存在した。平均としては 4.00 という値が示すように、本手法の撮影方法に臨場感を感じる学生が多い。

最後に、本手法で撮影した映像の学習効果についての評価は、ビデオ映像を見るだけでも講義と同程度に学習効果が得られた学生が多く、最低の評価値も 3 であり、学生による個人差も少ない。この学習効果の評価は総合的な評価であり、最も重要である。平均値 4.14 からも学習に対する本手法の貢献度が高いことが確認できた。

7 おわりに

本稿では、黒板講義を対象とした遠隔講義中継の際に画像処理技術を応用して講義撮影を自動化する手法を提案した。撮影を自動化することで人

間の手間を省くことが期待できる。実験結果をシステム内部および学生側から評価した結果、本手法の有効性が確かめられた。

今後の課題として以下の三つのことを考えている。

[OHPの対処] 大学の講義においては、黑板のみを使用する講義より黑板とOHPを併用する講義の方が多くなりつつある。この事実に対処すべく、教師がOHPを使用する際にはOHPを撮影の対象に選択する必要がある。

[過去の板書の参照] 本手法では教師の説明の対象が最も新しく書かれた板書である可能性が高いという事実に注目し、最新の板書領域を抽出することを考えたが、実際には教師が過去に板書を行った事柄について説明することもある。その際には過去の板書を撮影対象に選択すべきであり、この対処を考える必要がある。

[音声情報の利用] 現在、本手法では講義の撮影に必要な情報を動画像処理により獲得しているが、講義中に発生する音声も有用な情報である。最新の板書領域を抽出する際に、チョークの音で板書の書き終りを判断することができれば、より抽出精度が高くなると考えられる。音声情報の利用についても今後研究を進める予定である。

参考文献

- [1] Masahiko Yachida, Minoru Asada, Saburo Tsuji: "Automatic Analysis of Moving Images", IEEE Trans. on PAMI, Vol.3, No.1, pp.12-20, 1981.
- [2] 嶺直道, 八木康史, 谷内田正彦: "時系列差分画像を用いた複数移動物体の抽出および追跡", 情処研報 CV 81-7, pp.51-56, 1993.
- [3] Akira Suganuma, Shinichi Kuranari, Naoyuki Tsuruta, and Rin-ichiro Taniguchi: "Examination of an Automatic Camera Control System for Lecturing Scenes with CV Techniques", 3rd Korea-Japan Joint Workshop on Computer Vision, pp.172-177, 1997.
- [4] Ismail Haritaoglu, David Harwood and Larry S.Davis: "W⁴: Who?When?Where?What? A Real Time System for Detecting and Tracking People", International Conference on Face and Gesture Recognition, pp.14-16, 1998.