

Development and Validation of an Automatic Camera Control System Made with a Detection of a Chalking Sound

芦川, 平

九州大学システム情報科学研究所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所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所知能システム学部門

<http://hdl.handle.net/2324/5877>

出版情報 : 火の国情報シンポジウム. 2003, pp.60-67, 2003-03

バージョン :

権利関係 :



黒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

芦川 平, 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 福岡県春日市春日公園 6-1

TEL: 092-583-7618

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

あらまし 我々の研究室では、黒板を使用した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、その撮影方法は人間の目の動きとは違うため、不満を感じる生徒もいた。そこで本研究では、より人間の目の動きに近い撮影方法を考案した。その際に必要になるのが、教師が板書を開始したと終了したことの判定である。画像情報のみでは板書動作の判定を誤ることがあるため、板書動作時に生じるチョークの音を利用し、板書動作の判定を行った。

本稿では、実際の講義を対象にして ACE で撮影した映像と固定カメラで撮影した映像との比較、また従来の ACE 映像と本手法の ACE 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、チョーク音検出、画像処理、板書動作の推定、講義撮影法

Development and Validation of an Automatic Camera Control System Made with a Detection of a Chalking Sound

Taira Ashikawa, Akira Suganuma, Rin-ichiro Taniguchi

Department of Intelligent Systems, Kyushu University

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

TEL : 092-583-7618

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

Abstract We are developing an Automatic Camera control system for Education: ACE, which captures a traditional lecture with a blackboard. The early version of ACE focuses on the latest object written on a blackboard by a teacher when he finished writing it. This videoing strategy, however, is different from moving of student's eyes. In this paper, we present a new videoing strategy which is better than the old one. When our new strategy is realized, it is necessary for ACE to guess beginning and ending of teacher's writing. The early version of ACE mis-judges ending of his writing according to the region occluded by him because it guesses it only using image processing. We have improved, then, our guessing method with both image and sound processing. We have applied, furthermore, our new method to videoing a real lecture and validated it.

Key words Distance lecture, Detection of a chalking sound, Image processing, Guessing teacher's writing, Videoing strategy of a lecture

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、または人間がカメラを操作して行うかのいずれかが採られている。しかし、前者は固定した領域だけを撮影するため十分な情報が得られず、また後者はコストがかかるという問題がある。そこで、我々の研究室では、黒板を使って行う講義を自動的に効率よく撮影するシステム (Automatic Camera control system for Education 以下 ACE) を開発している^[1, 2]。

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、教師の身振りと教師が書いた板書が考えられる。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。ACEでは、遠隔地に伝える講義の撮影法として、教師と教師が新しく書いた板書に重点を置いて撮影する手法を採っている。

上記の撮影を自動で行うために、従来のACEでは、一旦教師が板書を行うと、教師が新しく板書した場所を板書後一定時間ズームして撮影していた。しかし、板書終了後のズームでは、なかなかズームしてくれない場合やズームしている時間が短すぎる場合などが起こった。そこで、本稿では板書を開始した時点で板書し始めた場所をある一定の範囲が映るようにズーム撮影することにした。この戦略を実現するには、教師が板書を開始したことで終了したことを判定する手法が必要である。ACEでは、この判定を画像処理と音処理を使って行っている。

本論文では、2節で今回実現するカメラワークを述べ、3節でACEで行っている画像処理および音情報を用いた板書動作の判定手法を説明する。4節で本手法の有効性を確認するための実験を行い、その結果と考察を述べる。最後に今後の課題を述べる。

2. ACEのカメラワーク

ACEでは動的にカメラを操作して講義の自動撮影を行う。その際にはカメラの撮影対象およ

び撮影方法が重要である。以下に、講義撮影中においてカメラが撮影する対象の選択と、映像を見る学生に対して効果的な撮影方法を説明する。

2.1 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であろうと考えられる箇所を撮影する。

実際の講義において、学生は黒板上に書かれた文字や図形、すなわち板書に注目していることが多い。これは、書いた板書の内容について教師が説明を行うため、教師が説明している内容を理解する手がかりとして板書の内容が重要であることと、学生が板書をノートに書き写すために板書の内容に注目することに起因している。

また、講義中は常に板書が書き加えられたり消されたりするが、その中で教師の説明内容になる可能性が最も高いのは、教師が最も新しく書き加えた板書(最新の板書)の内容である。そこでACEでは最新の板書を撮影の対象とする。

2.2 撮影対象の撮影方法

従来のACEでは、教師の板書終了後、教師が新しく板書した場所を一定時間ズームして撮影していた。しかし、教師が一度に多量に板書した場合、なかなかズームしてくれなかったり、ズームされている時間が内容を把握するには短すぎたりすることがあった。また、学生は講義では新しい板書が見え始めた時点でその板書を見ている。しかし、従来のACEでは板書が書き終わるまで、学生は板書がはっきりと見えない。これは、人間の目の動きとは違うため学生にとってあまり好ましくない。そこで、新しい板書が見え始めた時点で、板書にズームして撮影すれば、人間の目の動きと似たものになるため、学生にとってよい映像になると思われる。

講義を撮影する際に、ACEは次の6つの状態を遷移する。

- (1) 初期状態
- (2) 板書開始状態
- (3) 板書途中状態
- (4) 板書終了状態

(5) 板書終了後、一定時間以内の状態

(6) 板書終了後、しばらくたった状態

以下にそれぞれの状態におけるACEのカメラワークを述べる。

講義が始まった直後は、教室がどのような環境であるかを学生に知らせるため、状態(1)では、黒板全体が映るようにしておく。

状態(2)では、ACEは板書が見え始めた時点で新しい板書を中心にある一定の範囲が映るようにズームして撮影する。その後教師が長い文を書いた場合などは、新しい板書がカメラに映らない可能性がある。そこで状態(3)では教師が新しく書いた板書がカメラの画面内に収まっていない場合は、画面内に収まるようにカメラを調整する。

板書が終了した時点で、もう一度学生に最新の板書をすべて見せることによって、学生は板書の内容を把握しやすくなる。そこで状態(4)では、最新の板書がすべて映るように調整し、一定時間撮影する。

板書終了後のしばらくは、板書が新たに書き加えられた小黒板の内容が教師の説明の対象になる可能性が高い。そこで、状態(5)では、最新の板書を含む小黒板を画面に収まる程度にズームして撮影する。しばらくたった後は、板書の文字よりも教師の様子の方が重要になると考えられるため、状態(6)では教師が映るようにカメラを調整する。

3. 画像情報と音情報による撮影対象の抽出

2節で述べたカメラワークを実現するためには下の3つを抽出する必要がある。

- 教師領域(教師を含む矩形領域)
- 最新の板書領域(教師が黒板に板書した領域)
- 板書の開始と終了の判定

ACEでは黒板全体を映した映像と講義中に発生するチャーク音からこれらの情報を抽出している。^[4]

3.1 教師領域の抽出

ACEは、講義中に移動する物体は教師だけであると仮定し、フレーム間差分の手法を利用して教師領域を抽出している。しかし、教師が動かない場合には、フレーム間差分画像だけでは教師を抽出できない。そのため、1つ前のフレームで求めた教師領域を利用して以下のようにして求める。

- (1) フレーム間差分を行い二値化を行う
- (2) ノイズを除去する
- (3) 白色画素の重心を計算する
- (4) 重心の近傍領域を対象として重心を再計算する
- (5) 重心の近傍領域において白色画素の外接矩形を移動物体領域とする
- (6) 前フレームの教師領域と移動物体領域を合わせ、教師候補領域を求める
- (7) 教師候補領域において、背景差分画像を用いて教師領域を抽出する

3.2 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、下のような手順で板書領域を抽出している。

- (1) 背景差分を行い二値化を行う
- (2) ノイズを除去する
- (3) 教師領域をマスクする
- (4) マスク後に残った前景の外接矩形を板書領域とする

3.3 板書動作の開始と終了の判定

3.3.1 画像情報による板書動作の判定

黒板の上に書かれた板書は、講義の映像を画像処理することでその位置を得ることができる。板書が書き加えられると、新しい板書領域が得られる。そこで、ACEはある一定時間新しい板書領域が常に近い範囲で抽出されれば、板書開始と判定する。また教師は、板書を終えた後には、その板書から少し離れて学生にその内容を見せることが多い。このとき、板書領域には変化が起こらない。これを利用して、ACEは、板書領域が一定時間変化しなければ、板書を終了したと判定する。しかし、画像情報のみによる板書動作の判定には以下のような問題点がある。

- (1) 教師自身による文字の隠蔽により、板書中にもかかわらず板書領域が変化しないため、板書終了と判定してしまう。
- (2) 教師領域が正確に抽出されないと、教師領域のマスクがうまくできず板書領域が正確に抽出できないため、板書動作の判定を誤る。

そこで、板書動作の判定を正確に行うために、黒板講義における板書動作中に発生するチョーク音に着目した。

3.3.2 チョーク音検出による板書動作の判定

チョーク音は、

- 衝突音
- 発生間隔が一定時間以上開く

という特徴が見られる。また、講義中に発生する衝突音でチョーク音以外のものは少ない。そこで、チョーク音の検出は衝突音の発見で行う。

チョーク音は衝突音であるため、単位時間ごとのスペクトルの変化が大きい。そこで、スペクトルの変化を検出する指標である Spectrum Flux^[3]を用いる。また、なるべくチョーク音の変化のみで指標値が変化するように、チョークの特徴周波数帯が表れることが多い1-4kHzに周波数帯を限定した。これを帯域限定 Spectrum Flux(BPSF)と呼ぶ。BPSFを以下のように定義

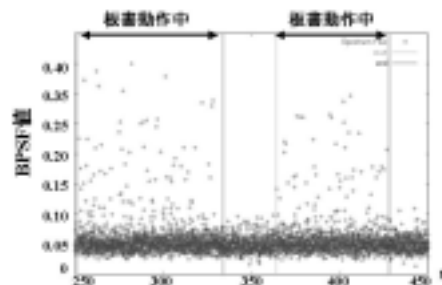


図 1: 帯域限定 Spectrum Flux

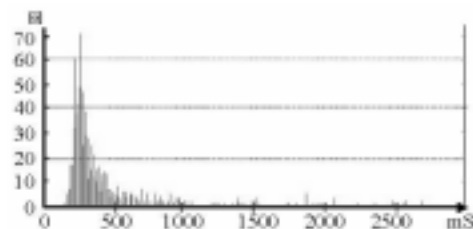


図 2: チョーク音の時間間隔

する。

$$BPSF = \frac{1}{(N-1)(k_2-k_1)} \sum_{n=1}^{N-1} \sum_{k=k_1}^{k_2} [\log(A(n,k)+\delta) - \log(A(n-1,k)+\delta)]^2$$

ただし、 $n = 0, 1, 2, \dots, N-1$ とし、 $A(n, k)$ は入力信号の n フレーム目の周波数 k の離散フーリエ変換値であり、 δ はオーバーフローを避けるためのシフト値である。 N はフレーム数、 $k_1=1\text{kHz}$ 、 $k_2=4\text{kHz}$ である。BPSFは、隣接フレーム間のスペクトル変動が求まる。チョーク音が発生した場合、隣接フレーム間でスペクトルの変化が大きいため、BPSFは大きくなると考えられる。

図1に講義中の音のBPSF値の時間変化を示す。BPSF値はフレーム幅を10ms、フレーム数を4として計算した。図の252-334秒と364-428秒間が板書動作中である。板書動作中のBPSF値は、通常の値より大きい値を示すことが多いことがわかる。これより、BPSFの値がある閾値以上であれば、その音をチョーク音とみなすことができる。

チョーク音による板書動作判定は、以下のように行った。

- チョーク音を α 秒以内に β 回検出したら板書開始と判定する

- 板書動作中で、 γ 秒間チョーク音が検出できなければ板書終了と判定する

実際にある教師の板書中のチョーク音の時間間隔の分布を調べたところ図2のようになった。チョーク音の時間間隔は最頻値 250ms、平均 380ms、標準偏差 280ms、最大値が 2960ms であった。このデータより、 $\alpha=1.0$ 、 $\beta=3$ 、 $\gamma=3.0$ として、板書動作判定を行う。

しかし、このチョーク音による判定手法では、板書動作の終了判定時刻が実際の板書終了後から最高 秒遅れてしまう。これはあまり望ましくない。音情報による板書中である確率を求め、画像情報の動作判定と融合することにより早く正確に板書動作の判定ならびに撮影対象の抽出を行う。

3.3.3 音情報による板書中である確率 P_b

前節の音情報による判定手法をもとに、音情報による板書中である確率 P_b を以下のように定めた。チョーク音の検出回数が、 α 秒以内に β 回に近付けば、確率 P_b は高くなり、また、チョーク音が検出できなければ、時間に比例して確率 P_b は低くなると考えられる。また、チョーク音は時間間隔をおいて発生するため、チョーク音が一度発生すると、一定時間 m だけは確率 P_b を減少しないようにした。 $\alpha=1.0$ 、 $\beta=3$ 、 $\gamma=3.0$ 、 $m=0.250$ とした場合の P_b の例を図3に示す。

3.4 画像情報と音情報による板書動作の判定ならびに撮影対象の抽出

板書動作の判定ならびに撮影する領域(以下、撮影領域)の抽出方法を具体的に述べる。

画像情報による板書動作の判定が開始と判定

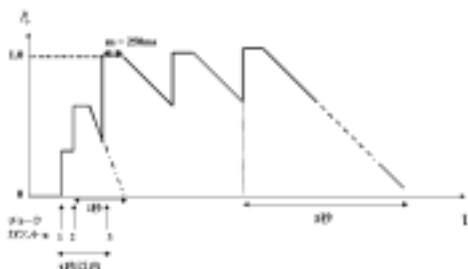


図3: チョーク音による板書動作中確率 P_b ($\alpha=1.0$ 、 $\beta=3$ 、 $\gamma=3.0$ の場合)

し、かつ音情報による板書中である確率 P_b が閾値 TH_p より高い場合、教師が板書を開始し、またその板書が見えていることになる。その場合、ACEは講義が板書開始の状態であると判定する。また、得られた板書領域を中心としてある一定の範囲を撮影領域とする。

板書開始と判定された後、板書終了と判定されるまでをACEは講義が板書中の状態であると判定する。画像情報から得られた板書領域が現在の撮影領域内になければ、現在の撮影領域内にはない板書領域を撮影領域とする。

板書終了の判定は、以下のようにして行う。確率 P_b が閾値 TH_p より高い場合には、画像情報から板書の終りを検出しない。逆に P_b が閾値 TH_p より低い場合には、板書が終了している可能性が高いため、画像情報から板書の終りの判定を行なう。そして、板書領域として TH_N 回以上同じような位置が抽出されれば、講義は板書終了の状態であるとし、その板書領域を、撮影領域とする。ここでは、判定の閾値を $TH_p = 0.5$ 、 $TH_N = 3$ とした。

4. 実験

本手法の有効性を確認するため、実際の講義にACEを適用し実験を行った。

4.1 実験の環境

実験で使用したシステムの構成を図4に示す。画像処理用の入力映像を撮影する固定カメラは、黒板全体が画面に収まる程度の距離に配置した。また、マイクは黒板の中央から約2m離れたところに固定して置いた。画像処理用の映像を撮影す

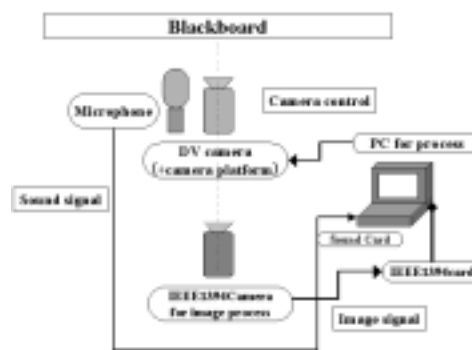


図4: システム構成

る固定カメラは Sony DFW-VL500、音処理用のマイクは Panasonic VW-VMS2 を用いた。処理用計算機として、CPU が Pentium III 1.20GHz、およびメモリが 650MB の計算機を用いた。また、画像をキャプチャーするために IEEE1394 カードを装着した。画像は 680 × 480 画素、5fps で画像処理を行い、音声は 48000Hz、16bit、モノラルで計算機に入力し処理を行った。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、処理用計算機から首振り台と DV カメラを制御した。

実際に九州大学の工学部 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、講義日の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、画像処理用カメラと同じ位置に配置した固定カメラで撮影した映像 (固定カメラ映像) とをそれぞれ DV テープに録画した。それら 2 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、学生に見てもらい評価を行った。また、板書終了時のみ板書領域をズーム撮影する従来の ACE を使って得られたアンケート結果^[1]と、本手法の ACE を使って得られたアンケート結果の比較も行った。

4.2 実験結果ならびに評価

学生が固定カメラ映像および ACE 映像によるそれぞれの講義映像を見終わった後にアンケートを行った。アンケートでは、学生に下の 6 つの質問に答えてもらった。

- (ア) 教師の様子がよく分かったか?
- (イ) 黒板の文字や図形は見やすかったか?
- (ウ) 自分の見たい所がよく見えたか?
- (エ) 講義の臨場感は得られたか?
- (オ) 講義としての総合的な評価はどうか?
- (カ) 通常の講義と比べて理解できたか?

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。

アンケートの結果を表 1、2 に示す。表 1、2 の数値は、(ア) ~ (カ) の質問に対する固定カメラ映像及び ACE 映像の評価値の人数の割合と、5 段階評価の平均値を表している。また、従来の ACE による (ア) ~ (オ) の質問に対するアンケート結果を表 3 に示す。(カ) については、従来の

ACE では学生に質問していなかったため、本手法と比較することができなかった。

固定カメラ映像と本手法の ACE 映像では、すべての質問に対して、ACE 映像の方が明らかに評価は高かった。また、従来の ACE の結果と本手法の ACE の結果を比較しても、本手法の ACE の方が評価の平均値は高かった。しかし、結果に明らかな差が見られない項目もあったため、t 検定を用いて従来の ACE 映像と本手法の ACE 映像に有意差があるかどうかを判断した。

それぞれの質問のアンケート結果に関する考察を以下に述べる。

4.2.1 固定カメラ映像と ACE 映像の比較

質問(ア)に関して、評価値 4 (大体よくわかった) 以上と答えた人が、ACE 映像では 41.7%、固定カメラ映像では 23.4%、であった。固定カメラ映像では黒板全体を画面に映すようにカメラを固定するとカメラのズーム率が低く設定されるため、教師の様子を確認しにくくなる。一方、ACE 映像では基本的に移動式黒板の小黒板を画面に捉える程度にズームしているため、教師の様子は確認しやすくなる。そのため、ACE 映像の評価は固定カメラ映像よりも高かったと考えられる。

質問(イ)に関して、評価値 4 (大体よく見えた) 以上と答えた人が、ACE 映像が 68.8%、固定カメラ映像が 2.1%、であった。黒板全体が常に画面に映されている固定カメラ映像では、板書を読み取るに十分な文字の大きさが得られず、大半の学生が見づらいという評価を行っている。それに対し、ACE 映像では板書が見やすかったという評価を行っている学生が多かった。これより、ACE の板書のズーム処理が有効であったと言える。

質問(ウ)に関しては、評価値 4 (大体よく見えた) 以上と答えた人が、ACE 映像が 57.9%、固定カメラ映像が 0%、であった。固定カメラ映像は黒板の文字が見づらかったので、教師が板書している内容および説明している内容を把握しにくい。そのため、評価は低かった。一方、ACE 映像は最新の板書をズームすることにより教師が板書している文字が適度な大きさで映っていたので固定カメラより評価は高かった。しかし、一

表 1: 固定カメラ映像

評価値	(ア)	(イ)	(ウ)	(エ)	(オ)	(カ)
5	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
4	23.4%	2.1%	0.0%	8.5%	0.0%	4.3%
3	31.9%	2.1%	10.6%	40.4%	8.5%	12.8%
2	25.5%	34.0%	31.9%	40.4%	63.8%	48.8%
1	19.1%	61.7%	57.4%	10.6%	27.7%	36.2%
平均	2.60	1.45	1.53	2.47	1.81	1.85

表 2: ACE映像

評価値	(ア)	(イ)	(ウ)	(エ)	(オ)	(カ)
5	2.1%	12.5%	2.1%	0.0%	2.1%	0.0%
4	39.6%	56.3%	45.8%	29.2%	27.1%	20.8%
3	41.7%	25.0%	39.6%	50.0%	56.3%	62.5%
2	14.6%	4.2%	12.5%	18.8%	14.6%	16.7%
1	2.1%	2.1%	0.0%	2.1%	0.0%	0.0%
平均	3.25	3.73	3.38	3.06	3.17	3.04

表 3: 従来の ACE映像

評価値	(ア)	(イ)	(ウ)	(エ)	(オ)
5	5.9%	1.2%	0.0%	1.2%	0.0%
4	40.0%	24.7%	8.2%	21.1%	8.2%
3	28.2%	32.9%	45.9%	47.1%	45.9%
2	18.8%	30.6%	32.9%	23.5%	32.9%
1	7.1%	10.6%	12.9%	7.1%	12.9%
平均	3.19	2.75	2.49	2.86	2.49

度見逃した板書を見ようとしてその部分が映っていない時、または教師が過去の板書を説明した場合説明対象がカメラに映っていない時などは、自分の見たいところが見れなかったという否定的な意見もあった。

質問(エ)に関しては、評価値4(大体よく得られた)以上と答えた人が、ACE映像が29.2%、固定カメラ映像が8.5%、であった。これはACE映像の方は、カメラがズームするなど動きのある映像になるため、固定カメラ映像より臨場感が得られたと考えられる。しかし、学生がいない場所で講義を行ったので、教師がカメラに向かって話しかけることがなかったため、臨場感は得られなかったという意見もあり、半数の人がどちらとも言えないという答えであった。

質問(オ)に関しては、評価値3(講義映像と

して妥協できる)以上と答えた人がACE映像が85.5%、固定カメラ映像が8.5%、であった。固定カメラ映像においては、90%近くの学生が不満が多いという評価結果が出たため、遠隔講義には非常に不適當であると思われる。一方ACE映像に関しては、講義映像として妥協できると答えた人が85%おり、ACE映像の方が固定カメラよりも、遠隔講義には適當であることがわかる。

質問(カ)に関しては、評価値3(通常の講義と同じくらい理解できた)以上と答えた人がACE映像が83.3%、固定カメラ映像が17.1%であった。これは、固定カメラ映像は、黒板の文字が見えにくい映像になってしまったので、学生は黒板の文字を見ることに集中してしまい、教師の話聞く余裕がなかったため、低い評価になったと考えられる。一方ACE映像の方は、板書開始

で板書をズームするため板書をノートに写すのが容易になり、教師の話を書くにも余裕ができたため、固定カメラ映像よりも高い評価になったと考えられる。

4.2.2 従来の ACE 映像と本手法の ACE 映像の比較

質問(ア)に対する評価に関しては、t検定を行った結果では有意差は見られなかったため、両映像とも同程度の評価といえる。本手法の ACE では板書中に教師がカメラに映っていない場合があり、その時の教師の様子はわかりづらい場合もあった。しかし、板書している時の教師の映像は学生にはあまり必要でないため、2つの評価は同程度だったと考えられる。

質問(イ)に対する評価に関しては、本手法の ACE 映像の方が平均も高く、また t検定を行った結果では有意差が認められた。これより、本手法の ACE 映像の方が優れていることがわかった。これは、従来の ACE の時に使った講義撮影用カメラよりも、今回の ACE の講義撮影用のカメラの解像度が高かったことが要因だと思われる。

質問(ウ)に対する評価に関しては、本手法の ACE 映像の方が平均も高く、また t検定を行った結果では有意差が認められ、本手法の ACE 映像の方が優れていることがわかった。これにより、本手法の撮影戦略が有効であることが示された。

質問(エ)に対する評価に関しては、t検定を行った結果では有意差は見られなかった。

質問(オ)に対する評価に関しては、本手法の ACE 映像の方が平均も高く、また t検定を行った結果では有意差が認められた。これにより、総合的な評価として本手法の ACE 映像の方が優れていることがわかった。

5. おわりに

本稿では、板書開始時に板書をズームして撮影するという新しい撮影戦略を採った。また、画像情報と音情報を用いて教師の板書動作の判定ならびに撮影対象の抽出を行った。ACE を実際の講義に適用し学生からのアンケートから、固定カメラの映像より ACE の映像の方が遠隔講義の映像として有効であることがわかった。また、従来の ACE と比較しても、今回の手法の方が総合的によいという結果が得られた。

今後の課題として以下のことを挙げる。

- 過去の板書の参照
本手法の ACE では、最新の板書だけに着目しているため、教師が過去の板書を参照した場合、説明対象がカメラに映らない場合があった。これに対する対処としては、音声認識による指示語の認識や、教師の状態を推定し、教師が黒板のどの部分を指しているのかといった情報を獲得することが考えられる。これに対応できれば、より学生が見たい映像を提供できると考えられる。
- 学生と教師のリアルタイムでの応答
現在は学生と教師はリアルタイムでの応答を行っていない。これは DVTS^[5]などを用いて動画像を送受信することにより、学生と教師がリアルタイムでインタラクションできると考えられる。これに対応できれば、学生はより臨場感が得られると思われる。

参考文献

- [1] 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.
- [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] Lie Lu, Hao Jiang and HongJing Zhang, “A Robust Audio Classification and Segmentation Method,” Microsoft-Research-TR-2001-79, 2001.
- [4] 芦川平, 菅沼明, 谷口倫一郎, “画像情報と音情報を用いた黒板講義における板書動作の判定,” 電子情報通信学会, 信学技報, Vol.102, No.509, pp.43–48, 2002.
- [5] WIDE project, “DV Stream on IEEE1394 Encapsulated into IP,” <http://www.sfc.wide.ad.jp/DVTS/>