

# 多様な板書環境に対応した講義自動収録システム

新井 崇也<sup>1</sup> 宮川 直人<sup>1</sup> 市村 哲<sup>1</sup>

**概要:** 筆者らは講義映像から講師の追跡動画と黒板全体の静止画を自動で作成するシステム ChalkTalk の開発を行ってきた。しかし、ChalkTalk は黒板の使用を前提としており、多様な環境下ではコンテンツが正常に作成できない問題が発生した。本論文では、ホワイトボードや電子プレゼンテーションを併用しているなどの多様な板書環境に対応させる手法について提案する。講師の位置を検出する方法の改善と誤認識の対策を実装し、講義動画の比較を行った。新システムで作成した講義動画と既存のシステムで作成した講師動画と比較した結果、新システムの方が講師位置の正確性が向上したことがわかった。

**キーワード:** e-ラーニング, 講義動画, 板書認識, ホワイトボード

## A Proposal of Automated Chalktalk Recording for Various Environment

TAKAYA ARAI<sup>1</sup> NAOTO MIYAKAWA<sup>1</sup> SATOSHI ICHIMURA<sup>1</sup>

**Abstract:** We had developed ChalkTalk which automatically produced E-learning materials including a lecture-tracking video and entire blackboard image from videotaped chalk talks. However ChalkTalk is based on the use of the blackboard, the problem do not function properly in various environment. In the paper, we propose a method to correspond to various chalk talks environment. The methods include improved method for detecting the position of the lecturer and so on. As a result of the evaluation, it was found that lecturer location accuracy was improved.

**Keywords:** e-learning, lecture video, blackboard recognition, whiteboard

### 1. はじめに

近年、コンピュータの普及にともなって電子プレゼンテーションを利用した講義が多くなっている。電子プレゼンテーションにて表や図、絵や映像などを用いて分かりやすく講義を行うことが可能となった。

しかし、黒板を利用することも一般的であり、講義内容や設備環境によっては未だに根強い支持を得ている。また、黒板ではなくホワイトボードを利用した講義も普及している。大学や高校は黒板を使用している場合が多いが、学習塾や企業の会議室ではホワイトボードを利用している事が多い。

加えてインターネットの普及により、ネットワークを利

用した WBT (Web Based Training) が注目されている。講師と受講者が一ヶ所に集まる必要がなくいつでもどこでも、各自の進捗状況にあわせて学習できるといった利点がある。

これらの背景に対し筆者らは、講義映像から学習用コンテンツを自動作成するシステム ChalkTalk[1][2] を開発した。講義中の板書全体をハイビジョンカメラで固定撮影した映像から、講師の姿を追従するような動画と、講師を除いた板書静止画を自動生成する講義自動収録システムである。現在、提携企業と共同で開発及び販売を行なっている [3]。そのため、多様な板書環境下で撮影した講義映像を用いて配信用コンテンツを作成することが増え、正常に作成されない場合が存在することがわかった。例えば、いまままで ChalkTalk は、黒板を使用した講義映像を用いて動作

<sup>1</sup> 東京工科大学  
Tokyo University of Technology

検証等を行なってきたため、ホワイトボード使用時等の他の環境下では講師が正しく認識されずに正常なコンテンツを作成できない問題が発生した。

そこで筆者らは、正常に処理できない講義動画について失敗原因を発見するための実験を行い、誤認識の原因を調査した。特に問題が多かったホワイトボードは、板自体が壁に固定されていない場合があり板全体が揺れたり、黒板より講師の影や照明が反射しやすいことが確認された。また、蛍光灯のちらつきなどの影響を受けやすいことがわかった。一方、電子プレゼンテーションを併用した授業では、スライドを切り替える際に講師は止まっていることが多く、そのためスライド切り替えに反応してしまうことがわかった。また、板書形式に関係なく受講者の頭が黒板やホワイトボードを隠すような場合に誤動作しやすいという問題を見出すことが出来た。

そこで本論文では、ホワイトボードや電子プレゼンテーションを併用する等の多様な板書環境にも対応した新しい方法を提案する。具体的には、講師の検出方法の改善と誤認識が発生した際の講師位置の安定化、講師の認識領域の指定を提案する。

新システムで作成した講義動画と既存のシステムで作成した講師動画と比較したところ、新システムのほうが講師位置の正確性が向上していることがわかった。

## 2. 2. 既存研究

固定ビデオカメラで板書全体を撮影した映像ではディスプレイのサイズや解像度の関係で、講師の顔や板書の文字など細かい部分が見えにくいという問題がある。そこで、専用のカメラマンがカメラを操作をして講師や板書に適宜向きを合わせたりズームをする必要があるが、カメラマンの導入にはコストがかかってしまう。そのため、講義映像を自動で収録する研究が従来から多数行われている。例えば、固定カメラとズーム可能な首振りカメラを併用して撮影する研究 [4][5][6] がある。しかし、装置が大掛かりであったりセットアップに時間がかかるなどという問題があった。また、講師を中心に撮影するため板書全体を見渡せない問題もある。

そこで筆者らは、ハイビジョンカメラ 1 台で板書を固定撮影した映像から、ネットワーク配信用の講義映像コンテンツを作成するシステムである ChalkTalk の開発を行ってきた。ChalkTalk は、板書映像から講師の姿を追従した動画および、講師の姿を除いた板書全体の静止画を自動的に作成する。更に、板書全体の静止画からは講師を削除して、板書のみを出力する。そのため、カメラマンが撮影時に追尾・ズームを手動で行う手間や、固定カメラ 1 台あれば済むため大掛かりな撮影機材を必要とせず、手軽で低コストに配信用の映像コンテンツが作成できるという利点がある。講義動画から講師の位置を検出する方法にフレーム

間差分法を用いており、移動物体領域として特定されたマクロブロックの数を黒板の横軸座標で整理してヒストグラム化し、以下のルールに従って講師領域を特定している。

- (1) 移動物体領域が横方向に最も広い範囲で存在する領域を講師の胴体部分と見なす。
- (2) 講師の胴体部分から左右に特定のピクセル以内に存在する動きの大きい移動物体領域は講師の手部分と見なして、胴体部分と領域をつなげて講師領域とし、講師領域の中心付近を撮影の中心点とする。
- (3) 映像酔いを防ぐため、講師領域の移動が少ない場合にはあえてカメラワークを行わず、講師の胴体や手が画面からはみ出しそうになったらパンをする。
- (4) 人間が行なうカメラワークに似せるため、講師が移動したタイミングから意図的に 0.5 秒間遅らせてパンをする。

これらの処理を経て、講師近傍領域の小さな動画を出力すると同時に H.264/MPEG4 AVC 技術を用いた動画圧縮を行うことで元動画（ハイビジョン）に比べ約 1/20 にコンテンツサイズを縮小している。また、移動物体の量（マクロブロックの数のヒストグラム）、講師領域、講師領域の中心点を静止画で出力するデバッグ画像出力機能を有している。

ChalkTalk は、黒板を使用した講義を前提としているため、ホワイトボードや電子プレゼンテーションを用いた講義では、今まで想定していなかった誤認識により、講師領域が変化して動画のカメラワークが不自然になる問題があった。

また、カメラの撮影する中心点は上記ルール (2) の通りに講師領域の中心が講師の中心位置としていたが、誤認識があった場合、講師領域が広がり講師の中心値がずれるため、動画の挙動がおかしくなるなどの問題が発生した。

## 3. 提案

多様な番所環境に対応した抗議自動収録システムを提案する。以下に詳しく述べる。

ホワイトボード環境下で誤反応を抑えるため、フレーム間差分法で検出した移動物体の量のヒストグラムを平滑化し、新しいヒストグラムを作成するようにした。講師領域が異常に大きくなった場合と、講師が動かないまたは講師領域が検出されない場合に対応するためである。また講師領域があまりにも大きくなった際と、逆に移動物体が検知されずに講師領域がない際は誤認識とみなし、前回の講師の中心点を現在の中心点にすることで対策するようにした。またヒストグラムの値が最も大きい場所を講師の中心位置と決定するように変更した。

また、講師が大きく移動すると移動前と移動後の場所に 2 つ以上のヒストグラムの山ができることがわかった。フレーム間差分法では、過去の山と現在の山の区別がつかない

ためである。その対策に複数個の山が現れた場合はそれぞれの山の中心を講師の地点とするようにした。

## 4. 実装 1

### 4.1 4.1 ノイズ除去と誤認識の対策

移動物体の値のヒストグラムを平滑化させるために単純移動平均を用いる。データ列  $p(x)$  で  $M$  番目の値において  $N$  個間の単純移動平均 SMA のを求める式を以下に示す。

平滑化したヒストグラムを画像出力させることで、誤認識が発生しないかどうかの確認を容易にした。

また、講師領域が大きすぎる誤認識が発生した場合と、講師が検出されない場合の対策として、図 1 のように講師領域が指定以上の大きさになった、または図 2 のように存在しない際は講師の中心位置を前回と同じ値とするようにした。



図 1 領域が大きすぎる例

Fig. 1 Example of area is too large

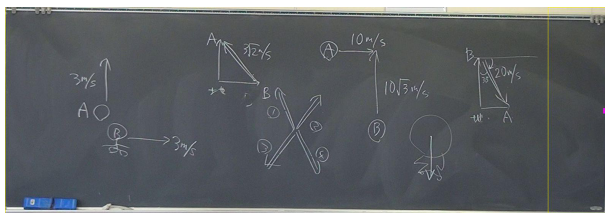


図 2 領域がない例

Fig. 2 Example of area is nothing

### 4.2 4.2 講師の中心位置の決定法

いままで講師領域の中心に設定していた講師の中心位置を、移動体の量が最大となる横軸上に変更した。また、講師が移動する際に複数山が出現する場合、同一時間上のすべてのヒストグラムの山と比べ、あきらかに小さすぎる山は誤認識としてその山は除外する。そして、すべての山の中心を講師の中心位置とした。提案 1 を全て実装したデバッグ画面を図 3 に示す。

## 5. 実験 1

### 5.1 書き方の基本

既存のアルゴリズム、ノイズ除去と誤認識の対策を施したアルゴリズム、講師位置の中心位置の決定法を改善したアルゴリズムの 3 種類の ChalkTalk で講師動画を作成した。そして、問題の発生した 3 種類の講義動画で講義動画

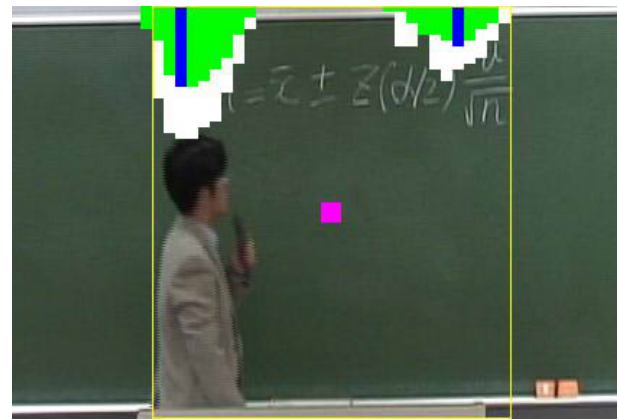


図 3 デバッグ画面の例

Fig. 3 Example of debug screen

を作成し、誤動作の数を比較することで、どのシステムが有効か評価を行った。使用した動画の詳細を以下に示す。

#### 動画 A

白板であり、受講者の頭と板書の文字に誤認識する

#### 動画 B

黒板であり、受講者の頭が誤認識し、講師が板書前から居なくなる

#### 動画 C

白板にスライドを併用しており、カメラと板書の距離が遠い

動画のサイズはすべて 1920x1080 であり、時間は 17 分 13 秒である。また、誤動作は大きく分けて以下の 3 種類ある。

#### 誤動作 A

講師は移動していないが画面が大きく移動する

#### 誤動作 B

講師が移動していないが画面が講師が映る範囲内で移動する

#### 誤動作 C

講師が移動中であるときに画面が追いつけない、または追い越す

### 5.2 5.2 実験結果

既存のアルゴリズムの実験結果を表 1、ノイズ除去対策を施したアルゴリズムを表 2、講師位置の中心位置の決定法を改善したアルゴリズムを表 3 に示す。動画 A は受講者の頭が板書をさえぎってしまうため誤動作 A が、講師付近の文字に反応してしまうため誤動作 C が発生していた。動画 B は講師が板書前にいない際に受講者の頭に反応することが多かったため誤動作 A が多数発生した。動画 C は板書が離れているため、誤動作 B が多かった。

既存のアルゴリズムと比べると、改良したアルゴリズム

表 1 測定結果 1 (誤動作回数)

Table 1 Result of measurement1 (The number of malfunctions)

	誤動作 A	誤動作 B	誤動作 C	合計
動画 A	15	5	7	27
動画 B	22	2	2	26
動画 C	3	9	0	12
平均	10	5.3	3	21.6

表 2 測定結果 2 (誤動作回数)

Table 2 Result of measurement2 (The number of malfunctions)

	誤動作 A	誤動作 B	誤動作 C	合計
動画 A	15	5	7	27
動画 B	22	2	2	26
動画 C	3	9	0	12
平均	10	5.3	3	21.6

表 3 測定結果 3 (誤動作回数)

Table 3 Result of measurement3 (The number of malfunctions)

	誤動作 A	誤動作 B	誤動作 C	合計
動画 A	15	5	7	27
動画 B	22	2	2	26
動画 C	3	9	0	12
平均	10	5.3	3	21.6

は誤動作がかなり改善されている。しかし、表 2 と表 3 を比較すると修正された誤動作の総数自体はほとんど変化がないが、対策できた誤動作の種類が異なっていることがわかる。ノイズ除去の対策を施したアルゴリズムは、大きなノイズで発生する誤動作 A の誤動作数が減っており、講師位置の中心位置の決定法を改善したアルゴリズムは小さいノイズで発生する誤動作 C の誤動作数が減っている。

しかし、どのアルゴリズムでも改善されなかった問題がある。それは講師が動いておらず板書の下部に映り込む受講者の頭や、電子プレゼンテーションのスクリーン、板書の文字に反応してカメラワークが誤動作する問題である。フレーム間差分法では講師かどうかを見分けることは難しく、そのため板書領域で移動物体を検出する範囲を別々に指定する必要があると考えた。

## 6. 実装 2

板書領域と移動物体検出領域を別々に設定できる機能を実装した。板書領域は講義映像から、長方形で板書の角の 4ヶ所を外部ツールで指定するようにしていたが、それと同様に板書領域の内側の 4ヶ所を外部ツールで座標指定することで、講師を認識する領域の指定が可能とした。図 4 に、板書領域と認識領域を指定する外部ツールを示す。認識領域を狭くすると移動物体の量のヒストグラムが平らに

なり、最大値が複数出現することが多くなる。そこで、ヒストグラムの最大値が複数ある場合、その中で最大値の平均を計算してから講師の中心位置を計算することにした。

## 7. 実装 2

### 7.1 実験内容

実験 2 は、板書領域全体を認識対象とする既存の Chalk-Talk と、講師を認識する領域を指定できる機能を組み込んだ ChalkTalk を比較し、講師追従の正確性に違いがあるのかを検証する。また、認識領域の最適範囲を調査するため、異なった 7 種類の範囲指定をしたうえで講師動画とデバッグ画像を作成して誤作動の比較を行った。

実験には動画 A を用いた。講師の認識領域の指定方法を以下に示す。領域指定は板書の左右辺と、図 5 のいずれかの 2 つの線を上下辺とする。

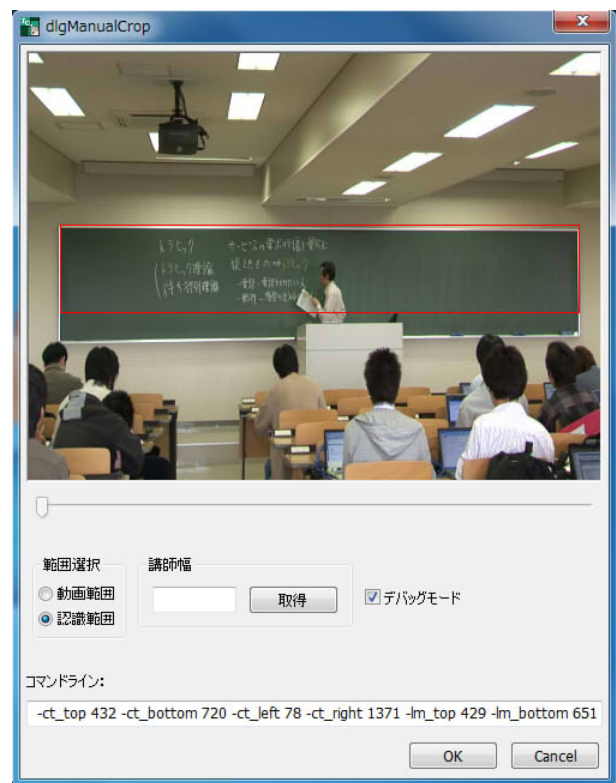


図 4 領域指定の例

Fig. 4 Example of specifying the area

表 4 認識領域による違いの分析

Table 4 Analysis of differences in recognition area

定義	講師動画	デバッグ画像
A	度々フォーカスから講師が外れる誤作動が発生	頻繁に受講者の頭に誤反応している
B	一ヶ所だけ誤作動が発生	講師領域が多少講師を捉えきれていない
C	良好である	頻繁に受講者の頭に誤反応している
D	一部左右に揺れる誤作動が発生	講師領域が講師を捉えきれていない
E	良好である	ほとんど良好である
F	良好である	良好である

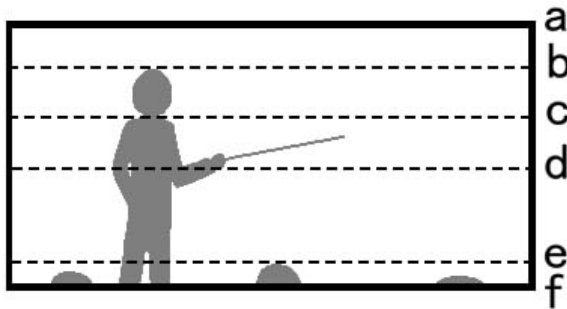


図 5 認識領域の例

Fig. 5 Examples of recognition area

動画 A

板書全体を指定 (a から f)

動画 B

受講者の頭を除く様に指定 (a から e)

動画 C

講師の頭のみを認識するように指定 (b から c)

動画 D

講師の肩のみを認識するように指定 (c から d)

動画 E

定義 B から講師より上を除くように指定 (b から e)

動画 F

定義 E から講師の頭を除くように指定 (c から e)

また、総合的な性能を比較するため、既存の ChalkTalk と提案 1 の機能を組み込んだ ChalkTalk, すべての提案機能を組み込んだ ChalkTalk で講師動画を作成し、誤作動の回数の比較を行った。用いた講師動画は動画 A である。認識領域の指定は、上記の中で最も精度の高い定義で指定するようにした。

7.2 実験結果

認識領域による誤認識の回数の違いを表 4 に示す。定義 A は既存 ChalkTalk と同等であるため、誤作動に変化はなく受講者の頭や板書の文字などに誤反応していた。定義 B は、受講者の頭を認識しなくなったがしばしば誤作動があった。定義 C, D は両方ともほとんど誤作動が発生しなかったが、講師の腕や体の一部を認識しておらず、講師領域がかかっていないため板書静止面に頻繁に講師の姿が

表 5 誤作動の回数

Table 5 The number of malfunctions

	既存	提案 1	提案 1 + 提案 2
回数	27	9	2

残っていた。定義 E, 定義 F はほぼ誤作動がなかったが、定義 E は多少板書の文字を誤認識している部分があった。そのため最も精度の高い定義は、定義 F であることと定めた。

誤作動回数の違いを表 5 に示す。提案 1 と提案 2 を組み合わせ、認識領域は定義 F の通りに指定することでかなりの誤作動を削減できた。

8. おわりに

8.1 実験結果

実験 1 の結果から、実装 1 でも受講者の頭に誤反応していた問題をある程度は対策ができていた場合があった。講師と受講者の頭の両方が移動物体として認識されると講師領域が大きくなり、誤認識として処理されたためと考えられる。しかし講師が動いていない、板書前にいないなど講師を移動物体と認識していない場合はどうしても誤認識を防止できなかった。しかし、実装 2 で受講者の頭や板書の文字など誤認識するものを認識対象から確実に外すことで誤作動数をかなり削減することができた。したがって、提案した手法が有効であると考えられる。

最良の認識領域の指定方法は、受講生の頭、板書の文字が比較的多い部分やスクリーンなどの誤認識が発生しそうな部分を除くように指定することであることがわかった。また、認識領域が狭すぎると講師の全体を認識しきない場合があるため、なるべく最も移動体が検出されやすく、かつ両腕を認識範囲に含めうように指定することが最良と考えられる。

8.2 実験結果

今回、問題が解決できなかった講義動画は板書前に講師とは別に人物が立って出てくる動画があった。フレーム間差分法では、どちらが講師か判断できないことができないため、講師位置が定まらなくなる。また講師と人物の背丈が

ほぼ同等なため、認識領域の指定でも対策することができなかつた。また、講義中に板書上にスクリーンを降ろしたりするようなことがあると対策することができないと思われる。この改善策として、講師が移動した軌跡を考慮して講師の現在の位置を把握する、または顔や色を利用した新しい講師の検出方法を模索する必要があると考えられる。

#### 参考文献

- [1] 市村, 井上, 宇田, 伊藤, 田胡, 松下: ChalkTalk: 講師動画と板書静止画の同時記録が可能な講義自動収録システム, 情報処理学会論文誌, Vol.47, No.3 pp.924-931(2006).
- [2] 市村, 福井, 井上, 松下: Web 学習用講義コンテンツを自動作成する板書講義収録システム, 情報処理学会論文誌, Vol.47, No.10, pp.2938-2946(2006).
- [3] 株式会社アーネット, [http://www.earnet.co.jp/jp/\(\)](http://www.earnet.co.jp/jp/).
- [4] 大西, 泉, 福永: 講義映像における板書領域のブロック分割とその応用, 電子情報通信学会論文誌 (D-I), Vol.J83-D-1, No.11, pp.1187-1195(2000).
- [5] Liu, Q., Rui, Y., Gupta, A., Cadiz, J.: Automating Camera Management for LectureRoom Environments, Proc. Of ACM Conference on Human Factors in ComputingSystem (CHI 2001), Vol.3, pp.442-449 (2001).
- [6] Zhang, C., Crawford, J., Rui, Y., He, L.: An Automated End-to-End Lecture Capturing and Broadcasting System, Proc. Of ACM Multimedia 05, pp.808-809 (2005).