

講師映像と板書静止画の記録が可能な講義自動収録システム

○市村哲; 富野剛; 井上亮文; 松下温
東京工科大学

電子プレゼンテーションが普及した現在でも、大学等の教育機関において黒板を用いた講義は根強い支持を得ている。著者らは、講師が講義前にビデオカメラを 1 台設置しておくだけで、その収録映像から自動的に講師の動画映像と板書の静止画とを作成し、即座にインターネット配信可能とする講義自動収録システム ChalkTalk を構築した。板書静止画については、画像処理を施して講師の姿を消去する工夫を施している。1 台の固定ビデオカメラのみで黒板全体を記録することができるように解像度の高い民生用ハイビジョンカメラを用いて実装した。

Recording Chalk Talks into a Movie and Still Images.

Satoshi Ichimura; Takeshi Tomino; Akifumi Inoue; Yutaka Matsushita
Tokyo University of Technology

Although lectures and seminars using presentation software became popular, a chalk talk is still commonly used among education institutions such as schools and universities. We developed ChalkTalk that automatically produces E-learning materials from videotaped chalk talks. The system separately extracts a lecturer's image and writing on blackboard from video images recorded with a single high-definition digital camcorder, and stores the former as a streaming video and the latter as a series of snapshot images.

1. はじめに

近年、インターネットを利用した新しい情報サービスとして WBT(Web-Based Training)などの E-ラーニングが注目されている。E-ラーニングの利点として、各受講者がいつでもどこでも、自分の進捗状況に合わせて学習を進められることや、通学費や施設費などの経費を削減できることがある。最近では、講師映像のストリーミングビデオファイルと電子プレゼンテーションファイルを予めサーバにアップロードしておき、受講者は、講師映像の再生に同期して進められるプレゼンテーションスライドを見て学習するという、マルチメディア対応のシステムも登場している。

しかしながら、電子プレゼンテーションやビデオが普及した現在でも、大学や企業教育機関等において黒板を用いた講義は根強い支持を得ている。電子プレゼンテーション中心の講義であっても、重要箇所の説明の際や、受講者の理解ペースにあわせて説明する必要がある際に板書が頻繁に用いられている。また、大学や予備校において、数学や英語では、ほとんどが黒板を用いて授業が行われているのが現

状である。著者らが実際に学内の大学講義を調査した結果、調査した講義の半分以上の講義において何らかの用途で黒板が利用されていることがわかった。また、調査した受講者から、電子プレゼンテーションのみの講義は、「進行スピードが早くノートがとれない」、「その場でわかった気になるが頭に残らない」、「短時間に多くのことを説明されるので理解が追いつかない」などの問題を指摘する意見が多く聞かれた。さらに、調査したある講師からは、「電子プレゼンテーションが学生から不評だったため今年度開講講義から板書に戻した」という意見も得られた。

以上のような背景から、著者らは、黒板の板書を用いた講義を E-ラーニング教材として提供できるようにすることが必要であると考え、これを目的として研究を実施した²⁾。本論文では、講師が講義前にビデオカメラを 1 台設置しておくだけで、その収録映像から自動的に板書の静止画と講師の動画映像とを作成し、即座にインターネット配信可能とする講義自動収録システム ChalkTalk について報告する。ChalkTalk の設計においては、専属カメ

ラマンを必要とせず、また、講義を担当する講師が極めて簡単に利用できることを重要目標とした。また実装にあたっては、可搬性を重視し、1台の固定ビデオカメラのみで板書全体を記録することができるよう、解像度の高い民生用ハイビジョンカメラを利用した。本論文では、システムデザイン、実装、および、評価について述べる。

2. システムデザイン

ビデオカメラによって講義映像を記録しようとすると、講師撮影映像の一部として板書を記録する必要があるが、黒板全体をカバーする領域を動画として保存またはインターネット配信しようとするネットワーク帯域を多く必要とするという問題点がある。また、1台のビデオカメラのみでは黒板全体を撮影するのに十分な解像度を得ることができないという問題がある。

このような理由から、講義における板書の自動撮影に関する研究の一環として、複数のカメラ映像の中から最も講義状況に適した映像を選択して記録する研究や、固定カメラとズーム制御可能なカメラとを併用し重要箇所に限定して板書を自動撮影する研究がなされている^{3,4,5}。しかしながら、固定カメラとズーム制御可能なカメラとを併用しなければならぬために装置が大掛かりとなり、どの教室にでも手軽に持ち運んで使えるシステムとすることや、ほとんどセットアップ無しで使うことができるシステムとすることが難しいという問題があった。

著者らは、講師の映像に比べて板書の映像は極めて変化が少ないことに着目し、講師の映像は動画として保存し、板書の映像は静止画として保存することで上記問題点を緩和できると考えた。しかしながら、板書を静止画として保存する際には、講師の陰となつて見えない部分が存在しないようにしなければならない。なぜならば、動画であれば、板書の一部が講師の陰となつて見えない場合にでも、少し待てば講師がその部分の前から立ち去るために問題とならないが、板書を静止画として記録してしまうと、その見えない部分がいつまでも読めないままになってしまうためである。このようなことから、収録映像から板書の静止画と講師の動画映像とを分離する際に、画像処理を施して板書静止画から講師の姿を消去し、講師の陰のない静止画を作成して保存できるようにした。

さらに、ハイビジョンカメラを用いることで黒板全体を撮影するのに十分な解像度を確保し、1台の固定

ビデオカメラ映像のみから講師の動画映像と板書の静止画とを作成できるようにすることができないのではないかと考えた。今回ハイビジョンカメラとしては、記録画素が1440×1080（約155万画素）確保できる民生用のSONY「HDR-FX1」⁶を使用した。HDR-FX1は、HDV規格（DVカセットテープにハイビジョン映像の記録再生を可能にするための規格）に準拠したデジタルビデオカメラであり、映像記録フォーマットはMPEG2、音声フォーマットはMPEG1 Audio LayerIIである。また、このハイビジョンカメラの撮影映像は16:9とワイドであり、一般的な4:3のDVカメラに比べて、横長の黒板を無駄なくカバーできる利点がある。

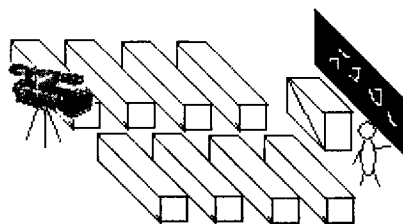


図1 教室内のビデオカメラ配置

システムの典型的な使用手順について述べる。講師は、講義開始前に教室の後ろにビデオカメラを設置し、黒板全体を撮影できるようにビデオカメラを調整して撮影を開始する（図1参照）（音声収録には、内臓マイクロフォンでは十分に音声を拾えないことが予測されたために、外部マイクを取り付けて利用した）。

講義終了後、講師は撮影に使ったビデオテープを巻き戻し、ビデオカメラとPCとをIEEE1394ケーブルで接続して、ビデオテープを再生しながら本ソフトウェアを実行する。これだけの操作で、ビデオテープに記録された講義映像より、講師の姿を消去して常に黒板の板書が見える状態を保った静止画として板書を保存し、同時に、講師を自動追尾した映像を動画として保存できるようになっている。なお、ビデオテープ中の映像を一旦すべてPC上にデジタルビデオキャプチャし、このPC上のキャプチャファイル（MPEG2-TS形式）を本ソフトウェアに入力することによっても同等の処理が可能となっている。

3. 実装

3.1. 講師位置の検出

講師の陰となって見えない部分が存在しない板書静止画を作成するために、映像から講師の存在する領域を検出し、次に、その領域の画像を、講師が存在しなかった時点の画像(講師がその領域に入る直前の画像)によって置き換えるようにした。ここでは、映像から講師の存在する領域を検出する処理について述べる。

講師の存在する領域を検出するために、まず背景差分法を導入することを試みた。背景差分法は、画像処理によって画像中の移動物体を検出する手法の1つであり、予め用意しておいた背景画像と入力画像の差分を取ることで動領域を抽出する方法である。背景差分法は高精度に動領域を抽出できるという利点がある反面、背景の明るさが時間的に変化する(すなわち背景画像が変化する)屋外などでは使用が難しいという欠点がある。しかしながら、講義収録の場合は、背景の明るさがほぼ一定と考えられる教室での撮影のため、著者らは、背景差分法が有効に働くであろうと判断した。

本実装では、講義開始前の映像を背景画像とし、講義中に撮影した入力画像と背景画像の輝度を比較して、その明度差に基づいて移動物体を抽出するようにした。実際には、背景差分法で得られた結果を閾値により2値化して移動物体の検出を行った。ここで、画像 $f(x,y)$ と画像 $g(x,y)$ との差分画像 $h(x,y)$ は以下の式で定義される。すなわち、2つの画像と同じ位置座標における輝度値の差により得られたものが差分画像である。

$$h(x,y) = |f(x,y) - g(x,y)|$$

具体的には、講師が黒板の前に立っていない画像を背景画像 g とする。そして、その背景画像 g と現在の画像 f との差分画像を求めることで、動く講師の領域を求めることができる。講師が映っていない領域の画素は基本的に背景画像 g と同一であるため、現在の画像 f から背景画像 g を引き算した差分画像 h の当該画素値は0となる。この差分画像 h に二値化を施せば、背景と異なるものが映っている画素は1に、背景とほぼ等しい輝度のものが映っている画素は0になる。実際にシステムに背景差分法を実装して講義撮影を行った結果から、ほぼ正確に講師位置を特定できることがわかった。

しかしながら、時間が経つと黒板消しの位置や板書

内容が変化するため、この方法のみでは講師位置を特定することは難しい。背景差分法を用いて黒板前の講師の位置を検出した結果を図2に示す(わかりやすいように、背景と異なるものが映っている画素は現在画像 f の画素を表示している)。これは板書が書き加えられた時や消された時、または、黒板消しが移動した時に算出された差分画像である。図の中央右側下部には動かされた黒板消しが移動物体として認識されており、また、講師が書き加えている板書も移動物体として認識されて講師と分離できなくなっている。



図2 背景差分法の利用

そこで、この問題を解決するために背景差分法に代えてフレーム間差分法によって講師位置を特定するようにした。フレーム間差分法は、撮影された時間の近い2つの画像で差分を計算する方法であり、画素変化のあった領域を移動物体が映っている領域と定める方法である。ここで、時刻 t に撮影された画像 $ft(x,y)$ とその1時刻前 $t-1$ に撮影された画像 $ft-1(x,y)$ との差分画像 $ht(x,y)$ は以下の式で定義される。

$$ht(x,y) = |ft(x,y) - ft-1(x,y)|$$



図3 フレーム間差分法の利用

フレーム間差分法を実装して講義風景を撮影する実験を行った。背景差分法を用いた時に問題となっていた、板書や黒板消しを移動物体として認識してしまう不具合は解消された。しかしながら、フレーム間差分法では、動いている物体の領域としては、時刻 t における位置と、時刻 $t-1$ における位置の両方が抽出されてしまうという問題がある。実際に、講師が早く移動したような場合に、時刻 t における位置と時刻 $t-1$ における位置の両方の領域が移動領域として抽出されてしまう不具合が観察された。ま

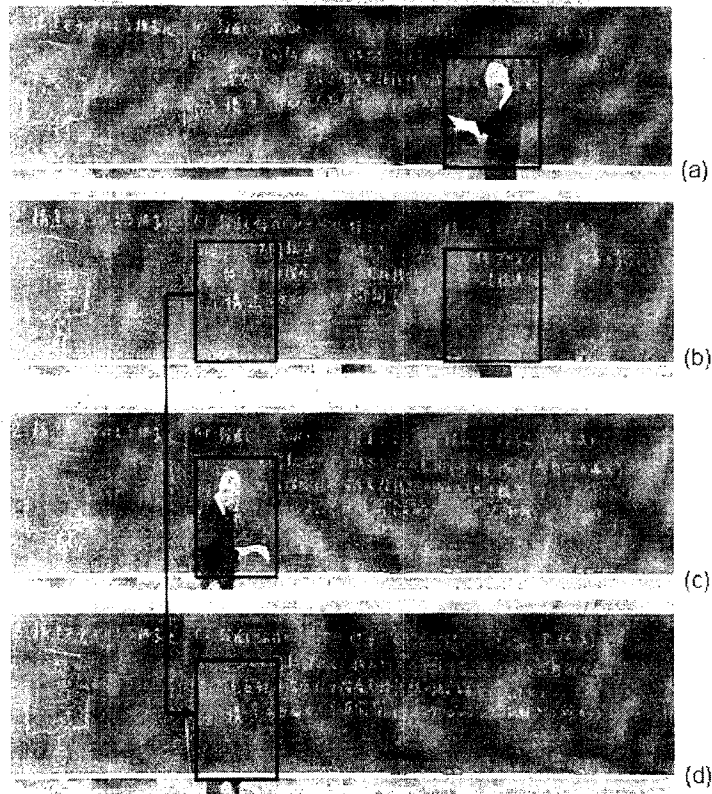


図 4 講師を消去した板書静止画の生成

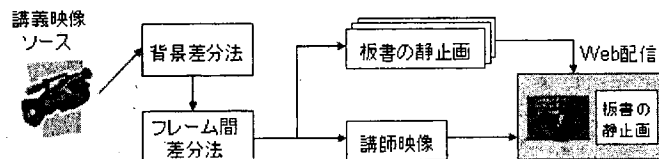


図 5 システム動作概要

た、講師の動きが遅い場合にも、移動物体が部分的に重なっているため、移動物体の輪郭が抽出されにくいという問題点があることが観察された（図3参照）。

そこで著者は、背景差分法とフレーム間差分法を併せて導入するようにシステムを変更した。具体的には、講義開始前の映像フレームを背景画像として背景差分法により差分画像を生成し、次に、その差

分画像と、その直前の差分画像とのフレーム間差分を求めて講師の輪郭（移動物体の輪郭）を抽出するようにした。また、背景画像と認識される画像が定められた時間内に変化しなかった場合に、背景画像をその時点の講師映像除去画像（次に述べる方法によって講師映像を除去した画像）に差し替える処理を行うようにした。

結果として、フレーム間差分法のみによる方法で検

出された講師位置よりも正確に講師位置を検出することができた。そしてシステムは、抽出した講師の輪郭（移動物体の輪郭）の外接四角形を求め、当該四角形内の画像を動画として記録すると同時に、当該四角形以外の画像を静止画として記録するようになっていた。

3.2. 映像からの講師画像の除去

映像から講師位置を検出した後、その領域の画像を、講師が存在しなかった時点の画像（本実装の場合は、8秒前の講師映像除去画像）によって置き換えるようにした。これによって、あたかもその時点で講師が存在しなかったかのような画像を作成している。

図4において、入力画像(a)の講師を消去した画像が(b)であり、入力画像(c)の講師を消去した画像が(d)である。すなわち、(a)と(b)、および、(c)と(d)はそれぞれ同一時刻の画像フレームである。また、(a)と(c)、および、(b)と(d)はそれぞれ時間的に隣接した画像フレームである。(c)は(a)の8秒前の入力画像であり、(d)は(b)の8秒前の講師映像除去画像である。

システムは、画像(a)から講師の領域を特定し、その領域にあたる一つ前に処理された画像((b)より8秒前の講師映像除去画像)を代わりに表示することで、(b)のような講師を消去した静止画を得ている。また次に処理された(d)の静止画でも、(c)の画像から講師の領域を特定し、ひとつ前に処理された(b)の該当箇所の画像((d)より8秒前の講師映像除去画像)を代わりに表示している。以上のような処理を繰り返すことで、常に、講師を除去した画像を作成・保存することが可能となっている。図5に実装システムの動作概要図を示す。

なお、8秒という時間間隔は実験によって求めた最適値である。静止画の保存容量のことを考慮すると、静止画を作成する時間間隔を極力長くすることが望ましいが、例えば、時間間隔を12秒に設定した場合には、講師が板書してそれを消すタイミングが速い場合に、保存されない板書が見受けられることがあった。すぐ消される板書であれば、重要度は低いと考えられるが、そのようなことが極力起きないようにする必要があると考えた。8秒というタイミングは、「講師が通常の板書をし、学生に説明して消すまでにかかった時間」の最短時間であり、講義撮影ビデオを観察しこの値を定めた。

3.3. 板書静止画と講師動画の作成と配信

本システムでは、講師映像除去画像が作成された瞬間に1つ前に処理された講師映像除去画像との比較が行われ、画像全体を対象とした差分値が閾値を超えた場合に、JPEG形式で板書静止画ファイルを保存するようになっていた。この際、時系列を反映させたネーミング規則でファイル名を変更し、順次同一ディレクトリに蓄積するようにした。

一方、講師の領域の動画ファイルは、Windows Media ストリーミング形式でファイルに保存されるようになっていた。板書を動画撮影する場合は、その板書が読めるようにするための高い解像度が必要となり、高いビットレートでエンコードしなければならないと想像されるが、本システムの場合には、動画に含まれるのは講師映像だけであるので、それ程高い解像度が求められない。本実装においては、画像サイズを320x240とし、700Kbpsのビットレートでエンコードしてストリーミング動画(30フレーム/秒)を作成するようにした。このストリーミング動画ファイルにはCD音質の講義音声も併せて記録されている。

講義コンテンツを受講者に配信する際には、基本的に、作成された静止画ファイルと動画ファイルを、ストリーミングサーバ上の公開ディレクトリにコピーするだけでよい。そして、受講者が講義コンテンツを利用する際は、各自のWebブラウザからインターネット経由で利用することができる。受講者が用いるWebブラウザ上で動作するJavaScriptが、講師映像の動画再生時点を一定間隔で検出し、この再生時点に対応したファイル名のJPGファイルをストリーミングサーバから随時ダウンロードして表示している。これにより、講師映像の再生に同期して板書スライドが自動的に切り替わるようになっていた。

ハイビジョンカメラからのMPEG2-TS映像入力をPC上で処理するために、Windows用マルチメディア拡張API群であるDirectShowを用いた。本システムの機能は、DirectShowの画像処理フィルタを実装することによって実現されている。また、ストリーミングサーバとしてはWindowsサーバ2003を用いた。

4. 評価

ChalkTalkでは、従来動画として記録されていた板書映像を、静止画で記録することによって記憶容量を削減することを主要な目的の1つとしている。

そこで、実際にどれだけの容量が削減できるかを実際の講義を撮影して評価した。

実験対象とした講義は、板書主体で進められた 30 分の講義である。撮影の条件としては、黒板の横幅全体がちょうどハイビジョンカメラ映像の横幅全体になるように画角を調節して撮影した。また静止画作成の条件としては、前述の通り静止画を作成する最短時間間隔を 8 秒と定めて処理を行った。また、ハイビジョン映像を MPEG2 動画(1440x1080 ピクセル)で記録した場合には記憶容量が 5.7GB になることを調べておいた。さらに、この 5.7GB の MPEG2 動画を単純に Windows Media ストリーミング形式に変換した場合、この動画ファイル(1440x1080 ピクセル)のサイズが 1.5GB になることを調べておいた(板書が読める解像度を得るために必要最低限の 7Mbps でエンコードした)。

実験の結果、30 分の講義の中で 80 枚の板書画像(1440x1080 ピクセル)が保存された。保存された全静止画ファイルのファイルサイズの合計は 40MB であった。一方、講師の領域の動画ファイルのサイズは 90MB となった。作成された全静止画ファイルと動画ファイルの合計 130MB と、MPEG2 動画で記録したファイルの 5.7GB とを比較すると、約 1/44 に帯域を圧縮できることがわかる。また、MPEG2 動画を単純に Windows Media ストリーミング形式に変換した 1.5GB の動画と比較しても、約 1/12 に圧縮できることがわかる。この結果から、ChalkTalk には帯域圧縮の効果があると判断した。

なお、比較実験の都合上、保存する静止画サイズを 1440x1080 ピクセルとしたが、実際には黒板以外の部分は必要でないために黒板の上下で切られた部分の静止画のみを保存するようになっている。また、受講者が用いるディスプレイによっては 1440 ピクセル幅の画像を表示できないため、1280 ピクセル幅や 1024 ピクセル幅などの縮小静止画を生成する機能も備えている。

次に、被験者に対し、講師と板書の両方を 1 台の DV カメラで手持ち撮影した映像と、ChalkTalk によって自動作成された講義コンテンツの両方を見せ、その後、それぞれの見やすさについてヒアリング調査を行った。

結果として、DV カメラで撮影した板書映像に比べ、ChalkTalk が作成した講義コンテンツの方が見やすいとの意見を被験者全員から得ることができた。

講師を消去することで黒板の死角をなくしていることや、ハイビジョンカメラで撮影しているため常に黒板全体を見渡せることが見やすさにつながったと推測できる。他にも、ChalkTalk が作成した講義コンテンツの利点として、「ちらつきなどが全くないので落ち着いて見られる」、「静止画をコピーして保存できるので便利」などの意見が得られた。講師動画については、320x240 と解像度をかなり落として記録したが、それを問題として指摘する被験者はいなかった。

5. おわりに

板書を使用する講義の映像から、マルチメディア WBT 教材を自動的に作成する講義収録システム ChalkTalk を提案した。

講師が殆ど動かなかった場合に、フレーム間差分法を利用している性質上その特定が曖昧になってしまうという問題がある。これの解決法としては、講師領域が急に消えたり減少してしまったりした場合に、その場所に講師が止まっていると予測をし、前に講師の存在した領域を再度講師領域と定めるなどの工夫が必要であると考えられる。また、講師が横に腕を大きく広げたような場合、その腕が表示されてしまうことがあった。板書を遮るほどの大きな問題はないものの、見え方に違和感があるため改善が必要であると考えられる。

参考文献

1. 板宮, 林, 千代倉: ワンマン録画可能な講義ビデオ作成システム, 情報処理学会コンピュータと教育 研究報告, No.70, pp.17-20, (2003).
2. 森田, 井上, 市村, 松下: 講義自動収録システムにおける板書静止画記録法, 情報処理学会全国大会, 3Z-5, (2005).
3. 大西, 村上, 福永: 状況理解と映像評価に基づく講義の知的自動撮影, 電子情報通信学会論文誌 (D-II), vol.J85-D-II, no.4, pp.594-603, (2002).
4. 大西, 泉, 福永: 情報発生量の分布に基づく遠隔講義撮影の自動化, 電子情報通信学会論文誌 (D-II), vol.J82-D-II, no.10, pp.1590-1597, (1999).
5. 宮崎, 亀田, 美濃: 複数カメラを用いた複数ユーザに対する講義の実時間映像化法, 電子情報通信学会論文誌 Vol.J82-D-II, No.10, pp.1598-1605, (1999).
6. <http://www.sony.jp/products/Consumer/handycam/PRODUCTS/HDR-FX1/>