イメージモザイキングによる講義のディジタルアーカイブと再生

横井 隆雄　　遠山 聖司　　藤吉 弘亘
中部大学工学部情報工学科

1　はじめに

近年，コンピュータとネットワークを用いた教育機会を提供する e-learning システムの導入が増加している．従来の講義と同様の教育効果を得るためには，講師の声と動き，板書，質問等の構成要素を e-learning 環境においても再現する必要がある．特に，板書は書くスピードや文字の大きさ等で内容の重要度を伝えるために必要不可欠な構成要素である．しかし，既存の e-learning システムにおける映像コンテンツでは，一台のカメラにより講師を中心に撮影されるため，黒板の板書文字を読みとることができない．本研究では，黒板の板書を含む講義のディジタルアーカイブとその再生を目的とする．本稿では，非同期カメラ (DV カメラ) で撮影された複数の映像からイメージモザイキングによる高解像度映像の生成と音声と講師の動きを用いた講義のインデキシングに関する基礎検討について報告する．

2　講義のディジタルアーカイブ

本システムでは，市販の DV カメラ複数台を用いて，黒板全体を含むように多視点映像を撮影し，イメージモザイキングにより一枚の高解像度映像を生成することで，板書を含む講義全体のディジタルアーカイブを行う．さらに，生成したモザイク映像から講師の動きや発話区間を検出し講義のインデキシングを行う．これらのデータをインターネットを介してストリーミング配信を行うことで，ユーザは講義をいつでもどこからでも再生することが可能となる (図 1 参照)．

2.1　音声信号による同期フレームの検出

撮影に用いる 3 台の DV カメラ間は同期されていないため，モザイク処理を施す前にカメラ映像間の同時刻のフレームを検出する必要がある．一般に動画像は 1 秒間に 30 フレーム，音響信号は 48kHz でサンプリングされるため，同期フレームの検出には時間分解能が高い音響信号を用いる．まず，基準カメラ映像の音響信号から，一秒間の平均パワーを求め，そのパワーが高い区間を講師の音声信号とする．そして，講師の音声信号が他のカメラ映像の音響信号内のどこに存在するかを探索する．探索の際には，音響信号を 11.025kHz にダウンサンプリングした信号からフーリエ変換により短時間パワースペクトルを求め，その類似度を計算する．類似度が最も高いフレームを同期フレームとする．

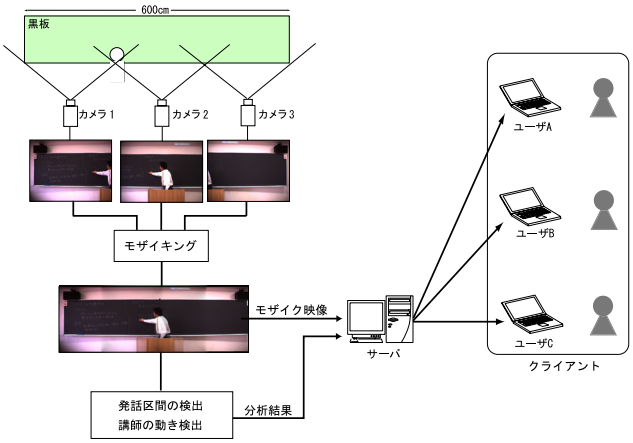

図 1: システム構成図

2.2　モザイク処理

各カメラ画像上の特徴点の対応より，平面射影変換行列 H を予め求めておく [1]．カメラより得られた 720×480 画素の動画像を平面射影変換により，1,600 × 480 画素のモザイク高解像度映像を作成する．隣接する画像の重複部分は，接合部が目立たないように濃度値をブレンドする．

3　講義モザイク映像の分析

ディジタルアーカイブされた講義映像データを閲覧する際に，全てを繰り返し再生すると非常に時間がかかるため，必要な部分を選択的に試聴する機能が必要となる．そこで，講義モザイク映像より講師の動きと発話区間の検出を行い，これらをインデックスに利用する．

3.1　講師の動きの検出

黒板領域の時系列画像に対して，フレーム間差分により，講師の動きを自動検出する．さらに，フレーム間差分により検出したピクセル群からセグメンテーションにより領域を求める．セグメンテーション領域が多く存在する場合は面積が最も大きい領域を講師領域と決定し，その座標を記憶する．この講師領域の座標点は，**4.1** で述べる講義の再生時の拡大表示にも用いる．

3.2　発話区間の検出

発話区間の検出には，ダウンサンプリングした音響信号の振幅対数パワーを求め閾値処理により検出する．3 人の講師による講義映像 (各 10 分間) の発話区間の検出

Digital Archiving by Image Mosaicing for Lecture and Its Replay
Takao Yokoi, Seiji Tohyama and Hironobu Fujiyoshi
Department of Computer Science, College of Engineering, Chubu University

結果を表 1 に示す．

表 1 発話区間検出結果

	再現率 (%)	適合率 (%)
Movie1	84.7	97.9
Movie2	78.4	99.2
Movie3	83.2	92.2
平均	82.1	96.4

実験の評価には，次式に示す再現率と適合率を用いる．

$$\text{再現率} = \frac{\text{正検出}}{\text{正検出} + \text{検出洩れ}}, \quad \text{適合率} = \frac{\text{正検出}}{\text{正検出} + \text{誤検出}}$$

表 1 より，約 96.4%の適合率を得ることができた．Movie3 の適合率がやや低い理由は，講義中にカメラの近くの生徒の咳や話し声を講師の発話と誤検出したことが原因である．

3.3 講義のインデキシング

講義における講師の状態は，“説明のみ”，“板書のみ”，“説明+板書”，“その他 (見回り)”の 4 状態を遷移していると考えられる．これらの状態遷移を抽出することは，講義のインデキシングに有効である．そこで，本稿では講義のインデキシングの基礎検討として，講師の動きと発話区間について検討した．図 2 に，講義の一部 (50[sec]) に対して振幅対数パワー，閾値処理により検出した発話区間とフレーム間差分で求めた講師の動きの有無を示す．図 2 より，発話の有無と講師の動きから “板書のみ”と “その他 (見回り)”の状態を推定することが可能であることが分かる．現在，黒板画像，講師の動き分析と音声信号を複合的に用いた “板書”状態の検出について検討している．

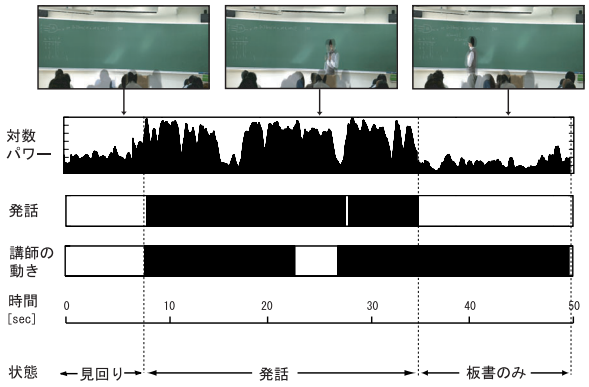

図 2: 講義における発話区間と講師の動き検出例

4 講義の再生

ノート PC の表示解像度は一般に XGA(1,024×768) が多く，高解像度のモザイク映像を表示することができない．そこで，図 3 に示すようにユーザが黒板全体と講師の位置関係が把握できるように縮小表示 (960×384 画素) する．同時に板書した文字が読めるように，ユーザがマウスクリックした点を中心に拡大表示と (図 3 右上) と講師の動きに追従して拡大表示 (図 3 左上) する．

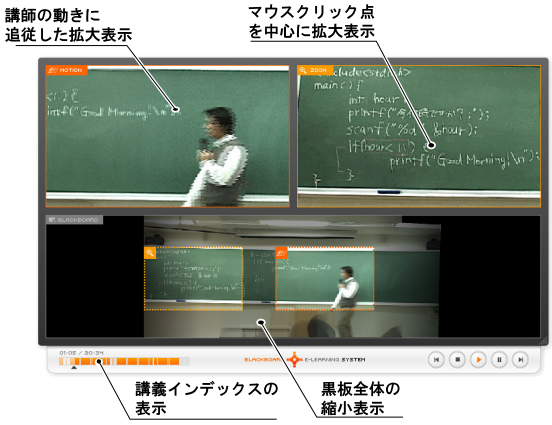

図 3: 講義の再生ビューア例

4.1 講師の動きに追従した拡大表示

3.1 で求めた講師の画像領域左端を中心に拡大表示する．この拡大中心点は，フレーム毎の検出結果を基に決定するため，講師の敏速な動きに追従して拡大画像が激しく揺れる場合がある．この急激な変化を抑制するため，拡大中心点の座標を次式に示す IIR フィルタを用いて計算する．

$$P_t = \alpha P_t + (1-\alpha)P_{t-1} \qquad (0 < \alpha < 1)$$

ここで P_t は拡大中心点の x 座標，P_{t-1} は 1 フレーム前の拡大中心点の x 座標，α は拡大中心点の座標をどれだけ反映させて更新するかを決定する定数である．図 4 は拡大中心点の x 軸座標を求める際に検出領域の左端とした場合と，IIR フィルタによる抑制をした場合の結果である．IIR フィルタ (α=0.05) を用いると拡大中心点が滑らかに変化しているのがわかる．作成した講義の再生ビューアでは，この α をユーザが任意に設定することができる．

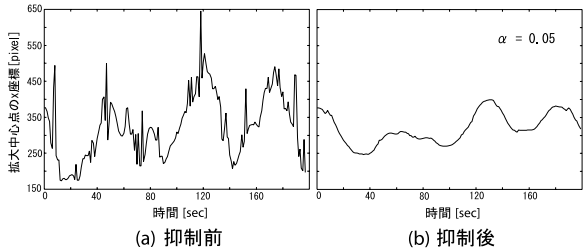

図 4: 拡大中心点の制御

5 まとめ

本稿では，DV カメラで撮影した多視点動画像からイメージモザイキングにより板書を含む講義のディジタルアーカイブとその再生について提案した．今後は，講義モザイク映像を分析し，より正確な講義のインデキシングと評価を行う予定である．

参考文献

[1] 金澤靖，太田直哉，金谷健一, “射影変換行列の最適計算によるモザイク生成”, 情報処理学会研究報告, 99-CVIM-116-2,pp. 9-16,1999.