

1. はじめに

e-learning において、撮影した講義の配信は、遠隔地等での教育格差の是正に期待できる。しかし、多くの講義をカメラマンが撮影することはコスト面に問題がある。この問題に対して、講義の自動撮影システムが期待されている。また、講義ビデオを復習のために再生する場合、重要な講義要素のみを視聴できることが望ましい。そこで、本研究では、HDV カメラで撮影した高解像度映像から、放送カメラマンが撮影したかのような講義映像の自動生成と講義イベント検出による短縮講義ビデオ生成法について提案し、その有効性を示す。

2. 講義アーカイブシステム

既に提案されている自動講義撮影システムでは、複数のカメラやセンサが用いられている [1, 2]。しかし、複数のカメラやセンサを設置するため、特別な講義室が必要となる。また、カメラの雲台を自動制御するため、その映像は無機質なものである。臨場感あふれる講義映像は学習効果に重要であり、滑らかな映像を生成する必要がある。本研究では、図 1 に示すように、高解像度映像から仮想カメラワーク生成と短縮講義映像生成により講義ビデオを生成する。

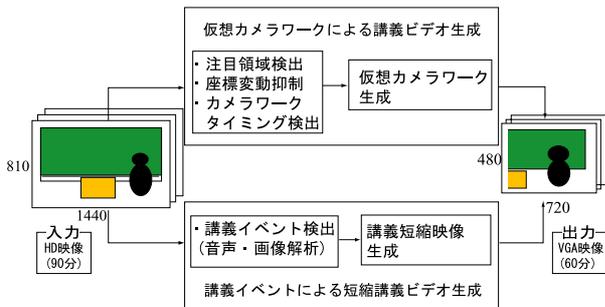


図 1: 講義アーカイブシステム

3. 高解像度画像からの仮想カメラワーク生成

本研究では、従来の問題点を解決するために、ハイビジョンカメラを用いて講義の撮影を行う。黒板全体が入り、かつ黒板の板書文字を読むことができるように一台の HDV(1080i) カメラを講義室の後方に設置し講義映像を取得する。しかし、閲覧者のユーザが持つノート PC の表示解像度は一般に XGA が多く、HDV カメラで撮影した高解像度映像 (1,440 × 810) をそのまま表示することができない。この問題を解決するために、図 2 に示すように講師等の注目対象に追従するように高解像度映像からトリミングを行い、講義ビデオを自動生成する。

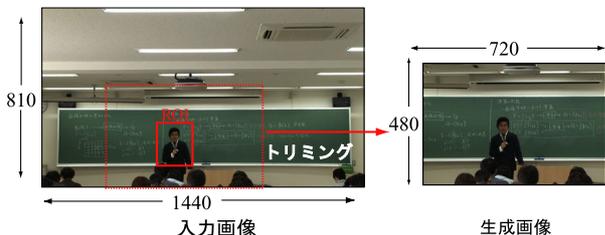


図 2: 高解像度画像からのトリミング画像生成

3.1 仮想カメラワーク生成

トリミングを行う際、放送カメラマンが撮影するような、仮想カメラワークを実現することで臨場感ある講義映像生成を行う。その手順を以下に示す。

step1 講師位置検出

高解像度映像から講師等の移動体領域をフレーム間差分法により求める。しかし、フレーム間差分によって得られた注目対象の位置座標は、講師の敏速な動きに追従して激しく変動する (図 3(a))。そのため、求めた座標値をそのままトリミングに使用するとゆれの激しい映像となる。そこで、

バイラテラルフィルタを用いて変動の抑制を行う。バイラテラルフィルタを用いることで、エッジを保存しつつ、細かな振動を抑制させることができる (図 3(b))。

step2 カメラワークタイミング

バイラテラルフィルタにより抑制された講師領域座標に対して、零交差処理を施し講師の動作特徴点を求める。隣り合う各特徴点間の位置座標の変化より、変化の激しい区間をパンニング区間、変化の少ない区間をズームング区間と判定する (図 3)。

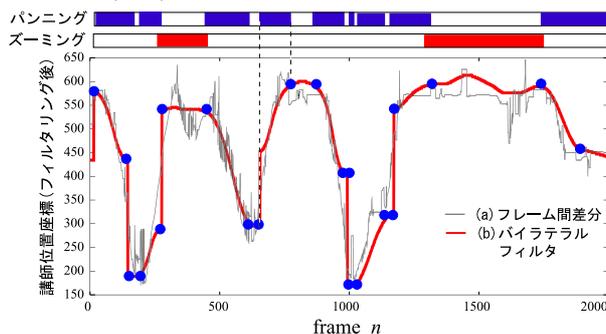


図 3: カメラワークタイミング

step3 仮想カメラワーク生成

検出した区間に対して、以下に示す放送カメラマンの撮影技術による知見から仮想カメラワークモデルを算出し、トリミング位置を決定する。それにより、臨場感あふれる講義映像を生成する

- パンニング速度曲線は非対称型である。
- 最大拡大変化率は、ズームイン時には後半に、ズームアウト時には前半に発生する。

4. 講義イベント検出による短縮講義ビデオ生成

3. で述べた高解像度映像からの仮想カメラワークによる講義映像の自動生成により、放送カメラマンの撮影技術に基づく臨場感ある講義映像の生成が可能となる。しかし、生成した講義ビデオを再生する際、重要な要素だけを視聴できることが望ましい。そこで、講義のイベント検出に基づく短縮講義ビデオの生成法について提案する。

4.1 講義におけるイベント検出

短縮講義ビデオを生成するために、講義における講師の発話と板書区間の検出を行う。

発話区間の検出

講義における発話検出は、事前に複数の講師の発話・無発話での音声特徴 (25 次 LPC ケプストラムとパワースペクトル 150Hz-900Hz) を抽出する。入力された音声データとのマハラノビス距離を求め、発話が無発話を判定する。

板書区間の検出

板書区間は、黒板上に新しく板書されたかどうかを判定する。新規板書の判定は、西口等が提案した入力画像から黒板上の講師部分を消去した生成画像間の差分を取ることで検出することができる [3]。しかし、[3] の手法は、フレーム間差分で得られた結果に対して、一定領域の大きさの矩形領域を講師領域とするため、得られる講師領域には背景ピクセルも含まれ、正確な板書タイミングを検出できない場合がある。そこで本研究では、以下に示す手順により板書検出を行う。

step1 現画像フレームに対して、フレーム間差分により得られた前景ピクセルを物体 seed とし、グラフカットにより講師領域のセグメンテーションを行い、物体マスクパターン O_M を得る (図 4)。

$$O_M = \begin{cases} 1 & : \text{object} \\ 0 & : \text{background} \end{cases} \quad (1)$$

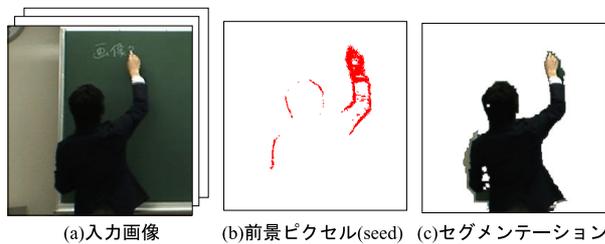


図 4: 講師領域のセグメンテーション
 step2 背景画像 I_B に対して, step1 で得られた講師領域以外のピクセル (黒板領域) のみを現ピクセルとして更新することで講師消去画像 I_C^t を次式により生成する (図 5(b)). ここで, I^t は現フレーム画像を表す.

$$I_C^t = I^t \cdot \bar{O}_M + I_B^t \cdot O_M \quad (2)$$

step3 過去の講師消去画像とフレーム間差分 Δ_t を求めることで新しく書かれた板書文字を検出する (図 5(d)).

$$\Delta_t = \sum_{(i,j) \in I} |I_C^t(i,j) - I_C^{t-1}(i,j)| \quad (3)$$

step4 step3 において, 一定時間, 黒板上に輝度変化が起これない場合板書が終了したと判断する. また, I_B^t は次式により更新する.

$$I_B^t = (1 - \alpha)I^t \cdot \bar{O}_M + \alpha I_B^{t-1}. \quad (4)$$

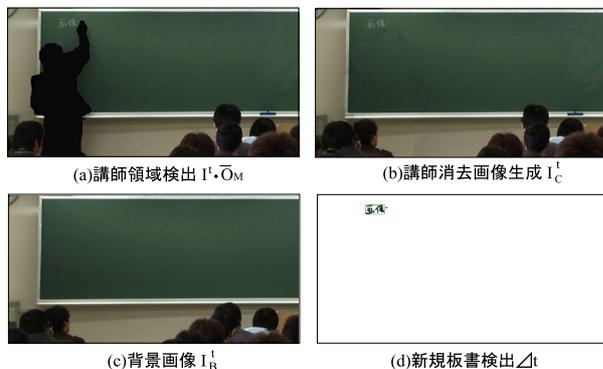


図 5: 板書検出

4.2 短縮講義ビデオの生成

4.1 で得られた講義イベントを用いて短縮講義ビデオを生成する. 短縮講義を生成するために, ブロック A や D (無発話, 無動作) のような何も講義要素が存在しない領域を消去する. また, ブロック F のように板書区間でも発話をしていない区間は, 講師の板書動作を確認できれば良く, 早送りが可能である. そこで, 検出された板書のみ区間は通常速度の 3 倍速でビデオの再生を行う.

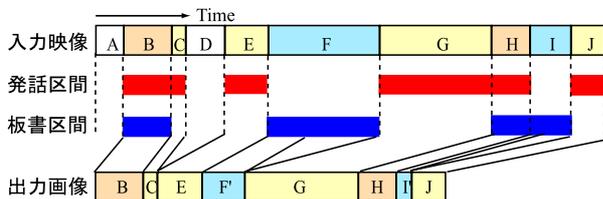


図 6: 講義短縮概要

図 7 に, 板書検出と発話検出に基づく講義短縮ビデオ自動生成例を示す.

4.3 評価実験

HDV カメラを用いて撮影した講義映像 3 本 (各 80 分) に対して, 本手法を適用し板書区間の検出を行う. 撮影した映像には, 板書や, 板書文字を消すという動作が含まれている. 本実験の評価には, 再現率と適合率を用いて評価を行う. 再現率は, 検出漏れの少なさを表す指標であり, 適合率は, 誤検出の少なさを表す指標である.

板書区間の検出結果

表 1 に, 板書検出結果を示す. 本手法は, 平均再現率が 85.1%, 平均適合率が 95.7% で板書区間を検出することができた. 講師領域を正確にセグメンテーションすることで,

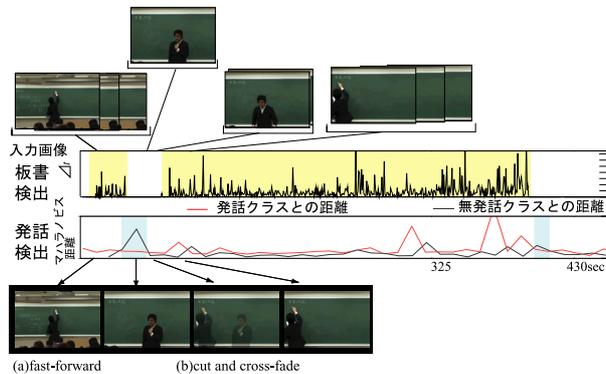


図 7: 講義短縮例

従来の板書検出手法 [3] より正確な板書区間検出が可能となる. 再現率が適合率に比べて値が低い理由として, 講師が文字を書く時に, 文字と講師が重なるために発生したオクルージョンが原因である. しかし, この原因による板書検出の遅延は一区間において 5 秒程度であり, 短縮する際に大きな問題とはならない.

表 1: 板書区間検出

	再現率 [%]		適合率 [%]	
	本手法	手法 [3]	本手法	手法 [3]
Movie1	84.8	53.9	96.7	84.6
Movie2	82.4	54.7	93.9	74.9
Movie3	88.1	57.0	96.4	87.5
Ave.	85.1	55.2	95.7	82.3

短縮講義ビデオ生成結果

図 8 は, 手動により生成した短縮講義ビデオの短縮率と, 本手法により生成した短縮ビデオの短縮率 (A) 及び従来の板書検出手法 [3] から生成した短縮講義ビデオの短縮率 (B) の類似度を表している. 図 8 より, 本手法により生成される映像の短縮率は, 従来手法に比べて手動での短縮率に近く, 手動での編集に近い短縮ビデオの自動生成が実現できた.

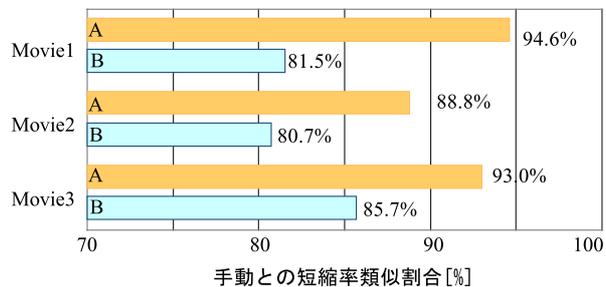


図 8: 手動との短縮率の類似割合

5. おわりに

本研究では, HD 映像からの仮想カメラワークによる短縮講義ビデオの自動生成法について提案した. 放送カメラマンの撮影技術に基づくトリミングにより, 臨場感あるカメラワークの実現が可能となる. さらに, 発話区間と板書区間の講義イベントの検出を行い, 人間が編集したビデオとほぼ同程度の短縮ビデオを自動生成することができた. 参考文献

- [1] 西口敏司ほか: “大学における実運用のための講義自動アーカイブシステムの開発”, 信学論 (D-II), Vol. J88, No.3, pp.530-540, 2005.
- [2] Y. Rui et al: “Videography for Telepresentations”, Proc. of the ACM Conference on Human Factors in Computing Systems, pp. 457 - 464, 2003.
- [3] 西口敏司ほか: “首振りカメラによる板書文字記録システム”, 信学総大, pp. 37-42, 1996.

研究業績

- [1] T.Yokoi and, H. Fujiyoshi: “Virtual Camerawork for Generating Lecture Video from High Resolution Images”, ICME2005, 2005.
- (他 研究会技術報告 2 件, 学会口頭発表 1 件)