高解像度映像からの自動講義ビデオ生成-仮想カメラワークの実現-

Automatic Lecture Video Generation from High Resolution Images

横井 隆雄† 藤吉 弘亘†
Takao Yokoi†, Hironobu Fujiyoshi†
†中部大学工学部情報工学科
†Department of Computer Science, Chubu University
E-mail: taka@vision.cs.chubu.ac.jp, hf@cs.chubu.ac.jp

Abstract

本稿では，高解像度映像から講師の動きに追従して領域を切り出すことで自動的に講義ビデオを生成する手法について提案する．提案手法は，HDV カメラ映像からフレーム間差分により検出した移動体領域座標に対して，バイラテラルフィルタを用いてカメラワークを決定するためのタイミングを検出する．検出したタイミング間において，対象領域のトリミング位置を変化させることで，仮想カメラのパンニングとズーミングを実現する．これにより，カメラマンが手動で撮影したかのような臨場感ある映像を生成することが可能となる．本手法により生成した映像を学生 20 人によるアンケート評価の結果，提案手法はカメラワークによる映像の見やすさに対して効果的であることを確認した．

1 はじめに

近年，教育機関において Web Based Training (WBT) のような e-learning の実施が増加している．特に，撮影した講義映像をインターネットを介して配信することで遠隔地等の教育格差の是正が期待でき，今後さらに需要が高まっていくことが予想される．しかし，高等教育機関等で多くの講義を専門のカメラマンが撮影し，編集を行うことはコスト面に問題がある．

このような問題に対して，講義を対象とした複数のカメラによる自動撮影の研究が行われている [1，2]．これらの手法は，複数のカメラ映像から発話等のイベントを抽出し，その情報を基にカメラワークや映像の切り替えを決定している．しかし，これらの手法で生成された編集映像は放送カメラマンのようなカメラワークではないため，臨場感溢れる映像とならない．

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法を提案する．本稿では，フレーム間差分を用いて移動体検出した結果に対して，バイラテラルフィルタを施すことで変動を抑制し，講師等の注目領域を追従する仮想カメラのパンニングとズームのタイミングの決定法について示し，放送カメラマンの撮影技術を考慮した仮想カメラワークについて述べる．

2 高解像度画像からの講義映像の生成

HDV(1080i) カメラを黒板全体が入り，かつ黒板の板書文字を読むことができるように講義室の後方部に設置し講義を撮影する．撮影した映像を配信する際，講義映像を受信する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) を表示することができない．

この問題を解決するために，図 1 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，講義ビデオを生成する．提案するシステムでは切り出すサイズを DVD フォーマットと同じ 720 × 480 画素とする．トリミングを行う際に，パンニングやズームにおいて放送カメラマンのカメラコントロールを再現することで臨場感あるビデオを生成することを目指す．このような臨場感ある講義映像を自動生成するには，パンニングやズーミングのタイミングの検出と仮想カメラワークの算出が必要である．

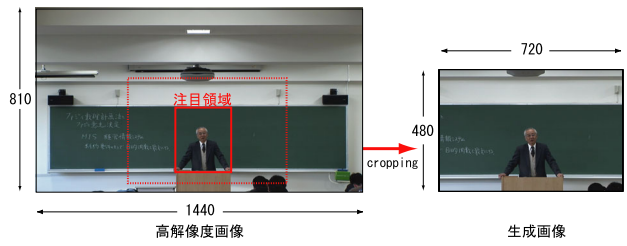

図 1 高解像画像からのトリミング画像生成

3　仮想カメラワークの生成

3.1　フレーム間差分による講師位置の検出

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 [6].

$$\Delta(n) = \max\{|\ I(n) - I(n-k)\ |\} \qquad 0 < k < 5 \quad (1)$$

$I(n)$ は現在の入力画像，$I(n-k)$ は k フレーム前の画像とする．急激な輝度値の変化がピクセル上に生じたとき，変化量 $\Delta(n)$ の値は大きくなる．ここでピクセルの状態を表す M は，輝度変化をしきい値処理することにより，移動体を判定する．

$$M(n) = \begin{cases} 1 : object & , \Delta(n) \geq Th \\ 0 : background & , \Delta(n) < Th \end{cases} \quad (2)$$

次に，フレーム間差分により移動体と判別したピクセル群をラベリングによりセグメンテーションを施し，各領域 $R_i(n)$ を求める．複数の領域が存在する場合，面積が最も大きい領域を講師領域 $R(n)$ と決定し，その重心座標 (x_c, y_c) を算出する．図 2 に，得られた講師領域 $R(n)$ の x 座標を示す．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場合がある．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ってしまう．

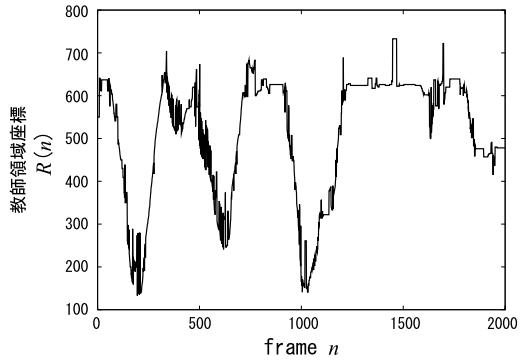

図 **2**　フレーム間差分による教師領域検出結果

3.2　バイラテラルフィルタによるタイミング検出

本手法では，上記に示した激しい変動に対してバイラテラルフィルタを用いて抑制を行う．バイラテラルフィルタは，注目画素からの距離による重みに加えて，注目画素との画素値の差に応じて，ガウス分布に従う重みを付けた平均化を行う [5]．ここでは，教師領域位置 $R(n)$ の変動を抑制するために，次式に示すように一次元に変形したバイラテラルフィルタを用いる．

$$\hat{R}(n) = \frac{\sum_{k=-w}^{w} W(n,k) \cdot R(n+k)}{\sum_{k=-w}^{w} W(n,k)} , \quad (3)$$

$$W(n,k) = \exp\{-\frac{k^2}{2\sigma_S^2}\} \cdot \exp\{-\frac{(R(n) - R(n+k))^2}{2\sigma_R^2}\}$$

図 3 は，フレーム間差分により得られた注目領域座標値に対してバイラテラルフィルタを 5 回および 10 回施した結果である．これより，エッジを保存しつつ細かな振動を抑制していることが分かる．

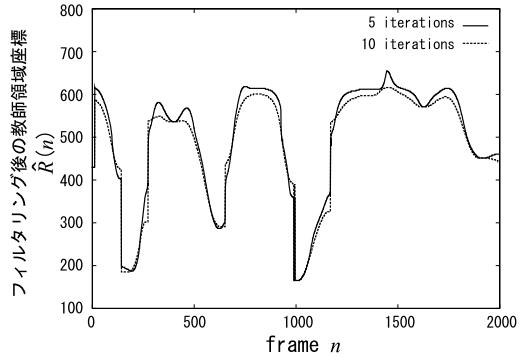

図 **3**　バイラテラルフィルタによる変動の抑制

バイラテラルフィルタにより抑制された教師領域座標 $\hat{R}$(n) に対して，零交差処理を施し特徴点を求める．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区間と判定する．図 4 に，検出したパンニング区間とズーム区間を示す．

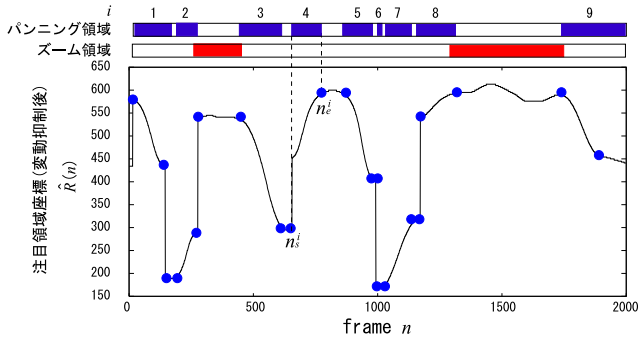

図 **4**　パンニングタイミング検出

3.3　仮想パンニングの実現

臨場感ある講義映像を生成するために，本システムでは，放送カメラマンの撮影技術を考慮した仮想カメラワークを実現する．文献 [3] では，放送カメラマンのカメラワークを解析した結果，物体を撮影する際のパンニングに以下の特徴があることを報告している．

- パンニング速度曲線非対称型で，減速時間が加速時間に比べて 6 割程度長い．
- 加速時の最大パンニング速度は，カメラの動きが最大速度に到達する直前に生じ，減速時には，最大速度から減速に入った直後に見られる．

上記に示した放送カメラマンの撮影技術を考慮した仮想カメラパンニングを実現するために，以下に示すア

ルゴリズムによりトリミング位置を決定する．

step1 パンニング区間の開始フレームを n_s と終了フレームを n_e としたとき，減速時間が加速時間に比べて 6 割程度長い特徴に従い，加速終了点 n_t を決定し，その座標値を $\hat{R}(n)$ を求める (図 5 参照)．

$$n_t = \frac{0.6n_s + 0.4n_e}{0.4 + 0.6} \quad (4)$$

$$\hat{R}(n_t) = \frac{0.6\hat{R}(n_s) + 0.4\hat{R}(n_e)}{0.4 + 0.6} \quad (5)$$

step2 パンニングの加速度 α は加速区間 ($n_s < n \leq n_t$) と減速区間 ($n_t < n \leq n_e$) に応じて以下の式により計算する．

$$\alpha = \begin{cases} \dfrac{2 \cdot (\hat{R}(n_t) - \hat{R}(n_s))}{(n_t - n_s)^2} & ,n_s < n \leq n_t \\ \dfrac{2 \cdot (\hat{R}(n_e) - \hat{R}(n_t))}{(n_e - n_t)^2} & ,n_t < n \leq n_e \end{cases} \quad (6)$$

step3 高解像度映像からの切り出し位置 $R'(n)$ を step2 で求めた加速度から計算し，仮想カメラパンニングを実現する．

$$R'(n) = \begin{cases} \frac{1}{2}\alpha \cdot (n - n_s)^2 + \hat{R}(n_s) & ,ns < n \leq n_t \\ \alpha \cdot (n_t - n_s)^2 + \frac{1}{2}\alpha \cdot (n_e - n_t)^2 + \hat{R}(n_t) & ,n_t < n \leq n_e \end{cases} \quad (7)$$

図 5 に，パンニングタイミングから求めた仮想カメラパンニングのための座標値と，その際のパンニング速度の変化を示す．

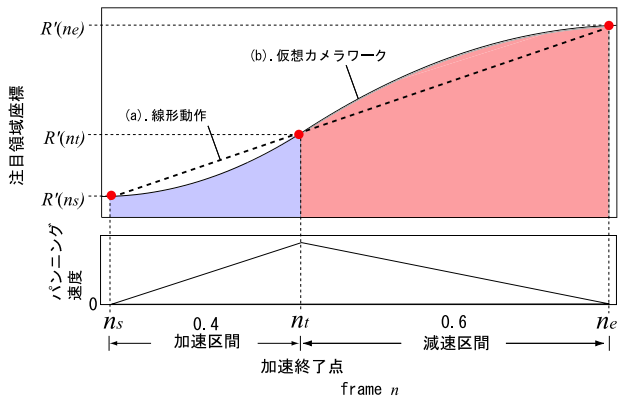

図 **5**　仮想カメラワーク **(**パンニング**).**

3.4　仮想ズーミングの実現

放送カメラマンのズーミングにおけるカメラワークの特徴として，文献 [4] では，ズーミングの加速時間・被写体のサイズ変化に以下の傾向があることを報告している．

- ズーミングの加速時間と減速時間の平均値は加速時間が 0.47~0.60 に対し，減速時間は 0.40~0.53 である．ズームアウトの場合，加速時間は 0.38~0.63 で減速時間は 0.37~0.61 の範囲である．
- 被写体サイズの変化率の最大値は，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ようにズーミングを行っている．

これらの特徴を考慮したズーミングを実現するために，フレームの拡大率を次式より決定する．ズームインもしくはズームアウトにかかる時間を t_{zoom} とし，ズーム最大拡大率を z_{max} とする．

$$Z(t) = \begin{cases} 1 + \sum_{i=-t_{zoom}}^{t} \frac{i^2}{M} e^{(\frac{-i^2}{N})} & ,-t_{zoom} < t \leq 0 (zoomin) \\ 1 + (z_{max} - \sum_{i=0}^{t} \frac{i^2}{M} e^{(\frac{-i^2}{N})}) & ,0 < t \leq t_{zoom} (zoomout) \end{cases} \quad (8)$$

M,N は，最大ズーム率・ズーミング時間により調整する．図 6 に提案する仮想カメラワークにより生成された講義ビデオの一部を示す．

4　実験結果

10 人の講師の講義を HDV カメラで撮影した映像から，以下に示す 4 手法により生成した講義ビデオ映像の比較評価を行う．

A. フレーム間差分

フレーム間差分によって得られた座標値を中心にトリミングした講義映像．

B. バイラテラルフィルタ

フレーム間差分の結果にバイラテラルフィルタを施して得られた注目領域座標値を中心にトリミングした講義映像 (図 3)．

C. 線形カメラワーク

3.2 で述べたパンニングタイミングに対して線形補間により得られた座標値を中心にトリミングした講義映像 (図 5(a))．

D. 仮想カメラワーク (提案手法)

3.3，3.4 で述べた放送カメラマンの撮影手法を基にして生成した仮想カメラワークにより得られた座標値を中心にトリミングした講義映像 (図 5(b))．

4.1　評価実験

上記の 4 手法により生成した各映像を被験者 (学生 20 人) に提示し，以下に示す各項目に対して評価をした．評価の尺度となる選択肢は (とても良い，良い，どちらでもない，悪い，とても悪い) とした．

Q1. 教師のいる部分が見やすかった?
Q2. 講師の動きがわかりやすかった?
Q3. 映っていないところをみたいと感じなかった?
Q4. カメラワークは自然だった?
Q5. 全体的に映像は見やすかった?

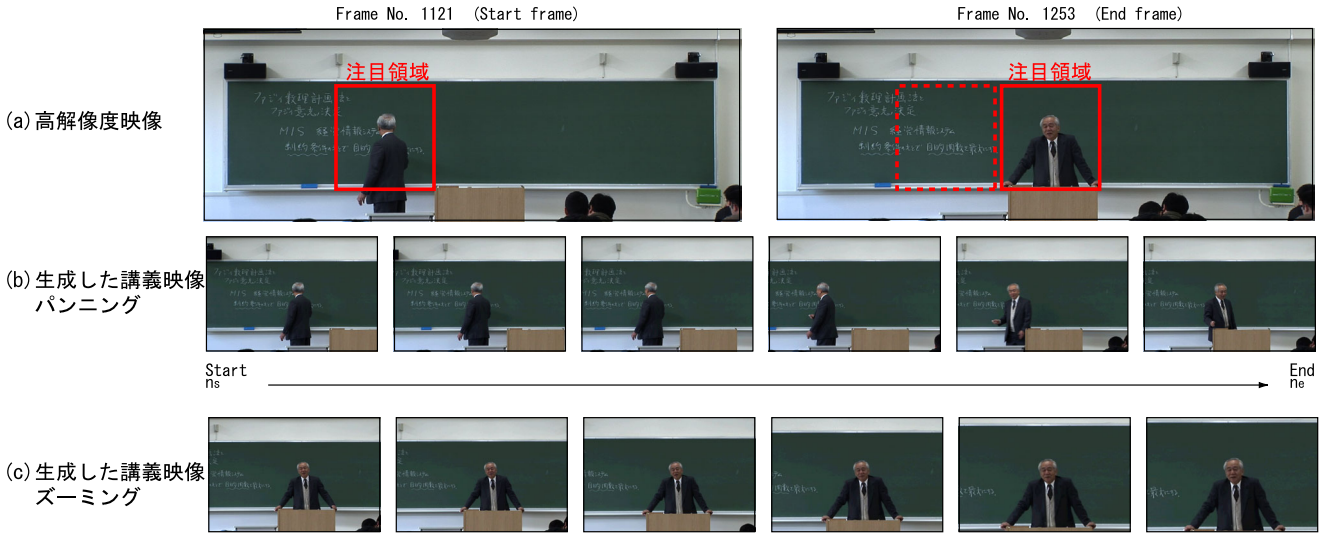

図 6　仮想カメラワークによる講義ビデオ生成

4.2　実験結果

評価尺度を-2 から 2 のスケールに変換し (-2=とても悪い，2=とても良い)，評価データ数 200 としたときの各手法の平均値を図 7 に示す．

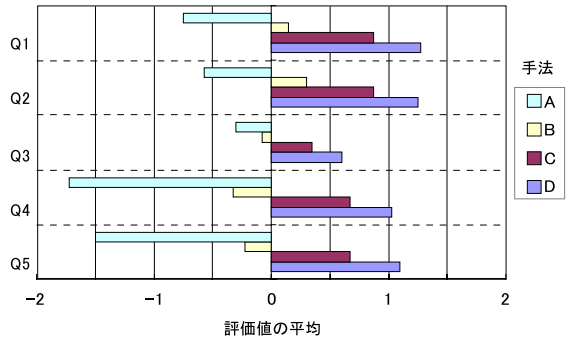

図 7　評価実験結果

図 7 に示すグラフより，最も評価が低かったのは，手法 A により生成された映像である．これは，フレーム間差分で得られた位置からトリミングを行うため，ゆれの激しい映像となるためである．一方，手法 C, D においては全ての質問項目に対してプラスの評価を得た．そこで，両者に対して有意水準 5% で t-検定を行った結果，質問 5 において優位差が得られた．従って，質問 5 において手法 C より手法 D の方が評価値が高いことから，放送カメラマンの撮影手法を考慮した仮想カメラワークは有効であるといえる．

5　まとめ

本稿では，高解像度映像からのトリミングによる講義ビデオ映像生成手法について提案した．本手法では，バイラテラルフィルタを用いてパンニングとズームのためのタイミングを検出し，放送カメラマンの撮影手法を考慮した仮想カメラワークを実現することで，より臨場感溢れる映像を生成することが可能となった．さらに，アンケートによる評価実験からその有効性を確認した．今後は，システムにユーザーインターフェースを加え，ユーザの入力に対してインタラクティブに映像を生成する手法ついて検討する予定である．

参考文献

[1] Y. Kameda, K. Ishizuka, and M. Minoh, “A Live Video Imaging Method for Capturing Presentation Information In Distance Learning,” Proc. of ICMCS’99, Vol. 2, pp. 897-902, 1999.

[2] 尾関基行，中村裕一，大田友一，“机上作業シーンの自動撮影のためのカメラワーク”, 信学論 D-II, Vol.J86，No.11，pp.1606-1617，Nov. 2003.

[3] 加藤大一郎，山田光穂，阿部一雄, “スタジオ番組における放送カメラマンのカメラワークと視線の動きの分析”, テレビジョン学会誌, Vol.49, No.8, pp.1023-1031,1995.

[4] 石川秋男, 加藤大一郎, 津田貴生, 福島宏, 下田茂, 阿部一雄, “放送カメラマンのズーミング計測法の検討と静止している被写体を撮影するときのズーミング解析”, 映像情報メディア学会誌, Vol.53, No.5, pp.749-757, 1999.

[5] C. Tomasi and R. Manduchi, “Bilateral Filtering for Gray and Color Images”, Proc. of ICCV, pp. 839-846, 1998.

[6] A. Lipton, H. Fujiyoshi, and R. Patil, “Moving target classification and tracking from real-time video”, Proc. of WACV, IEEE, pp. 8-14, 1998.