

キーフレーム間の物体追跡結果に基づく 仮想カメラワーク映像の生成法

篠木 雄大[†] 藤吉 弘亘[†]

[†] 中部大学工学部情報工学科

487-8501 愛知県春日井市松本町 1200

E-mail: †shinoki@vision.cs.chubu.ac.jp, ††hf@.cs.chubu.ac.jp

あらまし 近年では、個人で手軽に HD(High-Definition) 映像の撮影、映像のインターネット配信が可能となっている。しかし、撮影した HD 映像を直接配信することは、映像サイズやデータ容量の問題があり不可能である。本稿では、撮影した HD 映像中の 2 つのキーフレームに存在する対象物をユーザが指定し、2 つのキーフレーム間を双方向に追跡した結果を基に ROI(Region of Interest) のトリミングを行う事により、カメラワーク映像を自動生成する手法を提案する。ROI をトリミングする際には bi-directional tracking により得られた対象物の追跡結果をトリミング位置を基に、放送カメラマンの撮影特徴を利用したカメラワークモデルを適用することにより、滑らかで自然なカメラワーク映像の自動生成が可能である。カメラワークモデルを適用するカメラワークと適用しないカメラワークの比較を行い、主観評価を行う事により本手法の有効性を確認した。

キーワード キーフレーム、仮想カメラワーク、トリミング、双方向追跡

A Method for Generating Video with Virtual Camerawork by Object Tracking between Keyframes

Yudai SHINOKI[†] and Hironobu FUJIYOSHI[†]

[†] Department of Computer Science, Chubu University

1200 Matsumoto-cho, Kasugai, Aichi, 487-8501 Japan

E-mail: †shinoki@vision.cs.chubu.ac.jp, ††hf@.cs.chubu.ac.jp

Abstract Recently, personal video sharing services, such as Youtube, have become popular because it is easy to record videos in high-definition(HD) using a personal camcorder. However, it is impossible to broadcast an HD video via the Internet due to its large amounts of data. In this paper, we propose a method for generating a video with virtual camerawork based on object tracking technology. Once the user specifies positions of the region of interest(ROI) on the keyframes, the proposed method can be used to generate virtual camerawork between two keyframes in a row based on the results of bi-directional tracking. We evaluated our method with subjective experiments and confirmed its effectiveness.

Keyword keyframe, virtual camerawork, cropping, bi-directional tracking

1. ま え が き

近年、家庭用 HD ビデオカメラの低価格化により、個人で手軽に HD 映像の撮影が可能となっている。これに伴い、個人が撮影した映像をインターネットを介して配信するサービスの利用者が急増している。しかし、撮影した HD 映像を直接インターネットを介して配信することは、高解像度であるため映像サイズ・データ容量の問題があり非常に困難である。この問

題の一番容易な解決方法としてダウンサンプリングによる画像圧縮が多く利用されているが、画質を劣化させてしまうため、映像中の被写体を視認できない可能性がある。有効な解決法の 1 つとして、デジタルカメラワークによるカメラワーク映像の生成が挙げられる。デジタルカメラワークは、入力映像の注目領域 (ROI) をフレーム毎にトリミングし映像を生成する手法であるため、画質劣化をすることなく映像サイズ・データ容量を抑えることができる。従来のデジタルカメラワークを

用いた自動映像生成の研究では、手法 [1] [2] [3] [4] が挙げられ、[1] [2] では、パンニング・ズーム動作の速度曲線は十分な検討がされておらず、[3] では、HD 映像からのカメラワーク映像生成を行っているが、カメラが線形の動きをするため自然なカメラワーク映像とはならない。自然なカメラワーク映像を実現するため、[4] では、放送カメラマンの撮影特徴を基にした仮想カメラワークモデルを適用している。この手法により、自然なカメラワーク映像を生成することが可能であるが、限られた条件下での講義映像を入力とした自動生成法であり、一般映像には対応することが出来ない。しかし、一般映像の場合では様々な条件下で撮影された映像を対象とするため、完全な自動化は困難である。

そこで本稿では、ユーザが指定したキーフレーム内の ROI を基に対象物を追跡し、その追跡結果を基にカメラワーク映像生成を行う手法を提案する。提案手法では、ユーザの入力情報である ROI を基に bi-directional tracking [5] によるキーフレーム間の双方向物体追跡の結果よりカメラワークを設定するため、従来法では行うことのできなかつた各ユーザが望むカメラワーク映像を生成できる半自動システムとなる。また、カメラワーク映像を生成する際、放送カメラマンの撮影特徴を基にした仮想カメラワークモデルを適用させることにより、滑らかな自然なカメラワーク映像を実現する事が可能となる。

2. 提案手法

デジタルカメラワークを用いたカメラワーク映像の自動生成には、対象物の追跡手法が不可欠であるが、従来の手法では、時系列上に順方向に対象物の追跡を行った結果のみでカメラワークを施すため、1つの入力映像に対して1つのカメラワーク映像しか生成することができず、ユーザー一人一人が望んだ映像を生成することができない。しかし、本手法ではユーザの入力した複数のキーフレームにおける ROI の位置座標を基に、以下の3つの手法を使い分けることにより、仮想カメラワークを施した映像の自動生成を行うことが可能である。

- (1) キーフレームと ROI のみでカメラワークを設定
- (2) キーフレームの ROI 内の対象物を一定時間追跡
- (3) キーフレーム間の ROI 内の対象物を双方向追跡

(1) の手法では対象物の追跡は行わず、指定されたキーフレームの ROI 同士をつなぎ合わせ生成する。これにより、ユーザが追跡が必要ないと思う場合のカメラワークを生成することができる。(2) の手法では1つのキーフレーム内に存在する対象物を時系列上で順方向に追跡を行う。キーフレーム同士で対象物が異なる場合この手法を用いてカメラワークを生成する。(3) の手法では2つのキーフレーム内に存在する対象物の情報を基に、キーフレーム間を bi-directional tracking [5] により被写体である対象物を双方向から追跡する。双方向の物体追跡を行うことにより、重なりや類似物体への誤認識などへの対応が可能となり、長時間にわたるカメラワーク映像の自動生成に利用できる。この手法ではキーフレーム内の対象物が同じである場合に追跡が成功となり、その追跡結果に基づきカメラワークを生成する。

(1)(2)(3) 何れかの手法により得られた結果に対し、パラレルフィルタによる変動抑制、零交差法による特徴点抽出、カメラワークモデルを適用することにより、自然なカメラワーク映像を生成することが可能である(図1)。本稿では、(3)の手法及び有効性について述べる。

3. Bi-directional tracking による対象物の追跡

物体追跡法は既に多くの手法が提案されている。本手法では、対象物の追跡を行う際に発生する突発的な動き、不安定な動き、重なりで頑健に対応できる必要があるため、bi-directional tracking [5] を用いる。bi-directional tracking は映像中から2つのキーフレームを指定し、それぞれのキーフレーム内に存在する同一の対象物領域を指定する。その後、2つのキーフレーム間の対象物をそれぞれのキーフレームから双方向に追跡を行う。双方向に追跡を行うため、対象物が隠れた場合においてもそれぞれのキーフレームからの追跡結果からその隠れた部分の軌跡を推定することが可能である。

3.1 キーフレームと対象物の指定

Bi-directional tracking では、対象物は矩形領域 $R = \{p, s, \hat{w}, \hat{h}\}$ で表現される。ここで、 p は四角形の中心、 s はスケール、 w, h はそれぞれ矩形領域の幅と高さを示している。実際に対象物を表現する際には、矩形領域の中心とスケールを用いて $x = \{p, s\}$ と示し、状態空間へと投影する。各対象物の状態 x の表現にはカラーヒストグラムを用い、 $\mathbf{h} = \{h_1, \dots, h_H\}$ で表現され、 H は $8 \times 8 \times 8$ ピンの 512 次元 RGB カラー空間を示している(図2)。入力映像のシーケンス $Y = \{y_1, \dots, y_T\}$ 中の指定する2つのキーフレームそれぞれにおいて、対象物の状態 $\mathbf{h}(x_0), \mathbf{h}(x_T)$ を作成する。

3.2 フレーム内の対象物の位置の推定

時系列での位置を一度の計算で求めるには計算時間が膨大となるため、はじめに、各フレーム内の2次元座標中で対象物の位置の推定を行う。推定にはカラーヒストグラムの Bhattacharyya 距離によって得られる重みの極値を求めることにより行う。

3.2.1 Bhattacharyya 距離による重みの算出

フレーム内の対象物の位置の推定を行うには、フレーム内のある矩形領域 $\mathbf{h}(x')$ でラスタ走査し、対象物の状態 $\mathbf{h}(x_0)$ との比較を式(1)を用いて行うことにより、フレーム y_i 内での重み $\psi(y_i|x_i, x_1, x_T)$ を求める。

$$\psi(y_i|x_i, x_1, x_T) \sim \exp(-\min\{B^2[\mathbf{h}(x_i), \mathbf{h}(x_1)], B^2[\mathbf{h}(x_i), \mathbf{h}(x_T)]\}/2\sigma_h^2\}) \quad (1)$$

ここで、 $B^2[\mathbf{h}(x_i), \mathbf{h}(x_1)]$ は、矩形領域内のヒストグラム同士

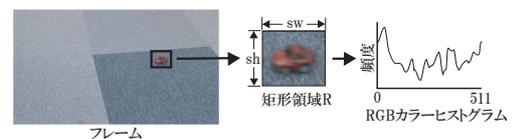


図2 対象物の状態

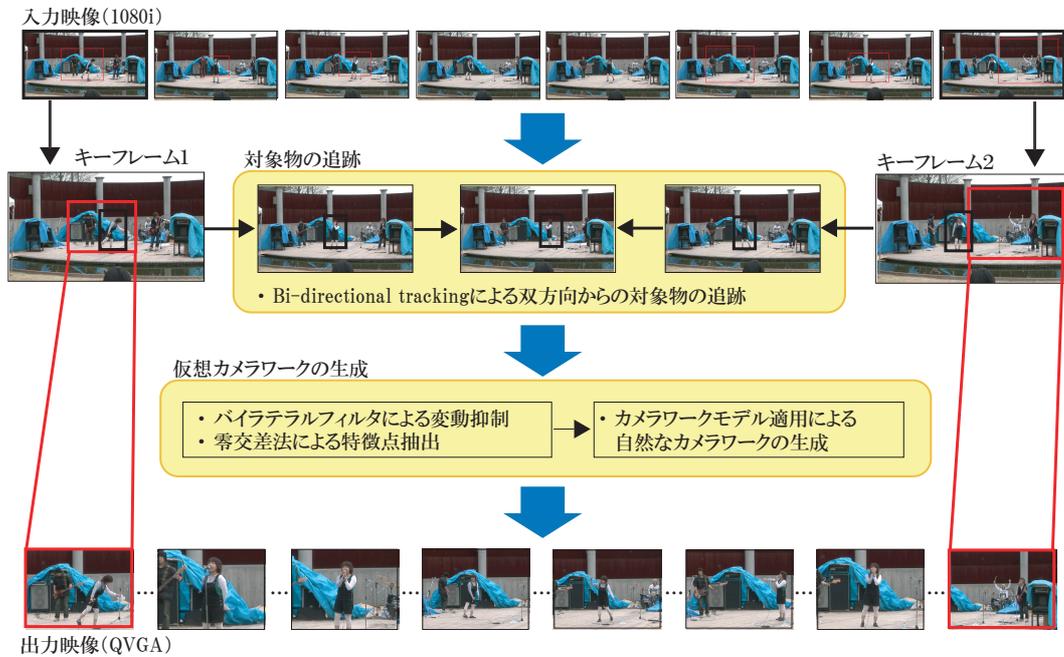


図1 キーフレームを基にしたカメラワーク映像生成

の Bhattacharrya 距離を表しており、次式により求められる。また、 σ_h^2 は分散パラメータである。

$$B^2[\mathbf{h}(x_0), \mathbf{h}(x_i)] = 1 - \sum_{j=1}^B \sqrt{h_j(x_0)h_j(x_i)} \quad (2)$$

求められた Bhattacharrya 距離による重みを可視化したものを図 3(b) に示す。図 3(b) では人の領域を対象物として指定している。この結果より、対象物である人の領域の重みの値が大きいことが確認できる。

3.2.2 Mean-shift 探索による極値の算出

求められた重み $\psi(y'_i|x_i, x_1, x_T)$ は、値が大きいほど対象物との類似度が高く、対象物がある位置に存在する可能性が高い。そのため、フレーム内の重み $\psi(y'_i|x_i, x_1, x_T)$ の極大値を計算することにより、対象物の存在する領域を推定する。極大値を探索する計算には、mean-shift [6] を用いて行う。mean-shift を用いることにより、効率的に極大値を求めることが可能である。通常 mean-shift で探索を行う場合、探索する領域は注目点を中心とした一定領域しか行わないが、ここでは、物体の重なりなどを考慮する必要があるため、フレーム画像全体を探索し極大値を求める。mean-shift 探索における初期点は格子状に配置をする。求められた極大値 $\psi(y'_i|x_i, x_1, x_T)$ の中で、ある閾値以上（ここでは 0.5）のものを対象物が存在する領域の重み $\psi(y_i|x_i, x_1, x_T)$ とする。mean-shift 探索による結果を図 3(c) に示す。

3.3 時系列での対象物の推定

すべてのフレームに対して重み $\psi(y_i|x_i, x_1, x_T)$ を計算した結果から、その座標を時系列を含む 3 次元空間へと投影する。その際の座標は $\mathbf{m}_n = [p_n, t_n]$ と表現し、 p_n はフレーム内での座標、 t_n は時系列での座標（フレーム番号）を示している。投影した座標をクラスタリングし、得られたクラスタ同士の繋がりから時系列での対象物の推定を行う。

3.3.1 Spectral clustering による分割

投影しただけの対象物の推定領域では、フレーム内に対象物の領域と類似した色情報を持つものが存在する場合、また、重なりにより対象物が隠れてしまう場合、時系列の対象物の軌跡を求める際に複数の物体が存在することになるため、正しく求めることができない。そこで、推定領域を時系列情報を含めたクラスタリングにより、対象物とそれ以外とを分割する必要がある。クラスタリングには spectral clustering [7] を用いる。spectral clustering は、クラスタリング対象となるある情報をそのまま利用してクラスタリングを行うのではなく、その情報内にある共通の成分を取り出し比較することにより行う。そのため、通常の k-means クラスタリングのような線形的なクラスタリングには不可能な、形状を考慮したクラスタリングが可能である。以下にその手順を示す。

Step1. クラスタリング対象となる座標 \mathbf{m}_i から相関行列 A を次式により求める。

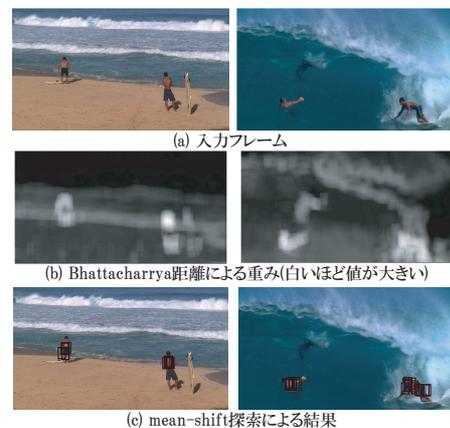


図3 対象物の状態

$$A_{ij} = \exp(-\|p_i - p_j\|^2/2\sigma_p^2 - \|t_i - t_j\|^2/2\sigma_t^2) \quad (3)$$

ここで、 σ_p と σ_t はスケールのパラメータである。次に、次式により相関行列 A にアピアランスの情報を含める。

$$A'_{ij} = \alpha A_{ij} + (1 - \alpha) \exp(-B^2[h(m_i), h(m_j)]/2\sigma_h^2) \quad (4)$$

ここで、 $B^2[h(m_i), h(m_j)]$ は、クラスタリング対象となる各座標 m_i の色情報を用いた Bhattacharyya 距離による重みである。また、 α はそれぞれの比率を表す係数である。

Step2. 求められた相関行列 A を用いて、相関行列 $L = D^{-1/2}AD^{-1/2}$ を求める。ここで、 D は $D_{ii} = \sum_{j=1}^N A_{ij}$ で求められる対角行列である。 $D^{-1/2}$ を両側からかけることにより、 A の各要素に対して、その要素がクラスタリング対象となる座標 m_i 同士の類似性全体の中での割合を求めることができる。

Step3. 次に、 L の固有値問題を解き、その固有ベクトルから行列 $E = [e_1, \dots, e_K]$ を求める。各要素の e_k は L の最大固有ベクトルを K 個使用する。この K はクラスタ数として指定する値である。ここで求められる E は各座標 m_i の類似性から得られる主成分を表しており、成分の大きなものを K 個使用することにより、クラスタリングに必要な成分のみを利用することができる。そして、k-means クラスタリングで E をその要素の行ごとにクラスタリングを行い、その結果を行番号を基に元の空間に反映させることにより、対象となる座標 m_i が k クラスタに属するかを判定する。

クラスタリングされた結果を図 4 に示す。通常の k-means クラスタリングでは不可能な形状を考慮したクラスタリングを行えていることが確認できる。また、k-means クラスタリングでは設定する初期シードの位置により、クラスタリング結果が変わるとい問題があるが、spectral clustering では主成分でのクラスタリングを行うため、初期シードに依存せず、全て同じ結果が求められる。

3.3.2 Tree-growing algorithm によるクラスタの結合

クラスタリングされた各座標 m_i のクラスタから対象物が含まれるものを選択し結合する。[5] では、クラスタの分析を行った結果から以下に示すことを定義している。

- (A) クラスタにはキーフレーム内の対象物を含む
- (B) 類似した対象物を含むクラスタは除外する
- (C) 時間軸上で 2 つのクラスタ間では重なりは存在しない
- (D) 重なりが存在する間は一定の速度と時間制限がある

これらを定義した上で、[5] では tree-growing algorithm を利用し、重なりを推定およびクラスタの結合を行う。まず、 T_A, T_B のツリーを用意し、2 つのキーフレームの対象物の座標 m_i を active ノードとしてそれぞれに入れる。また、クラスタを次の

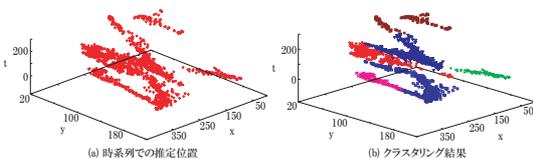


図 4 spectral clustering による結果

ノードの候補とする。次に (B) に従って対象となるクラスタを除外する。その後 (C) と (D) に従って T_A のノード候補が T_B の子ノードを 1 つつくわえる。これは Q-best hypothesis で行われる。Q-best hypothesis は $(L_o + \gamma S_o)$ の最小となる値を取ることで求められる。ここで L_o は二つの部分軌跡の間の長さ (pixel) であり、 S_o は最大速度 (pixel/frame) である。以上の操作を T_B に関しても同様に行う。もし active ノードが見当たらない場合は処理を終了し、まだ存在する場合は (B) に従ってクラスタを除外する部分へ戻る。tree-growing algorithm によるクラスタの結合を行った結果を図 5(a) に示す。この結果より、形状を考慮したクラスタリングにより対象物の軌跡である部分と、それ以外の部分が分割されていることが確認できる。

4. 追跡結果に基づくカメラワーク生成

前章でのクラスタの結合により 1 つに求められたクラスタを対象物の軌跡とし、カメラワーク生成に用いる。カメラワーク生成には得られたクラスタに対して、フィルタリング処理を行い、カメラワークに適した軌跡へと補正を行う。

4.1 対象物の軌跡の生成

クラスタに存在する座標 m_i が同一フレーム内に複数ある場合は、その座標間で重心を求めフレーム内での対象物の座標を 1 つにまとめる。また、フレーム内に対象物の座標が存在しない場合は、前後の座標 m_{i-t}, m_{i+t} を基にした線形補間により座標を生成する。これにより生成される軌跡はこのままの状態では激しく変動している。このため、本手法ではバイラテラルフィルタを用いて急激な変化を残しつつ、微細な変動を抑制し軌跡の補正を行う。バイラテラルフィルタによる変動抑制の結果を図 5 に示す。

4.2 カメラワークモデルの適用

バイラテラルフィルタを適用した座標に対して、零交差法により特徴点を求め、カメラワークタイミングを決定する。本手法では、[4] で用いられている仮想カメラワークモデルを適用することにより、カメラワークを生成する。この仮想カメラワークモデルは [8] に報告されている、放送カメラマンのカメラワークを解析した結果から得られるパンニング特徴を基にしており、以下に示す特徴がある。

- パンニング速度曲線は非対称型で、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加速度は、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。

このモデルを利用することにより、実際のカメラに近いカメラ

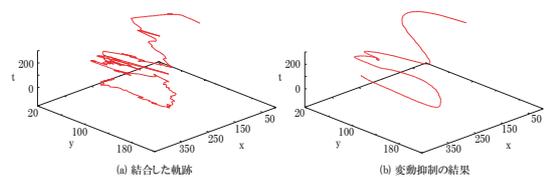


図 5 バイラテラルフィルタによる変動抑制

ワーク映像を生成することが可能である．以下にその手順を示す．

Step1. パニング区間の開始位置を t_f と終了位置を t_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，加速終了位置 t_c を決定し，その座標値 \hat{m}_c を求める．

$$t_c = 0.6t_f + 0.4t_e \quad (5)$$

$$\hat{m}_c = 0.6\hat{m}_f + 0.4\hat{m}_e \quad (6)$$

Step2. パニングの加速度 α は加速区間 ($t_f < t \leq t_c$) と減速区間 ($t_c < t \leq t_e$) に応じて以下の式により計算する．

$$\alpha = \begin{cases} \frac{2 \cdot (\hat{m}_c - \hat{m}_f)}{(t_c - t_f)^2} & , t_f < t \leq t_c \\ \frac{2 \cdot (\hat{m}_e - \hat{m}_c)}{(t_e - t_c)^2} & , t_c < t \leq t_e \end{cases} \quad (7)$$

Step3. 高解像度映像からのトリミング位置 m'_i を，step2 で求めた加速度から計算し，仮想パニングを実現する．

$$m'_i = \begin{cases} \frac{1}{2}\alpha \cdot (t - t_f)^2 + \hat{m}_f & , n_s < n \leq n_t \\ \alpha \cdot (t_c - t_f)^2 - \frac{1}{2}\alpha \cdot (t_e - t)^2 + \hat{m}_c & , t_c < t \leq t_e \end{cases} \quad (8)$$

図 6 に仮想パニングモデル，図 7 に仮想カメラワーク適用後の軌跡を示す．

5. 実 験

本手法の有効性を確認するために，生成したカメラワーク映像の評価を行う．

5.1 実験環境

図 8 に示す入力映像に対してキーフレームと ROI を指定した後，追跡結果をそのまま用いた補正なしのカメラワーク，追跡結果に Zhang 等による手法の線形カメラワーク [9]，追跡結果に仮想カメラワークモデルを利用したカメラワークを用いて映像を生成する．また，追跡結果を用いず，2 つの ROI を繋ぎ合わせたのみのカメラワーク映像，約 50 フレーム毎に ROI を指定し，手動で繋ぎ合わせたカメラワーク映像も生成し，それらの比較評価を行う．追跡結果に仮想カメラワークを利用したカメラワーク映像の例を図 8 に示す．

5.2 評 価

5.1 節の実験環境に基づき生成したカメラワーク映像 (約 10 ~ 20 秒) を，順不同で被験者 (学生 20 人) に提示した．各

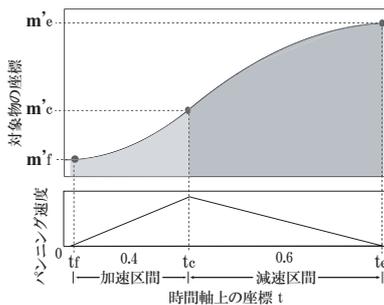


図 6 仮想パニングモデル

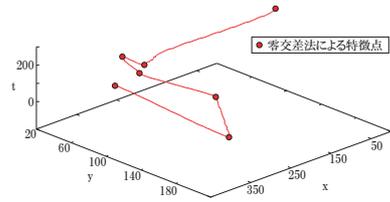


図 7 仮想カメラワーク適用後の軌跡

手法の映像の視聴後に，以下に示す項目に対するアンケート評価を行った．被験者は各映像生成手法についての事前の予備知識はない．

Q1. カメラワークは自然だった?

Q2. 全体的に映像は見やすかった?

評価の尺度の選択肢とスケールは，(とても良い:2，良い:1，どちらでもない:0，悪い:-1，とても悪い:-2) の 5 段階で評価を行った．各手法の平均値を図 9 に示す．2 つの ROI の繋ぎ合わせのみのカメラワークでは，質問 1，2 共にマイナスの評価となっている．これは対象物がカメラワーク映像から外れてしまうためと考えられる．これに対し，追跡結果に Zhang 等の線形カメラワークと提案手法である仮想カメラワークを利用した手法では，質問 1，2 共にプラスの評価を得られ，対象物の追跡結果からカメラワークを生成することにより，見やすい映像を生成できたことが確認できた．また，図 9(a) に示すように，移動カメラでは追跡結果とカメラワークモデルを用いた両手法の間に大きな差は見られなかった．これは，カメラワークを施す際の座標の与え方が理由と考えられる．本手法では特徴点間のカメラワークの座標を求める際に，x 座標，y 座標別々に行っているため，カメラワークモデルの適用を行うことにより加速度の面では実際のカメラワークに近づくが，特徴点間の動きは直線となっている．この時，入力となる映像に施されている元々のカメラワークが影響し，カメラワークが施されている映像からのトリミングを行う際に，同じような動きとなってしまうと考えられる．一方，図 9(b) に示すように，固定カメラでの入力映像ではこの影響が少ないため，提案手法である仮想カメラワークを用いた手法が，両質問共に高い値を得た．追跡結果にカメラワークモデルを当てはめるカメラワーク映像は，手動のカメラワーク映像と同等の結果を得た．これにより，追跡結果を用いることにてカメラワークを施すフレームの ROI をユーザが全て指定すること無く，適した仮想カメラワーク映像を生成することを確認した．また，追跡結果のみのカメラワークでは，質問 1，2 共にマイナスの評価となり，追跡結果をカメラワークへ直接利用することは難しいという事を

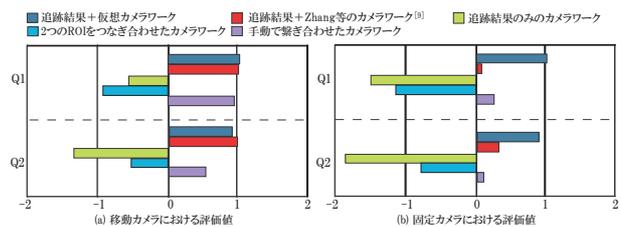


図 9 評価実験結果

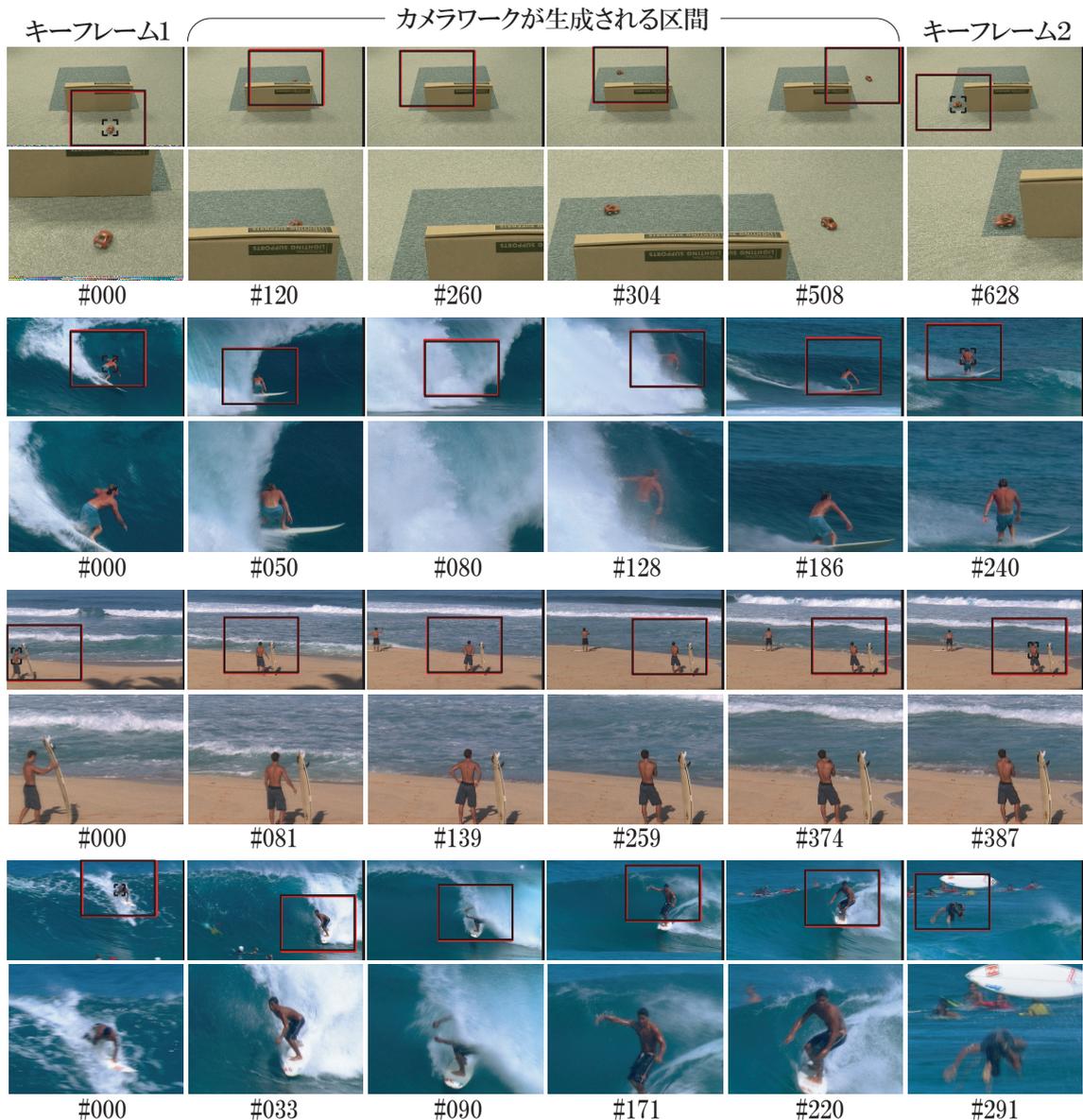


図 8 追跡結果に仮想カメラワークを利用した映像 (上：入力映像，下：出力映像)

確認した。しかし、実際に生成したカメラワーク映像は手ぶれのような動きをとり、評価の中には、実際に撮影している感じが出ているという意見も得られた。

6. む す び

本稿では、ユーザが指定するキーフレーム内の対象物を基に、双方向の対象物追跡を行った結果よりカメラワーク映像を自動的に生成する手法を提案した。その際、仮想カメラワークモデルを適用することにより、線形なカメラワークにはない自然なカメラワークを実現し、評価実験よりその有効性を確認した。今後は、ズームを含めた仮想カメラワーク映像の生成を検討する予定である。

文 献

- [1] 大西 正輝, 泉 正夫, 福永 邦雄, “デジタルカメラワークを用いた自動映像生成”, 画像の認識・理解シンポジウム (MIRU'2000) 講演論文集 I, pp. 331-336(2000).
- [2] 大西 正輝, 村上 昌史, 福永 邦雄, “状況理解と映像評価に

基づく講義の知的自動撮影”, 信学論 (D-II), Vol.J85, No.4, pp.594-603(Nov.2002).

- [3] 熊野 雅仁, 有木 康雄, 塚田清志, “ボールと選手に着目したデジタルカメラワークの実現法-デジタルシューティングによるサッカー解説映像生成システムに向けて-”, 映像情報メディア学会誌, Vol.59, No.2, pp.271-278(2005).
- [4] 篠木 雄大, 藤吉弘巨, “高解像度映像からの視聴者の注目点を考慮した講義映像の自動生成”, 映像情報メディア学会誌, Vol.62, No.2, pp.240-246(2008).
- [5] Jian Sun, Weiwei Zhang, Xiaou Tang, Heung-Yeung Shun, “Bi-directional Tracking using Trajectory Segment Analysis”, IEEE ICCV 2005, Vol.1, pp.717-724(Oct.2005)
- [6] D.Comanicu, V.Ramesh, P.Meer, “Real-time tracking of nonrigid objects using mean shift”, CVPR 2000.
- [7] Andrew Y.Ng, Michael I.Jordan, Yair Weiss, “On Spectral Clustering: Analysis and an algorithm”, NIPS 2004.
- [8] 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”, テレビ誌, Vol.49, No.8, pp.1023-1031(1995).
- [9] C. Zhang, Y. Rui, L. He and M. Wallick, “Hybrid Speaker Tracking in an Automated Lecture Room”, Proc. of IEEE ICME 2005, July, 2005.