追跡結果に基づく仮想カメラワーク映像の自動生成に関する研究

画像認識講座 篠木 雄大 指導教授 藤吉弘亘

1.はじめに

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 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.

2. 双方向の対象物追跡

今回は双方向追跡に利用する手法として bi-directional tracking[2] を用いた、双方向の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る、2.1.キーフレームと対象物の指定

Bi-directional tracking[2] では対象物は,矩形領域の中心 p とスケール s を用いて $x=\{p,s\}$ と表す.各対象物の状態 x はカラーヒストグラムを用いた $\mathbf{h}=\{h_1,\cdots,h_H\}$ で表現する.H は 512 次元 RGB 空間 (各 8 ビン) である.入力映像 $Y=\{y_1,\cdots,y_T\}$ 中の指定する 2 つのキーフレームにおいて,対象物の状態 $\mathbf{h}(x_0)$, $\mathbf{h}(x_T)$ を作成する. 2.2.フレーム内の対象物の位置の推定

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ため , フレーム内をある矩形領域 $\mathbf{h}(x')$ 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 , フレーム y_i 内の重み $\psi(y_i|x_i,x_1,x_T)$ を求める .

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

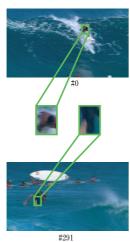
ここで, $B^2[\mathbf{h}(x_i),\mathbf{h}(x_1)]$ は,矩形領域内のカラーヒストグラム同士の Bhattacharyya 距離であり,次式により求められる.また, σ_b^2 は分散パラメータである.

$$B^{2}[\mathbf{h}(x_{0}), \mathbf{h}(x_{i})] = 1 - \sum_{j=1}^{B} \sqrt{h_{j}(x_{0})h_{j}(x_{i})}$$
 (2)

求められたフレーム内の重み $\psi(y_i'|x_i,x_1,x_T)$ の極大値を mean-shift 探索により求め,対象物の存在する領域を推定 する.極大値 $\psi(y_i'|x_i,x_1,x_T)$ の中で,ある閾値以上(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 1 に示す.

2.3. 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,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 は時系列での座標を示している。投影した座標



(a) キーフレームと指定する対象物



(b) 入力フレーム #59



(c) Bhattacharyya距離による重み (輝度値が高いほど大きい)



(d) mean-shift探索による結果

を spectral clustering[3] により,3次元空間でのクラスタリングを行う.以下にその手順を示す.

Step1. 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 .

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

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

ここで, $B^2[\mathbf{h}(\mathbf{m}_i),\mathbf{h}(\mathbf{m}_j)]$ は,各座標 \mathbf{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.

Step3. 次に,L の固有値問題を解き,その固有ベクトルから $E=[e_1,\cdots,e_K]$ を求める.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,結果を行番号を基に元の空間に反映させ,対象となる座標 \mathbf{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各座標 \mathbf{m}_i のクラスタから,tree-growing algorithm により重なりの推定およびクラスタの結合を行う.

3. 追跡結果に基づくカメラワーク生成

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. 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 \mathbf{m}_{i-t} , \mathbf{m}_{i+t} を基にした線形補間により座標を生成する. 軌跡の変動を抑えるために, バイラテラルフィルタを適用する(図 2).

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し、文献 [1] 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(図3).

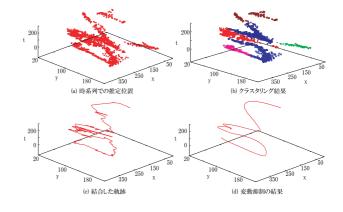


図 2: 軌跡の生成

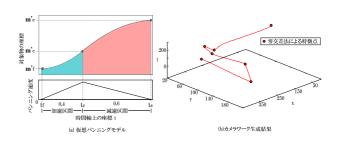


図3:カメラワーク適用結果

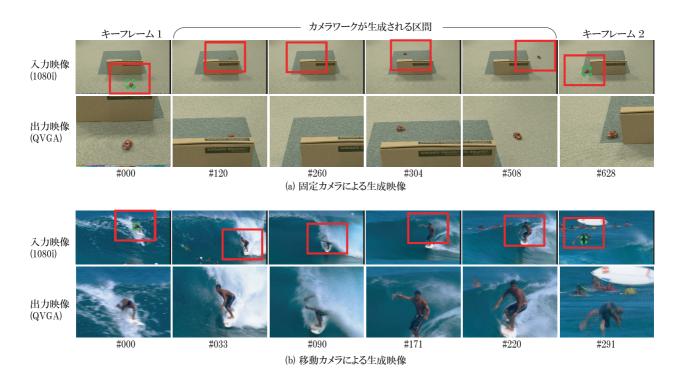


図 4: 仮想カメラワーク映像生成結果

4. 実験

入力映像に対してキーフレームと ROI を指定した後,以下のカメラワークで映像を生成した。

- A 双方向追跡 + 仮想カメラワーク (提案手法)
- B 双方向追跡のみ
- C 手動のカメラワーク

これらの映像に対して評価を行う.提案手法によるカメラワーク映像の例を図4に示す.

5.評価

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.各手法の映像の視聴後に,以下に示す項目に対しての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 ~ 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 5 に示す .

提案手法である A のカメラワークでは,両質問でプラスの評価を得られた.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.一方,B の追跡結果のみのカメラワークでは,両質問でマイナスの評価であることから,A のカメラワークで適用している仮想カメラワークの有効性を確認できた.

また,提案手法である A のカメラワークは,手動のカメラワークである C と同等の結果を得た.この結果より,追跡結果を用いることで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,自然なカメラワーク映像が生成できることを確認した.

6.おわりに

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,自然なカメラワークを実現し,アンケートによる評価実験より,提案手法の有効性を確認した.今後は,特徴点追跡を基にした手法の検討と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.

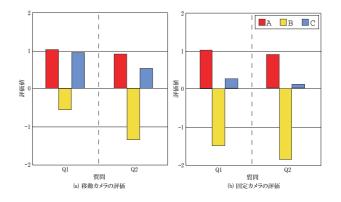


図 5: 評価実験結果

参考文献

- [1] 篠木 雄大,藤吉弘亘. "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.
- [2] Jian Sun, Weiwei Zhang, Xiaoou Tang, Heung-Yeung Shun. "Bi-directional Tracking using Trajectory Segment Analysis", IEEE ICCV 2005, Vol. 1, pp. 717-724, Oct.2005.
- [3] Andrew Y.Ng, Michael I.Jordan, Yair Weiss. "On Spectral Clustering: Analysis and an algorithm", NIPS 2004.
- [4] C. Zhang, Y. Rui, L. He and M. Wallick. "Hybrid Speaker Tracking in an Automated Lecture Room", Proc. of IEEE ICME 2005, July.2005.

研究業績

- [1] 篠木雄大,藤吉弘亘. "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",第4回デジタルコンテンツシンポジウムDCS'08,10-3.
- [2] 篠木雄大, 藤吉弘亘. "キーフレーム間の双方向物体追跡結果に基づく仮想カメラワーク映像の生成法", 映像情報メディア学会誌 Vol. 63, No. 2, pp. 209-215, 2009.

(他 学会誌論文1件,学会口頭発表1件)