

H-034

# 移動カメラ動画像からの手振り動作認識法の多人数動作への拡張法の検討

Study of Recognizing Hand Gesture from the Video Sequence Acquired by a Moving Camera  
- Extension to Multiple People's Gestures from a Single Person's Gestures -

羅 丹†  
Luo Dan

大谷 淳†  
Jun Ohya

## 1. はじめに

人間と移動ロボットの共存環境を実現するためには、人間がロボットに適宜リアルタイムに指示を与えられることが必要である。手振り動作は人間にとって、直感的で使いやすいインターフェースと考えられる[4]。従来、筆者らは、一人の手振り動作を対象に、移動カメラの位置変動を抑制した人物追従型座標システム(HFLC: Human Following Local Coordinate)による手振り動作軌跡モデルの抽出法と抽出したモデルに適用した Condensation アルゴリズムを提案した。ただ、不特定多数の人物手振り動作に対する有効性の検証は行っていない。本論文では、多人数動作データを対象にし、認識手法の有効性を検討する。

## 2. 研究手法

本研究では、「領域分割」、「手振り軌跡の抽出」、「手振りの認識」という流れで処理を行う。

「領域分割」という段階では、サンプルデータから色データベース(図1に示す)を作成した色情報を用いて、入力動画像のフレームごとに対して人間の肌、体と背景の各領域分割を行う。「手振り軌跡の抽出」の段階では、人物像を動画像中で追跡するとともに、人物像の重心点を原点とする座標系「人物追従型局所座標系 HFLC (Human-Following Local Coordinate)システム」を構築し、これを利用した安定高速かつ簡易な軌跡の抽出法を提案する。

### 2.1. 色情報による領域分割

異なる人種や多様な照明条件下で収集した肌色・非肌色(背景と服)サンプルに対して、RGB表色系で表す色情報をHSV表色系へ変換する。ガウス混合モデル(GMM)により肌色、背景色、服の色をモデリングする。モデリングでできた色相モデルに基づいて領域統合することにより顔、手、ボディの領域はメディアンフィルタでノイズ除去後抽出される[1]。図1にて1番の原画像に対し、2番の肌領域と3番のボディ領域の結果を示す。

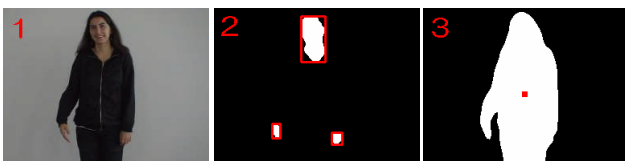


図4 肌領域とボディ領域の抽出結果

### 2.2. 手振り軌跡の抽出

手振り軌跡の抽出法は、人物追従型局所座標系(HFLC)の構築により、手振りの軌跡の抽出を提案する[2]。領域抽出で得られた顔領域の中心座標(F)とボディ領域の中心座標(G)の分布をローパスフィルタで平滑化したあと、Fを原点としてFGを軸とする人物追従局所座標(HFLC)を構築する。図2にて、座標のイメージ図を示す。図3の左図では「秋」というジェスチャ(320×240 pixel, 150フレーム[30fps])の顔とボディ領域の中心座

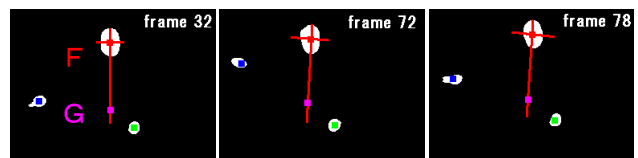


図3 適応的局所座標

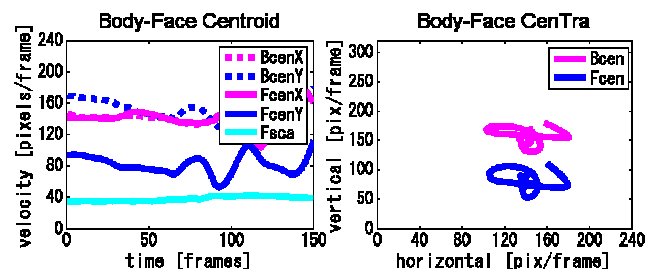


図3 中心座標の安定性と手振り移動軌跡

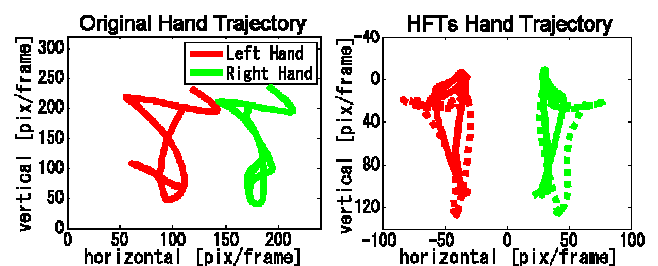


図3 オリジナル手振り軌跡とHFTs手振り軌跡

標の安定性を示す。図3の右図では中心座標の軌跡からカメラと人物の相対移動軌跡が分かる。

オリジナル手領域の移動軌跡HT<sub>o</sub>をHFLCに対して座標変換を行い、局所相対手振り移動軌跡HT<sub>f</sub>は算出される。さらに、カメラと人物の相対距離を一定にするために顔領域の長さF<sub>sca</sub>をスカラー量としてスケール変換し、スカラー局所相対手振り移動軌跡HFT<sub>s</sub>

( $HFT_s = HT_f * \{F'_{sca} / F_{sca}\}$ ) を計算する。最後に、 $HFT_s$  分布をローパスフィルタにより平滑化処理を行う。図3では  $F_{sca}$  の分布となる。図4にて、図3のジェスチャの手振り運動軌跡を表す。左図では  $HT_0$  を示す。右図では虚線が静止カメラで取った単純手振り軌跡  $HT$  となり、実線が  $HFT_s$  の結果を示す。動的カメラ画像から検出した軌跡  $HFT_s$  の結果を静止画像のときの軌跡  $HF$  と比べると、とても近い軌跡が検出されたことがわかる。

2.3. 手振りの認識

Condensation Algorithm を手振り認識に応用するために、Black と Jepson[3]の方法を拡張した手法で行う[2]。

まず、観測対象の状態を  $s$ 、観測結果を  $z$  とした場合の、時刻  $t-1$  における位置推定対象の確率分布  $P(s_{t-1} | z_{t-1})$  と、 $t-1$  から  $t$  までの状態推移確率  $P(s_t | s_{t-1})$  により、 $t$  における事前確率  $P(s_t | z_{t-1})$  を求める。次に、この事前確率に基づき、実際の観測結果から尤度  $P(z_t | s_t)$  を算出し、 $t$  における確率分布  $P(s_t | z_t)$  を求める。 $t+1$  においても、同様に、 $t$  における確率密度に従い、計算を繰り返す。この状態推移確率を求めるモデルをシステムモデル、尤度を求めるモデルを観測モデルと呼ぶ。従来追跡の分野で利用されたが、本研究では、認識アルゴリズムとして利用する。

3. 実験評価

本研究では、通常照明条件の下で、3人の被験者を対象として、日本とアメリカの手話に対応する35種類の手

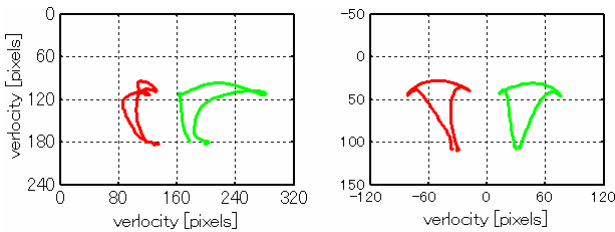


図6 手振り軌跡の抽出結果例：「大きい」

振り動作を利用した。被験者に色分けがない服の着用と単純な背景を設定する。

手振り軌跡モデルの抽出実験結果を図5に示す。

左図はカメラの動きを含まれる手振りの軌跡を表す。右図はカメラの動きを抑制した軌跡である。

Condensation Algorithm を用いて認識を行う際に、手振り軌跡モデルのデータベースを片手データベースと両手データベースを作成する。片手と両手の手振り軌跡モデルに対する認識結果例を図6に示す。

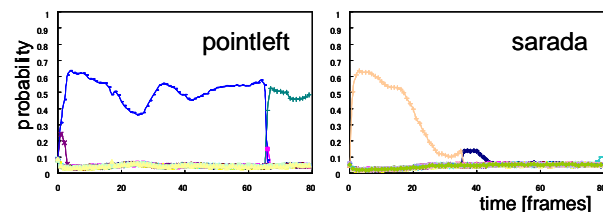


図7 片手と両手の認識の結果例

それぞれの被験者はトレーニングデータベースに登録するため、ひとつの手振り動作を5回行う。図7から片手の手振り軌跡モデルを用いた認識では、100%の認識率が得られた。両手の場合は90.9%となる。手振りの動作の

Input Database	(1) asa	(2) beautiful	(3) douita	(4) hat	(5) hello	(6) natu	(7) pointleft
(1) asa	8.33	0	0	0	0	0	0
(2) beautiful	0	13.66	0.4894	0	0	1.46	0
(3) douita	0	0	15.26	0	0	0	0
(4) hat	0	0	0	1.72	0	0	0
(5) hello	0	0	0	0	10.24	0	0
(6) natu	0	0	0	0	0	42.29	0
(7) pointleft	0	0	2.76	0	0.32	5.73	14.53

図5 認識の結果例 (片手)

動画画像中で、ひとつの手振り動作の始まりと終わりは、手の初期位置から手振りを発生し、また手の初期位置まで戻るまでの時間的区間と判断する。実験結果からわかるように、手ぶり動作の始まりと終わりに対応する軌跡は含まれるので、データベースの中に、始まりの時間的区間と終わりの時間的区間では、ほかの手振りモデルの尤度が高く評価してしまうケースもある。データを分析するところ、入力モデルと違うモデルが時間的区間における軌跡の分布が、急に高い確率が発生してしまう誤動作も発生する。この場合に対応するために、認識区間の自動的な判別が今後の課題である。

4. まとめ

本研究で、提案した人物追従型局所座標系 HFLC による抽出法を用いて、従来のアーム型のモデル抽出より、安定でかつ自由度が大きいジェスチャに対応できる。従来、静止カメラ画像からの片手の手振りしか適用できなかった Condensation アルゴリズムを動的カメラ画像中の片手と両手の手振りジェスチャの認識に適用し、良好な認識率を得た。また、一人の被験者で実験を行った[2]に対し、多人数の検証を行った。

今後の課題として、手振り動作の軌跡は手動で区間を区切ることを行ったことに対して、今後自動的に区切ることを目指す。また、手振り動作の自動手振り動作を分類し、手振りデータベースをより充実を増やし、本研究で提案した手法の問題点と有効性を検証することを考える。

参考文献

[1] M.-H. Yang and N. Ahuja, "Gaussian Mixture Model for Human Skin Color and Its Applications in Image and Video Databases", SPIE Storage and Retrieval for Image and Video Databases, vol. 3656, pp. 45-466, Jan. 1999.

[2] D. Luo, Jun Ohya, "Hand-gesture extraction and recognition from the video sequence acquired by a dynamic camera using condensation algorithm", San Jose, California, USA, 2009, IS&T/SPIE Electronic Imaging, (2009.1).

[3] Black M., Jepson A.: Recognition Temporal Trajectories using the Condensation Algorithm, Int'l Conf. on Automatic Face and Gesture Recognition, Japan, pp.16-21 (1998).

[4] Pavlovic, V.I.; Sharma, R.; Huang, T.S., "Visual interpretation of hand gestures for human-computer interaction: a review," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol.19, no.7, pp.677-695, Jul 1997.