2008－CVIM－162 (3)
2008／3／10

移動カメラ動画像からのCondensation Algorithmを用いた手振り認識の検討

羅 丹† 大谷 淳†
†早稲田大学 国際情報通信研究科

概 要 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，移動ロボットの動作に関する指示を与えられることが重要である。このようなヒューマンロボットインターフェースには，手振りが有効と考えられる。従来の手振り認識の研究は，静止カメラ画像を用いる場合がほとんどであり，動的カメラ画像の例は少ない。本研究では，動的カメラ画像中で人物追従型局所座標を構築することで，カメラの位置変動を抑制した安定かつ単純な手振り軌跡抽出法を提案する。これにより，Condensation アルゴリズムを，手振り動作認識手法として適用可能となる。日本とアメリカ手話に含まれる語彙に対応する35種類の動作を対象として，実験的な検討を行う。

Study of Hand Gesture Recognition from the Video Sequence Acquired by a Dynamic Camera Using Condensation Algorithm

Ra Tan† Jun Ohya†
†Graduate School of Global Information and Telecommunication Studies, Waseda University

Abstract The recognition of human gestures in image sequences is an important and challenging problem that enables human robot co-existence Environment as a human-computer interaction application. There already are many researches working on hand gesture recognition from image sequences acquired by still camera. But using Dynamic Camera is few. Our focus here will be on recognition of gestures from video sequences acquired by a Dynamic camera. We created a very simple and stable extracting method of hand motion model using Human-Following Local Coordinate system (HFLC), and we using condensation algorithm to recognize the hand motion model. We demonstrated 35 kinds of Japanese and American sign languages actions of gesture recognition.

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，人間とロボットの衝突回避や，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。このような，人間とロボットのインターフェースのためには，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手振り動作を利用するのが有効である。

従来，手話などの手振りを認識しようとする研究は，静止カメラ画像を用いる場合がほとんどであり，移動するカメラにより獲得された動画像からの手振り認識に関する研究例は少ない。静止カメラ画像からの手振り認識手法には，代表的なのは Starner ら[1,2] によって提案された HMM(Hidden Markov Models)と Waibel ら[3]が開発した TDNN(Time-delay Neural Networks)を用いる手法がある。また，Black ら[4,5]により提案した Condensation Algorithm はトラキングによく使われるが，観測処理に強い点で片手の簡単な手振り動作の軌跡の認識にも応用された。本研究は位置情報

の観測に強いCondensation Algorithmに基づいて，動的カメラ画像から片手だけではなく両手の手振り軌跡の認識法を提案する。

動的カメラ画像からのジェスチャ研究は数少ないですが，NASA Johnson Space centerのEric[6]らとカーネギーメロン大学のWaldherrら[7]はロボットに指示を与えるようなヒューマンロボットインターフェースのために，ロボットのステレオカメラ画像から手振りを人物の肩モデル情報を利用した。これらの研究における問題点として，肩の情報の抽出が容易でないことがあげられる。また，手振り動作によって肩の特徴点が消失したりするので，不自然な動作と限られる手振りを要求する。本研究では，より自由度が大きい自然な手振り動作（片手と両手の双方場合）を対象にし，正確かつ安定な抽出法を提案する。さらに，静止カメラ画像にしか応用できなかった認識アルゴリズムを動的カメラ画像から抽出した手振り（片手も両手もの場合）の軌跡モデルの認識に適用する。日本とアメリカ手話に含まれる語彙に対応する35種類の手話動作を対象として，実験的な検討を行う。

2　研究手法

本研究では，「手の発見」，「手振り動きの抽出」，「手振りの認識」という流れで処理を行う。

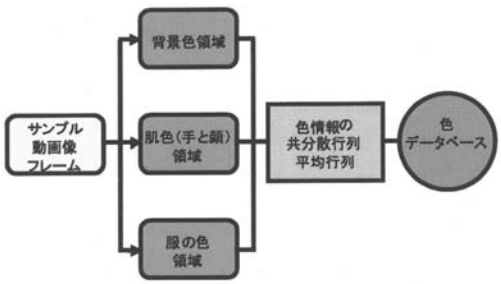

図1　領域分割のための色データベース

「手の発見」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領域分割を行うことで，胴体領域・顔領域・手領域を取得する。「手振り動きの抽出」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。

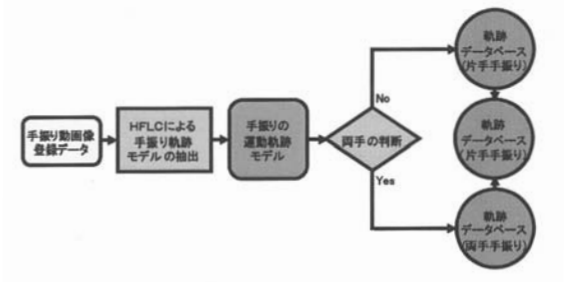

図2　手振り軌跡の学習データベース

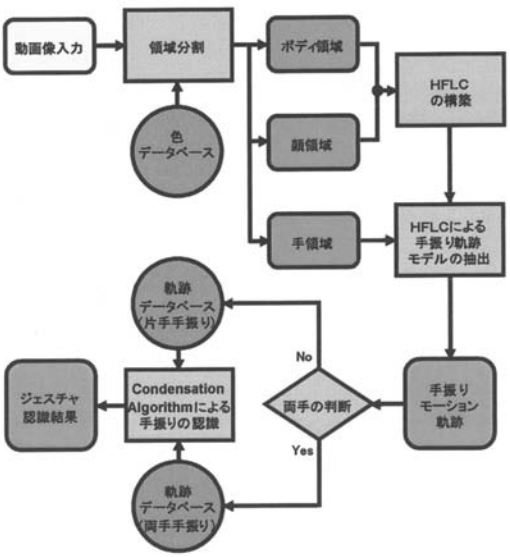

図3　アルゴリズムの流れ

最後，［手振りの認識］の処理では，HFLCにより抽出した手振り移動軌跡モデルを

Condensation の参照データベース（図２に示す）に登録し，新たな入力手振りの軌跡モデルを Condensation Algorithm を用いて認識する。

上記に本研究全体の処理の流れについて述べた。図３に示す。

３　手振りの抽出

本研究で提案した手振り抽出法は，色情報による領域分割，人物追従型局所座標系の構築，手振りの軌跡の抽出の順で行う。

3.1　色情報による領域分割

肌色を抽出しやすくなるために，原画像に RGB 空間から，肌色の彩度が高く表す HSV 空間への色相変換を行う。サンプル画像から，肌・背景・服のガウス分布モデルを求める。トレニングデータはこれらのモデルの集合となる。

本研究では，肌色と非肌色（服色と背景色）を同時に抽出する必要があるので，混合ガウス分布を利用した[8]。混合ガウス分布による領域分割を行った結果を図４に示す。原画像(a)に対して，顔領域・手領域(b)，胴体領域(c)が得られた。

(a)　原画像

(b)　肌領域の抽出結果

(c)　胴体領域の抽出結果

図４　領域分割の結果

3.2　人物追従型局所座標系

本研究では，胴体領域の中心座標，顔領域の中心座標を利用して，人物像に対して定められる局所的な座標系―「人物追従型局所座標系 HFLC(Human-Following Local Coordinate)システム」を提案した。

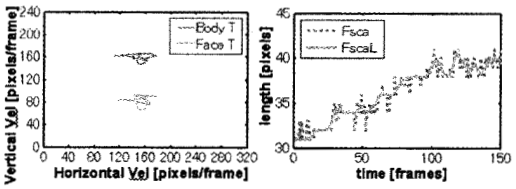

図５　胴体と顔スケールの安定性

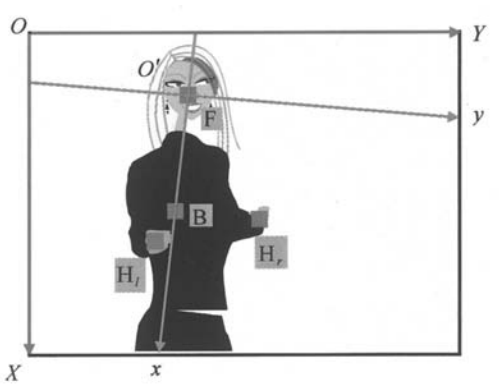

図６　HFLC システムの構築

ここで，画像座標系$O(\mathrm{X},\mathrm{Y})$と人物追従型局所座標系$O'(\mathrm{x},\mathrm{y})$を定義する。画像座標系$O(\mathrm{X},\mathrm{Y})$中の顔領域の重心 F を座標系$O'(\mathrm{x},\mathrm{y})$の原点とし，F と胴体領域の重心 B を通る直線を縦軸とする座標系$O'(\mathrm{x},\mathrm{y})$（図６に示す）を人物追従型局所座標系 HFLC とする

図５には，左図が胴体の重心と顔の重心軌跡となり，右図が顔の長さで表すスケールの分布である。そこからカメラの動きがわかることができる。顔スケール量を用いて，手振り軌跡の位置情報を正規化する。

3.3　手振りの軌跡の抽出

画像座標系$O(\mathrm{X},\mathrm{Y})$での両手の重心H_{LR}を

HFLC 座標系$O'(x,y)$への座標変換を行うことで，両手の中心座標H'_{LR}を式(1)で求める。

$$\begin{pmatrix} H'_{LR\text{-}x} \\ H'_{LR\text{-}y} \end{pmatrix} = \begin{pmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{pmatrix} \begin{pmatrix} H_{LR\text{-}x} \\ H_{LR\text{-}y} \end{pmatrix} \quad (1)$$

ここで，顔領域の長さF_lを用いて，両手の中心座標H'_{LR}スケール変換をし、ローパスフィルタによる（平滑閾値 0.13）平滑化処理を行う。

図7には，手振りの抽出の結果を示す。HFLC座標変換後の抽出結果は左上となり，顔スケールで相対距離を正規化した結果は右上となる。下図らはカメラの動きを抑制した手振りの軌跡の分布である。

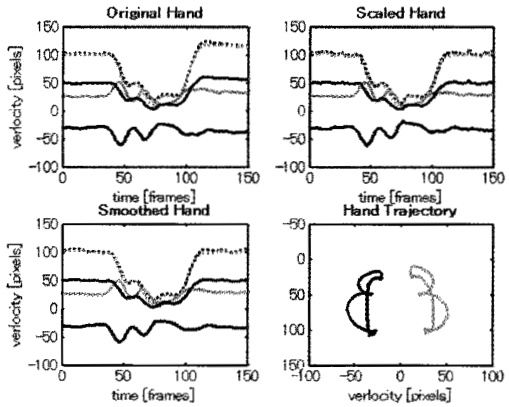

図7　手振り軌跡の抽出結果例

4　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[9]の方法を拡張した。

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$P(s_{n-1}|z_{n-1})$と，$t-1$からtまでの状態推移確率$P(s_n|s_{n-1})$により，tにおける事前確率$P(s_n|z_{n-1})$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$P(z_n|s_n)$を算出し，tにおける確率分布$P(s_n|z_n)$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。

- **初期化**

時刻tにおける状態をパラメーターベクトル$s_t=(\mu,\phi^j,\alpha^j,\rho^j)$と定義する。

μは認識対象の動作を表す。

ϕ^jは軌跡モデルの位相を表す。

α^jは手振り動作の振幅を表す。

ρ^jは時間においての伸縮さを表す次元のスケール要素である。

ここで，$i\in\{l,\gamma\}$，lは左手で，γは右手を表す。

- **観測状態の推測**

予測の段階では，ランダムに選択された状態sの各パラメータは次の時刻におけるサンプルの状態推定(式(2))に用いられる。

$$\begin{aligned} \mu_{t+1} &= \mu_t \\ \phi^j_{t+1} &= \phi^j_t + \phi^j_t + N(\sigma_\phi) \\ \alpha^j_{t+1} &= \alpha^j_t + N(\sigma_\alpha) \\ \rho^j_{t+1} &= \rho^j_t + N(\sigma_\rho) \end{aligned} \quad (2)$$

$N(\sigma_\star)$は標準偏差$\sigma_\star$による正規分布に従ってランダムに選択された数を示す。ここで，$\sigma_\phi=\sigma_\alpha=\sigma_\rho=0.1$と設定する。

- **アップデート**

尤度は観測データ$Z_t=(z_t,z_{t-1},\cdots)$の近似度を与える。本研究では，時間においてiの共分散には，観測データ$Z_{t,i}=(x_{t,i},x_{(t-1),i},\cdots)$を定義する。

$Z_{t,1}, Z_{t,2}, Z_{t,3}, Z_{t,4}$ はそれぞれが左手の水平位相の分布，左手の垂直位相の分布，右手の水平位相の分布，右手の垂直位相の分布となる。Black ら[22]の拡張で，尤度の計算は式(3)(4)のようになる。ここで，時間的な窓掛け ω（$\omega=10$）とする。

$$p(Z_t \mid s_t) = \prod_{i=1}^{4} p(Z_{t,i} \mid s_t) \qquad (3)$$

$$p(Z_{t,i} \mid s_t) = \frac{1}{\sqrt{2\pi}} \exp \frac{-\sum_{j=0}^{\omega-1} (x_{(t-j),i} - \alpha^* m^{(\mu)}_{(\phi^*-\rho^* j),i})^2}{2(\omega-1)} \qquad (4)$$

ここで、α^*，ϕ^*，ρ^* は領域の観測モデルの適切なパラメータ値を示す。$\alpha^{*(\mu)}_{(\phi^*-\rho^* j),}$ は時刻 $\phi^* - \rho^* j$ において内挿され，α^* 倍されたモデル μ の i 番目の倍数に与えられる値である。

Condensation の基本的な考えはもっとも近似しているモデルを認識結果とする。本研究では，0.3 以上の尤度を用いて，35 種類の動作の合計（時間の全域について）を比較し，最大値を与える動作を認識結果とする。

5　実験評価

本研究では、通常照明条件の下で、日本とアメリカの手話に対応する 35 種類の手振り動作を利用した。

手振り軌跡モデルの抽出実験結果は図８に示す。

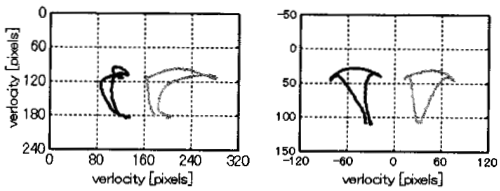

図８　手振り軌跡の抽出結果（例:「大きい」）

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９に示す。

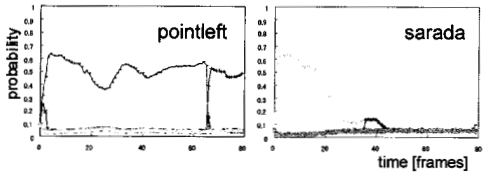

図９　片手と両手の認識の結果例

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１０　認識の結果例（片手一部）

図１０から片手の手振り軌跡モデルを用いた認識では，ほぼ，100%の認識率が得られた。両手の場合は 90.9%となる。手振りの動作の動画像中で，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。

6 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から撮った片手と両手の手振りジェスチャの認識に適用できた。今後の課題として、手振り動作を分類し、データベースをより充実に増やすと考える。

参考文献

[1] Starner, T. and Pentland. 1995. Real-Time American Sign Language Recognition from Video Using Hidden Markov Models, TR-375, MIT Media Lab.

[2] T.E. Starner and A. Pentland, "Visual Recognition of American Sign Language Using Hidden Markov Models", Proc. First Int'l Workshop Automatic Face and Gesture Recognition, pp. 189-194, 1995.

[3] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, and K. Lang, "Phoneme Recognition Using Time-Delay Neural Networks", IEEE Trans. Acoustics, Speech, and Signal Processing, vol. 37, no. 3, pp. 328-339, 1989.

[4] Michael J. Black and Allan D. Jepson, "A probabilistic framework for matching temporal trajectories: Condensation-based recognition of gestures and expressions." In Proceedings 5th European Conference Computer Vision, volume 1, pages 909-924, 1998.

[5] M. J. Black and A. D. Jepson, "Recognizing temporal trajectories using the condensation algorithm," in Int. Conf. Automat. Face and Gesture Recognition, Nara, Japan, Apr. 1998, pp. 16–21.

[6] Korten Kamp, David,Huber, Eric, Bonasso, R Peter,"Recognizing and interpreting gestures on a mobile robot ",The 1996 13th National Conference on Artificial.

[7] Stefan Waldherr, Roseli Romero, Sebastian Thrun, "A Gesture Based Interface for Human-Robot Interaction ",Springer Netherlands, Autonomous Robots,p151-173,2000.

[8] 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.

[9] 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).