

## 投票機構を用いた人物動作の認識手法

大藤 篤史<sup>†</sup> 渡辺 俊典<sup>†</sup> 古賀 久志<sup>†</sup>

<sup>†</sup> 電気通信大学大学院情報システム学研究科

E-mail: †{a-ofuji,watanabe,koga}@sd.is.uec.ac.jp

**あらまし** 人間の胴体・下半身の形状などをシルエット画像から自動抽出し、投票機構を用いてモデル動作データベースとのマッチングをとることによって、動画像から人物の基本動作を自動認識するシステムを構築した。投票機構は、動画の各フレームや身体の特徴量で個々にマッチングを行い最後に足し合わせるという手法で、時間や画像にノイズが存在していてもロバストな認識を行うことができる特長がある。本システムを用いることにより、人物の基本動作を認識できることを確認した。

**キーワード** 動作認識、投票機構、スター型スケルトン

## Human Motion Recognition With Voting

Atsushi OFUJI<sup>†</sup>, Toshinori WATANABE<sup>†</sup>, and Hisashi KOGA<sup>†</sup>

<sup>†</sup> Graduate School of Information Systems, University of Electro-Communications

E-mail: †{a-ofuji,watanabe,koga}@sd.is.uec.ac.jp

**Abstract** We built the system which recognizes basic motion of a human automatically from video by extracting the shape of human from silhouette and comparing it to human motion database with voting. The recognition method with voting is performed by summing up calculated votes obtained from the comparison of human body skeletons. It enables robust recognition against noisy condition. We confirmed that our system can recognize the basic behavior of human.

**Key words** Motion recognition, Voting, Star skeleton

### 1. はじめに

近年、計算機やネットワークの高性能化にともない、さまざまな動画データが大量に流通するようになってきた。今後、これらの膨大な情報の中から自分の見たいものをすばやく探し出す内容検索や、映像内容の自動要約・分類ができるシステムの構築が重要になってくると思われる。特に、人物動作を認識する技術は監視カメラやスポーツの実況、またフォームの解析など、広い範囲での応用の可能性を持っている。

人物の動作認識には、静止画の特徴を一枚ずつ見ていく方法と、動作の変化を見ていく方法に大きく分類することができる。前者には Yang らが手振り

認識に用いたニューラルネットなどがあり [?], 後者には Schlenzig らが "Hello" と "Good-bye" のジェスチャーを見分けるのに用いた隠れマルコフモデル (Hidden Markov Model:HMM) などがある [?]. しかし、多様な動作の認識には至っていないのが現状である。

Ben-Arie らは静止画の特徴を見る方法と動作の変化を見る方法を融合させ、投票機構によって認識を行うことにより 8 種類もの動作のロバストな認識に成功した [?]. だが、投票機構に入力する特徴量を抽出するのに動作の初期点を与える必要があるなど自動システムに応用させるには改良が必要である。

本研究では、完全に自動抽出が可能な特徴量を Ben-Arie らの投票機構に与えることにより、動画

像からの人物動作を自動的に認識するシステムの実現を目指す。

以下、第2節では Ben-Arie らによって提案された投票機構について述べる。第3節では特徴量自動抽出に利用した技術について述べ、第4節では実際にシステムに用いた特徴量について述べる。また第5節では特徴量を投票機構に適用する手法について述べる。第6節では、実験を通して提案手法の性質を検証し、第7節でまとめを行い、今後の課題を述べる。

## 2. Ben-Arie らの動作認識手法

ここでは、Ben-Arie らによって提案された投票機構による人物動作認識手法[?]を紹介する。

彼らの手法は、まず人体を図??のように9つの部分に分割し、これらの「現在の角度」および「変化量(角速度)」を抽出する。そして、あらかじめ登録されているモデル動作と未知入力動作とで、投票機構を用いて比較をするという流れで、人物の動作認識を行う。

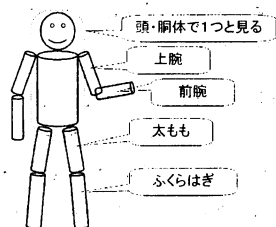


図1 分割された人物の身体モデル

### 2.1 投票機構

#### 2.1.1 モデル動作登録テーブル

図??のように、各身体部分について個別にテーブルを作成する。テーブルは、

- 左右の手・足について4次元テーブル4つ
- 胴体について2次元テーブル1つ

の計5つを用いる。例えば、手には「上腕と前腕」の「角度と角速度」の合計4つの要素が含まれるので、4次元テーブルを用いている。

各テーブルには、同じテーブルエントリに関係する全ての動作と時刻がセットで格納されている(図??)。

#### 2.1.2 投票による認識手順

認識手順を以下に示す。なお、式(??)~(??)に

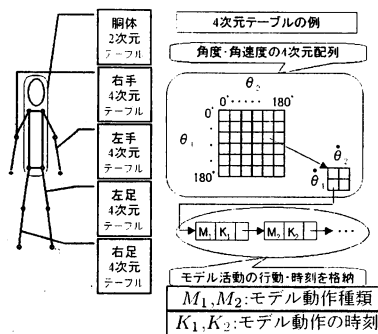


図2 左:人間の身体部分における5つのテーブル

右:4次元テーブルの例

※「 $M_1, K_1$ 」のように、各々のセルに該当するモデルと時刻はセットで格納される。

において、 $m$  はモデルナンバー、 $t$  はモデル動作のフレームナンバー、 $k$  はテスト動作のフレームナンバー、 $K$  はテスト動作のフレーム数、 $i$  はテーブルナンバー、 $N$  はテーブルの数を表している。

**Step1** 各々のテーブルにおいて投票値の計算。

投票関数は多次元正規分布モデル(式??)に従う。なお、 $d$  は次元数を表している。式(??)の場合は、2または4となる。また、値域は  $0 \leq V_{mk}^i(t) \leq 1$  である。

$$V_{mk}^i(t) = \exp \left[ \frac{-1}{2} \sum_{i=1}^d \left( \frac{a_i^m - a_i}{\sigma_{a_i}} \right)^2 \right] \quad (1)$$

ここで、 $a_i^m$  モデル動作の特徴量値、 $a_i$  などはテスト動作の特徴量値、 $\sigma_{a_i}$  は標準偏差を表す。

**Step2** Step1における結果を足し合わせ、1フレームにおける投票結果を求める。(式??)

$$V_{mk}(t) = \sum_{i=1}^N V_{mk}^i(t) \quad (2)$$

ここで、1フレームあたりの投票値の満点 ( $V_{mk}(t)$  の最大値) は  $N$  となる。

**Step3** Step2の結果をテスト動作のフレーム数だけ足し合わせ、最終投票結果とする。(式??)

$$V_m = \sum_{k=1}^K V_{mk}(L_k) \quad (3)$$

ここで、 $L_k (k = 1 \dots K)$  はテスト動作の  $k$  番目に相当するようなモデル動作のフレーム番号であり、 $V_m$  を最大とするように選択する。ただし、フレームの前後関係が逆転しないように  $L_i < L_j; i < j$  を満た

している必要がある。また、最終投票値の満点 ( $V_m$  の最大値) はテーブルの数  $N$  とテスト活動のフレーム数  $K$  の積になることができる。

最終投票結果で最多得票を獲得したモデル活動が「認識されたモデル」となる。

## 2.2 利点・欠点

Ben-Arie らの実験によると、本手法により 8 種類の基本動作を全て認識することが確認されている。多少の角度変化にも対応していて、高いロバスト性を持つ手法といえる。

しかし彼らは、人物画像から特徴量を抽出するのに EXM(EXpansion Matching) フィルタ [?] を用いているとしている。これは、人間の手によって人物の関節初期点を与える必要があり、完全な自動システムになっていないという欠点がある。

## 3. 特徴量の自動抽出

本研究では、前節で述べた投票機構に与える特徴量を自動抽出するために、スター型スケルトン [?] を利用する。

スター型スケルトンとは、人物のシルエット領域の重心および輪郭線上に自動的に抽出された特徴点で構成されているスケルトンである (図??)。

スター型スケルトンの抽出手順は、

- Step1 原画像から人物のシルエットを抽出 (図??(b))
- Step2 シルエットから輪郭を抽出 (図??(c))
- Step3 輪郭から重心点を求める
- Step4 輪郭点と重心点の距離を計算 (図??(d))
- Step5 距離が極大になる点が輪郭特徴点 (図??(d))
- Step6 重心点と輪郭特徴点を結ぶ (図??(e))

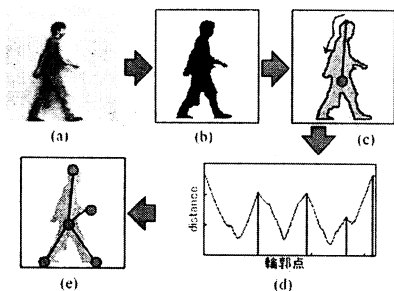


図 3 スター型スケルトンの取得手順:

- (a) 原画像、(b) シルエット画像、(c) 重心点と輪郭線、
- (d) 重心と輪郭の距離グラフ、(e) スター型スケルトン

このスター型スケルトンから人物の頭部および足部の特徴点を抽出することができる。(図??) すなわち、図??のように人物の大まかな形状を把握することが可能になる。

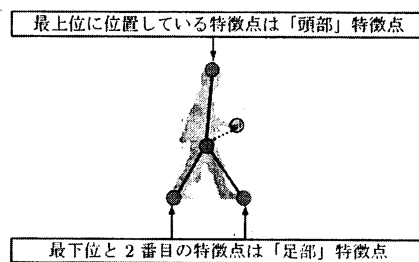


図 4 スター型スケルトンから得られた頭部と足部の特徴点

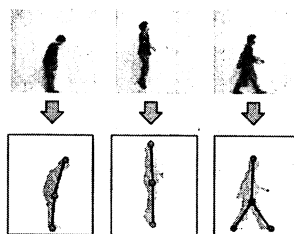


図 5 さまざまな身体形状から得たスター型スケルトン  
左:お辞儀、中:ジャンプ、右:歩き

## 4. 動作認識に用いる特徴量

ここでは、スター型スケルトンからどのような特徴量を得ることができるかについて、検討する。

### 4.1 上半身・下半身の傾き角度と角速度

図??のように、重心点と頭部特徴点から上半身(胴体)の傾き角度 ( $\phi$ )、重心点と足部特徴点から下半身(足)の傾き角度 ( $\theta_1$  と  $\theta_2$ ) を求めることができる。また、求めた角度で前フレームとの差分を求めれば角度変化の大きさを求めることができ、これを角速度として考えることができる。

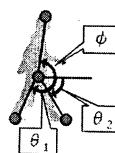


図 6 傾き角度と角速度

#### 4.2 外接長方形の縦横比

図??のように、頭・両足の3点を囲む外接長方形を考える。縦横比 ( $y/x$ ) を求めることで身体形状の広がりの変化を観察することができる。また、 $x, y$  の値ではなく比を用いることで、人物のサイズ変化に対応できるようにした。

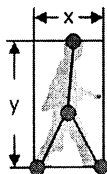


図7 外接長方形の縦横比

#### 4.3 上半身と下半身の高さ比

図??のように、頭と重心点 (A) および最下点と重心点 (B) の距離の比を考える。高さ比 ( $A/B$ ) を求めることで身体の中での重心位置の変化とすることができる。また、縦横比と同じく、 $A, B$  の値ではなく比を用いることで、人物のサイズ変化に対応できるようにした。

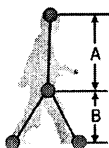


図8 上半身と下半身の高さ比

#### 5.2 特徴量をテーブルに登録

認識の際、テーブルに登録するため、各特徴量を離散化する必要がある。胴体・足の角度は1度ずつ180段階、角速度については4段階に分割を行った。また、外接長方形の縦横比および高さ比は10段階への分割とした。

各テーブルには、同じセルに關係する全てのモデル活動と時刻がセットで格納されている。(図??)

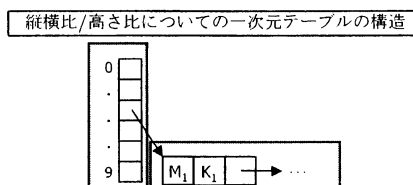
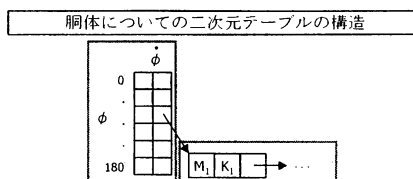
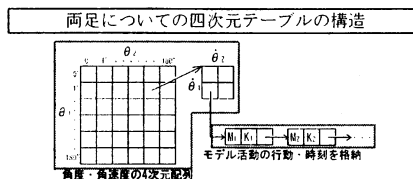


図9 動作認識に用いるテーブルの形状

### 5. 投票機構による動作認識

#### 5.1 テーブルの変更

第2節で紹介した投票機構に第4節で紹介した特徴量を与えるために、テーブルに変更を加えた。

図??のように、特徴量ごとについて別々にテーブルを作成する。テーブルは、

- 左右の足の角度・角速度で4次元テーブル
- 胴体の角度と角速度で2次元テーブル
- 外接長方形の縦横比で1次元テーブル
- 高さ比について1次元テーブル

計4つのテーブルを用いる。

また、認識の手順 (2.2節、式??~??) についての変更点はない。

### 6. 実験

提案手法の有効性を確認するために、実動画像を用いて認識実験を行った。

#### 6.1 モデル動作データベースの作成

モデルとなる動作は、4種類の基本動作 (歩き、お辞儀、物を拾う、ジャンプ) を設定した。これらに用いるサンプル画像の集合は、時間間隔を等間隔にした15枚の静止画像を用いた (図??)。この図では、15枚から8枚を抜粋している。各動作によってかかる時間が異なるため、それぞれの動作における時間間隔は異なっている。

これらの静止画像から人物部分のシルエットを抽

出し、スター型スケルトンを求めた。スター型スケルトンから人物の重心、頭部および足部の特徴点を求め、第4節で紹介した特徴量(胴体/両足の角度・角速度、外接長方形の縦横比、上下の高さ比)を計算した。実動画像(図??)から得たスター型スケルトンを図??に示す。

これらの計算した特徴量を用いて図??に表したテーブルに格納し、モデル動作データベースを作成した。

### 6.2 認識に用いる動作

認識に用いる動作は、モデルとなる動作とは別に撮影した基本動作(歩き、お辞儀)を設定した。特徴量を得る手順は4.1節と同じ方法であるが、画像の枚数は10枚とした。よって、投票値の満点は $4 \times 10 = 40$ となる。

### 6.3 実験結果

下表に、認識実験結果を示す。

モデル活動	歩き	拾う	お辞儀	ジャンプ
テスト:歩き1	35.19	23.09	26.50	26.26
テスト:歩き2	32.53	24.86	26.49	26.20
テスト:歩き3	32.09	24.54	27.74	27.58
テスト:お辞儀	27.30	19.49	33.71	31.23

こちらを見ると、全ての場合においてテストと同じモデルが投票最大値を示している。よって活動認識が正しく行えていることは言える。

しかし、テスト入力を「お辞儀」とした場合に認識結果が際どくなってしまっている。これは、図??の通り、テスト入力の「お辞儀」は会釈程度の浅いものであるのに対し、モデルの「お辞儀」は深いものであることが原因であると考えられる。浅いお辞儀は、「お辞儀」特有の「胴体角度」の変化が小さくなってしまい、認識に何らかの影響を与えたことが考えられる。

## 7. まとめと今後の課題

本稿では、Ben-Arieらによって提案された投票機構による動作認識手法の自動システム化を試みた。自動抽出可能なスター型スケルトンから得られたさまざまな特徴量を、投票機構によってあらかじめ教示したデータとのマッチングを取ることにより、人物の動作認識を行うシステムを構築し、基本動作の認識ができることを確認した。しかし、認識結果が

際どい場合も存在している。

今後の課題としては、特徴量や投票値の足し合わせ方法の見直しをして、より安定に活動認識が行えるシステムの実現を目指す。

## 文 献

- [1] H.Fujiyoshi and A.J.Lipton, "Real-Time Human motion Analysis by Image Skeletonization," Proc. Workshop Application of Computer Vision, Oct.1998.
- [2] J. Ben-Arie, Z. Wang, P. Pandit, and S. Rajaram, "Human Activity Recognition Using Multidimensional Indexing," IEEE Trans. Pattern Analysis and Machine Intelligence, vol.24, No.8, pp.1091-1104, 2002.
- [3] J. Schlenzig, E. Hunter, and R. Jain, "Vision Based Hand Gesture Interpretation Using Recursive Estimation," Proc. 28th Asilomar Conf. Signals, Systems, and Computers, 1994.
- [4] M.-H. Yang, and N. Ahuja, "Recognizing Hand Gesture Using Motion Trajectories," Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp.466-472, June 1999.
- [5] Z. Wang, and J. Ben-Arie, "Optimal Ramp Edge Detection Using Expansion Matching," IEEE Trans. Pattern Analysis and Machine Intelligence, vol.18, No.11, pp.1092-1098, 1996.

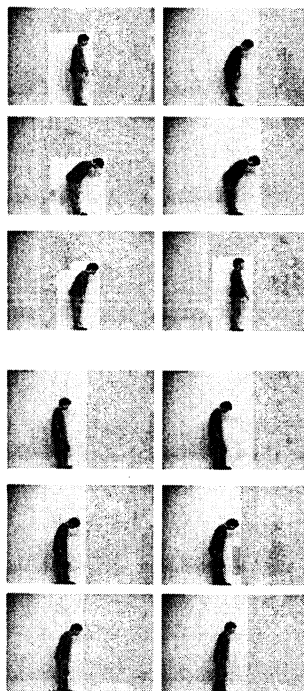


図10 上:モデルに用いた「お辞儀」  
下:テストに用いた「お辞儀」

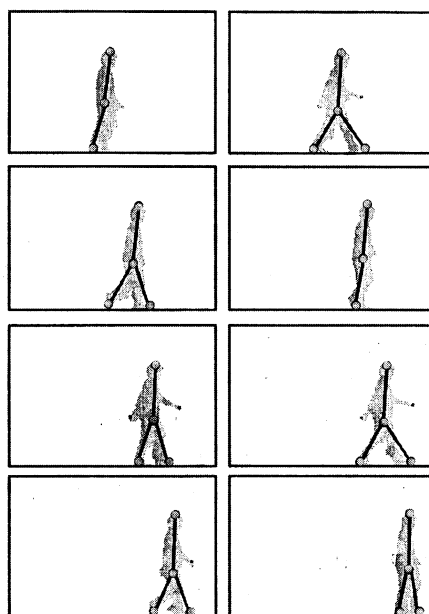


(a) 歩き

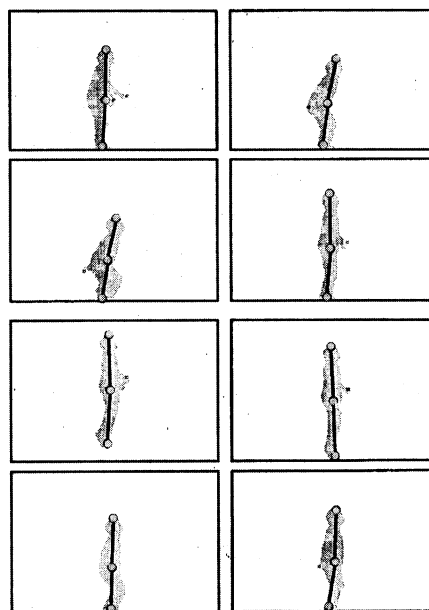


(b) ジャンプ

図 11 モデルとなる活動  
(活動は「歩き」と「ジャンプ」・8 フレームを抜粋)



(a) 歩き



(b) ジャンプ

図 12 実画像から得たスター型スケルトン  
(活動は「歩き」と「ジャンプ」・8 フレームを抜粋)