Vol.2010-CVIM-170 No.30
2010/1/21

ハッシュ関数を用いた行動認識

植浦　総一郎[†1]　岩井　儀雄[†1]　石黒　浩[†1]

本研究では，画像によるモニタリングの自動化を目的として，動画像中の人物の行動を認識するシステムの構築を目指している．一般に画像と姿勢は一対多の関係であり，入力画像に対し複数の姿勢が候補として存在する．そこで本研究では，入力画像に対し複数の姿勢を候補として保持しつつ，時間軸方向に確率を展開していくような行動認識のフレームワークを提案する．画像空間から姿勢空間への写像をハッシュ関数を用いることで学習し，姿勢が遷移するパターンをモデル化することで行動認識を実現する．本稿では，画像変数と姿勢変数を用いた行動認識モデルと，ハッシュ関数を用いた行動認識システムについて提案し，実環境における実験を通して提案システムの有用性を評価する．

Action Recognition By Using Hash Function

Soichiro Ueura,[†1] Yoshio Iwai[†1] and Hiroshi Ishiguro[†1]

We propose a system for human action recognition from image sequences to realize automatic monitoring. Because the relationship between images and postures is one-to-many correspondence, there are multiple postures as candidates for a input image. So we propose a framework to recognize actions keeping multiple postures for an input image and expanding probability as the time advances. We realize action recognition by learning mapping between images and postures by using a hash function and by modeling of posture transition patterns. We show the effectiveness of our system through an experiment in a real environment.

1. はじめに

近年，各種センサの発達により身体動作をリアルタイムで獲得し，行動を理解しようという研究が積極的に行われている．特に，非接触で計測できる画像処理による身体動作の認識は，ユーザへの負担とならず自然に行えるので様々な方法が提案されている．例えば医療の場において，患者のベッド周辺で起こる事故，あるいは事故につながりかねない危険な行動を自動検出するシステム[1]が提案されている．他にもセキュリティ[2]やジェスチャ認識[3]など様々な応用が考えられている．そこで，本研究では画像処理により人物行動を認識するシステムの構築を目指す．

画像処理による行動認識の手順は概して，(1) 入力画像から画像特徴量を抽出し，(2) 画像特徴空間における特徴点の時間的変化をモデルを用いて学習及び認識するというものである．言い換えると，(1) において画像特徴量により人体姿勢の近似を行い，(2) で姿勢の時間変化を画像特徴空間における特徴点の遷移として捉えることで行動のモデル化を行っている．この際，(1) の画像特徴量による姿勢の近似が行動認識の精度に関わる重要な問題であると言える．画像特徴量にはシルエット[4]や輪郭[5]，オプティカルフロー[6]などといったものが一般的であるが，いずれも姿勢パラメータと必ずしも相関が高いわけではなく，それが原因で認識精度の低下が発生する場合がある．つまり，上記のような画像特徴量を扱う場合，一つの画像に対し複数の姿勢の候補が存在する可能性を考慮しなくてはならない．

そこで本研究では，入力画像に対し複数の姿勢を候補として保持しつつ，時間軸方向に確率を展開していくような行動認識のフレームワークを提案する．まず (1) において，画像対姿勢の一対多の関係を写像として学習する．写像学習の基本的なアイデアは，大量の画像データと姿勢データを事例データとして取得し，画像特徴を基に候補となる姿勢を絞り込むというものである．候補の絞り込みには，部分空間法に基づくもの[7]や，グラフベースのもの[8]や，2 分木を用いるもの[9]や，LSH と呼ばれるハッシュ関数を用いたもの[10),11)]などが提案されており，いずれも高速近傍探索手法として注目されている．しかし，これらの手法は入力空間（画像特徴空間）における近傍探索手法であり，パラメータ空間（姿勢空間）を考慮したものではない．そこで本研究では，画像特徴からの候補姿勢の絞り込みを目的として Parameter Sensitive Hashing(PSH)[12)–14)] と呼ばれる近傍探索手法を導入する．PSH の特徴として，パラメータ空間における類似性を入力空間に反映させたハッシュ関数を設計することが可能であり，入力画像から精度の良い姿勢推定が実現できる．またハッシュ関数の特性から，画像探索における距離計算が不要となり認識速度が高速になるという効果も期待できる．次に，(2) においては，姿勢変数と画像変数を用いたマルコフ連鎖モデルにより行動パターンを学習する．その際，姿勢変数と画像変数間のパラメータ学習にハッシュ関数を利用できるようにする．

†1 大阪大学大学院基礎工学研究科システム創成専攻
Graduate School of Engineering Science, Osaka University

©2010 Information Processing Society of Japan

以下，本稿では PSH の概要と提案する行動認識モデルを説明した後に，ハッシュ関数を用いた行動認識システムについて述べる．次に実験とその結果及び考察について述べ，最後に結論としてまとめる．

2. Parameter Sensitive Hashing (PSH)

本研究では，画像特徴から候補となる姿勢の分布を学習するために Parameter Sensitive Hashing (PSH)[12] と呼ばれる高速近傍探索手法を導入する．PSH は主に事例ベース推論に基づく人物姿勢推定の分野で用いられる高速探索手法である．事例ベース推論に基づく姿勢推定では，入力特徴ベクトルと事例データの特徴ベクトルとの類似度導出計算を行い，類似度が最大となる事例データの姿勢を推定結果として出力するが，事例データ数の増加に従い，類似度導出計算の回数が増え，計算コストが増加してゆく．PSH は，この類似度導出処理の対象（正解姿勢候補）をハッシュ関数により絞り込むことで計算コストの削減を行っている．また，PSH の特徴としてハッシュ関数により絞り込まれたデータはパラメータ空間（本研究では姿勢空間）において類似性が保証されており，本研究の目的の一つである画像特徴空間から姿勢空間への正確な写像学習に適している．以下では，PSH のハッシュ関数について説明する．

PSH のハッシュ関数は式 (1) のように L 個 ($L \leq \frac{XY}{xy}$) の decision stump $\{d_{a_{(l)}}(f(\mathbf{z}))\}_{l=1}^{L}$ を bit 連結したもので，$\{d_{a_{(l)}}(f(\mathbf{z}))\}_{l=1}^{L}$ の出力に従ってハッシュ値が決定される．なお，decision stump $d_{a_{(l)}}(f(\mathbf{z}))$ は式 (2) のようにバイナリ出力の関数で，参照する特徴ベクトル $f(\mathbf{z})$ の要素 $a_{(l)}$ と閾値 T_l によって定義される．

$$h(f(\mathbf{z})) = [d_{a_{(1)}}(f(\mathbf{z})), ..., d_{a_{(L)}}(f(\mathbf{z}))]^T \tag{1}$$

$$d_{a_{(l)}}(f(\mathbf{z})) = \begin{cases} 1 & if\ f_{a_{(l)}}(\mathbf{z}) \leq T_l \\ 0 & otherwise \end{cases} \tag{2}$$

PSH では，同一のハッシュ値を持つデータペアは類似した姿勢であり，異なるハッシュ値を持つデータペアは類似していない姿勢であることを特徴として持つ．つまり，類似ペアのハッシュ値が等しくなる確率が高く，非類似ペアのハッシュ値が異なる確率が高くなるように，閾値 T_l を決定することが求められる（アルゴリズムは図 1 参照）．

N_T : *the number of test threshold samples for* $a_{(l)}$
N_{all} : *total data number in database*
r_1, r_2 : *risk parameter*
T_l : *optimal threshold for* $a_{(l)}$

for $k = 1$ *to* N_T **do**
  $T_k \leftarrow \frac{k}{N_T}$
  $\mathrm{R}(T_k) \leftarrow 0$
  for $i = 1$ *to* N_{all} **do**
    for $j = 1$ *to* N_{all} **do**
      if $\mathrm{sim}(\Theta_i, \Theta_j) = -1$ and $d_{a_{(l)}}(f(\mathbf{z}_i)) = d_{a_{(l)}}(f(\mathbf{z}_j))$
      then $\mathrm{R}(T_k) \leftarrow \mathrm{R}(T_k) + r_1$
      if $\mathrm{sim}(\Theta_i, \Theta_j) = 1$ and $d_{a_{(l)}}(f(\mathbf{z}_i)) \neq d_{a_{(l)}}(f(\mathbf{z}_j))$
      then $\mathrm{R}(T_k) \leftarrow \mathrm{R}(T_k) + r_2$
    end for
  end for
end for

$T_l \leftarrow \arg\min_{T_k} \mathrm{R}(T_k)$

図 1 最適な閾値 T_l 導出のためのアルゴリズム

このアルゴリズムでは，類似ペアに異なるハッシュ値が割り当てられた場合か，非類似ペアに同一のハッシュ値が割り当てられた場合に，値が増加するようなリスク関数 R が設計されており，リスク関数の値を最小化する閾値 T_l を最適な閾値としている．

ここで，$\mathrm{sim}(\mathbf{\Theta}, \mathbf{\Theta}')$ は式 (3) のように類似・非類似を表すラベルを与える関数である．

$$\mathrm{sim}(\mathbf{\Theta}, \mathbf{\Theta}') = \begin{cases} +1 & if\ d_\theta(\mathbf{\Theta}, \mathbf{\Theta}') < R/(1+\epsilon_\theta) \\ -1 & if\ d_\theta(\mathbf{\Theta}, \mathbf{\Theta}') > R \\ \text{not defined} & otherwise \end{cases} \tag{3}$$

ただし，類似度を定める基準として，関節角度ベクトル $\mathbf{\Theta} \in \mathrm{R}^{r_\theta}$ の関節角度平均距離 $d_\theta(\mathbf{\Theta}, \mathbf{\Theta}')$(式 (4) 参照) を用いている．

$$d_\theta(\mathbf{\Theta}, \mathbf{\Theta}') = \frac{1}{r_\theta/2} \sum_{i=0}^{r_\theta/2-1} \sqrt{\sum_{j=1}^{2} (\theta_{2i+j} - \theta'_{2i+j})} \tag{4}$$

ⓒ2010 Information Processing Society of Japan

また，R は類似・非類似を識別する閾値で，ϵ_θ は識別器の locality-sensitive[10] な性質を保証するための変数である．

上記の処理を特徴ベクトルの各要素 $a_{(l)}$ に対し行うことで，decision stump の集合 $\{d_{a_{(l)}}(f(z_t))\}_{l=1}^{\frac{XY}{xy}}$ が求まる．この集合から類似・非類似に対する分離性能の高い上位 L 個の decision stump を抜き出して，それらを bit 連結することでハッシュ関数を構築する．

なお，PSH は姿勢推定の精度向上のために L 個の decision stump からランダムに $s(< L)$ 個を選ぶという操作を K 回繰り返し，K 通りのハッシュ値に変換を行うが，写像が複雑化するため本研究では複数のハッシュ値への変換は行わない．つまり，L 個の decision stump により構築した単一のハッシュ関数から得られたハッシュテーブルを基に，画像特徴空間から姿勢空間への写像を学習する．ハッシュテーブルを用いた写像学習の詳細については 4.5 節で説明する．

3. 行動認識モデル

3.1 姿勢遷移モデル

本研究で行動認識に用いる確率モデルを姿勢遷移モデルと呼び，そのモデル図を図 2 に示す．ここでは，横軸方向で姿勢の時間変動を表現しており，縦軸方向で姿勢から画像への空間変動を表現している．姿勢遷移モデルで用いられる変数集合は，姿勢クラス $\mathbf{M} = \{m^1, ..., m^M\}$，画像クラス $\mathbf{I} = \{i^1, ..., i^N\}$ であり，モデルパラメータ集合は $\boldsymbol{\lambda} = \{\boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\phi}\}$ を用いて以下のように記述される．

- 初期姿勢確率 $p(\mathbf{m}_1|\boldsymbol{\pi})$：
  $\boldsymbol{\pi} = \{\pi_i | \pi_i = p(\mathbf{m}_1 = m^i)\}(i = 1, ..., M)$
- 姿勢遷移確率 $p(\mathbf{m}_t|\mathbf{m}_{t-1}, \mathbf{A})$：
  $\mathbf{A} = \{a_{ij} | a_{ij} = p(\mathbf{m}_t = m^j | \mathbf{m}_{t-1} = m^i)\}(i, j = 1, ..., M)$
- 画像出力確率 $p(\mathbf{i}_t|\mathbf{m}_t, \boldsymbol{\phi})$：
  $\boldsymbol{\phi} = \{\phi_{jk} | \phi_{jk} = p(\boldsymbol{\phi}_t = i^k | \mathbf{m}_t = m^j)\}(j = 1, ..., M, k = 1, ..., N)$

なお，姿勢遷移モデルは一次のマルコフ性を仮定しており，ある時刻での姿勢の状態は一時刻前の姿勢の状態にのみ依存している．また，各画像クラスが出力される確率は同時刻における姿勢の状態にのみ依存すると仮定している．姿勢遷移モデルでは学習時において，事例データの姿勢クラス間の遷移確率と姿勢クラスと画像クラスの確率密度分布から容易にモデルパラメータを推定することができる．

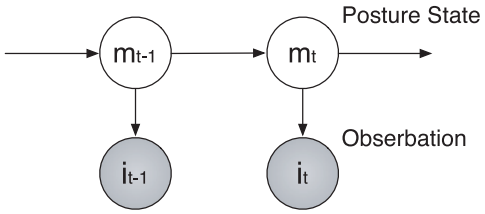


図 2 姿勢遷移モデル

3.2 学　　習

姿勢遷移モデルによる認識系では各カテゴリごとに一つのモデルから構成される．認識対象の各カテゴリ C_i ごとのモデルパラメータを $\boldsymbol{\lambda}^i (= \{\boldsymbol{\pi}^i, \mathbf{A}^i, \boldsymbol{\phi}^i\})$ とする．$\boldsymbol{\lambda}^i$ の学習を C_i の学習用パターンを用いて行う．なお，簡単のため以下ではカテゴリ C_i のモデルパラメータを $\boldsymbol{\lambda} = \{\boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\phi}\}$ とする．カテゴリ C_i の学習用パターン数を L，パターン l のフレーム数を T_l としたとき，事例データを $\{(\mathbf{i}_\tau^l, \mathbf{m}_\tau^l)_{\tau=1}^{\tau=T_l}\}_{l=1}^L$ により表現する．このとき，初期姿勢確率 $p(\mathbf{m}_1|\boldsymbol{\pi})$ の要素 π_i は式 (5) により得られる．

$$\pi_i = \frac{\sum_{l=1}^L \delta(\mathbf{m}_1^l = m^i)}{L} \tag{5}$$

ただし $\delta(\cdot)$ は引数の条件を満たした場合に 1 を返し，それ以外の場合に 0 を返す関数とする．これは，学習用パターンの行動開始時 ($t = 1$) における姿勢クラスの確率密度分布を表している．

次に，姿勢遷移確率 $p(\mathbf{m}_t|\mathbf{m}_{t-1}, \mathbf{A})$ の要素 a_{ij} は式 (6) により得られる．

$$a_{ij} = \frac{\sum_{l=1}^L \sum_{t=2}^{T_l} \delta(\mathbf{m}_t^l = m^j, \mathbf{m}_{t-1}^l = m^i)}{\sum_{l=1}^L (T_l - 1)} \tag{6}$$

これは，学習用パターンの一時刻前の姿勢 i が現時刻の姿勢 j へ遷移する頻度を表している．

最後に，画像出力確率 $p(\mathbf{i}_t|\mathbf{m}_t, \boldsymbol{\phi})$ はハッシュ関数を用いて学習する．ハッシュ関数により画像空間から姿勢空間への写像 $p(\mathbf{M}|\mathbf{I})$ を学習し，ベイズルールを用いて逆写像 $p(\mathbf{I}|\mathbf{M})$ を導出する．任意の時刻において画像出力確率はモデルパラメータ $\boldsymbol{\phi}$ にのみ依存することを前提としているので，本研究ではここで得られた逆写像 $p(\mathbf{I}|\mathbf{M})$ を画像出力確率 $p(\mathbf{i}_t|\mathbf{m}_t, \boldsymbol{\phi})$ として用いる．詳しくは 4.5 節で説明する．

なお，モデルパラメータの学習手法は以上のように事例データの分布に基づくものであるが，事例データ数が不十分な場合，推定されたパラメータ π_i, a_{ij}, ϕ_{jk} の値が 0 になり，各時

ⓒ2010 Information Processing Society of Japan

刻ごとに計算される尤度が，順次 0 になることがある．この問題を回避する方法として，事例データに含まれない観測に対しても，ある程度の確率値を与えることが考えられる．そこで，$\pi_i, a_{ij}, \phi_{jk} \gg \epsilon \approx 0$ となるような ϵ をオフセットとして与える．以上の手法で得られた 3 つのモデルパラメータを用いて観測画像系列に対するカテゴリ C_i ごとのモデルの尤度計算を行う．

3.3 認　　識

入力として画像クラス系列 $\mathbf{i}_1, \mathbf{i}_2, ..., \mathbf{i}_t$ が観測されたとする．このときカテゴリごとのモデルの尤度は式 (7) により求められる．

$$P(\mathbf{i}_1, \mathbf{i}_2, ..., \mathbf{i}_t|\lambda^l) = \sum_{j=1}^{M}\sum_{i=1}^{M} P(\mathbf{m}_{t-1} = m^i, \mathbf{i}_1, \mathbf{i}_2, ..., \mathbf{i}_{t-1}|\lambda^l) a_{ij}\phi_{jk}(\mathbf{i}_t) \tag{7}$$

ただし，$\phi_{jk}(\mathbf{i}_t)$ とは時刻 t での観測 $\mathbf{i}_t$ に対応する $\boldsymbol{\phi}$ の要素とする．ここで式 (7) の左辺を周辺化して，

$$P(\mathbf{i}_1, \mathbf{i}_2, ..., \mathbf{i}_t|\lambda^l) = \sum_{j=1}^{M} P(\mathbf{m}_t = m^j, \mathbf{i}_1, \mathbf{i}_2, ..., \mathbf{i}_{t-1}|\lambda^l) = \sum_{j=1}^{M} \alpha_j(t) \tag{8}$$

とおくと，式 (7) は以下のように書き換えることが出来る．

$$\alpha_j(t) = \sum_{i=1}^{M} \alpha_j(t-1) a_{ij}\phi_{jk}(\mathbf{i}_{t-1}) \tag{9}$$

よって，式 (9) により逐次的に尤度を計算することが可能となる．なお，初期値は $\alpha_i(0) = \pi_i$ により得られる．このとき，モデルの事前分布が一様分布と仮定すると，式 (10) のように尤度が最大となるカテゴリ C を選択することで行動を認識する．

$$C = \arg\max_l P(\mathbf{i}_1, \mathbf{i}_2, ..., \mathbf{i}_t|\lambda^l) \tag{10}$$

4. ハッシュ関数を用いた行動認識

4.1 提案システム概要

3 章で提案した姿勢遷移モデルを用いた行動認識システムの処理の流れを図 3 に示す．まず，学習段階において学習用パターンである同期撮像された時系列画像データと時系列姿勢データから各フレームを切り出すことにより画像データと姿勢データのペアを事例データとして取得する．3.1 節で述べたように，姿勢遷移モデルにおける変数は離散値をとるため，画像データと姿勢データをそれぞれクラス分類する必要がある．そこでまず，姿勢データをクラスタリングすることにより．姿勢クラス $\mathbf{M} = \{m^1, ..., m^M\}$ を得る．そして，事例データの時間的な姿勢遷移の情報から 3.2 節の手法に基づきモデルパラメータ $\boldsymbol{\pi}, \mathbf{A}$ を学習する．次に，画像データから画像特徴量抽出を行い，画像特徴量を基にハッシュ関数を用いて画像クラス $\mathbf{I} = \{i^1, ..., i^N\}$ を得る．なお，同一のハッシュ値を持つ事例データは同一の画像クラスに属するものとする．そして，ハッシュテーブルの各インデックスに対応する事例データをリストとして保存し，リスト内の姿勢クラスの確率密度分布 $p(\mathbf{M}|\mathbf{I})$ を基にモデルパラメータ

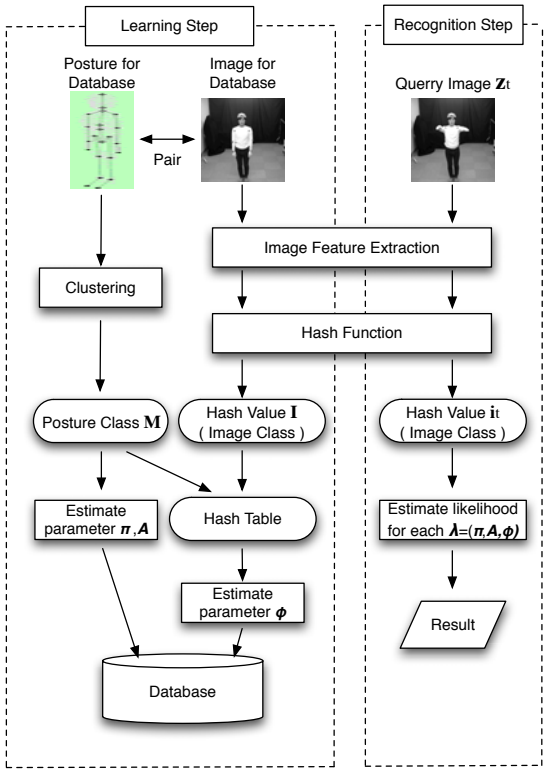


図 3　提案システムの処理の流れ

ⓒ2010 Information Processing Society of Japan

$\boldsymbol{\phi}$（画像出力確率）を学習する．認識段階においては，観測画像 $\mathbf{z}_t$ から画像特徴量を抽出し，得られた特徴量からハッシュ関数により画像クラス $\mathbf{i}_t$ を得る．得られた画像クラス $\mathbf{i}_t$ から 3.3 節の手法を用いて，各カテゴリごとのモデル尤度の計算を行うことで行動認識をする．

以下の節で，姿勢データのクラスタリング，本研究で使用する画像特徴量，ハッシュ関数の設計手法，及びハッシュテーブルを用いた画像出力確率の学習について述べる．

4.2 姿勢データのクラス分類

姿勢パラメータをクラスタリングすることで姿勢データのクラス分類を行う．ここでは，姿勢パラメータとして図 4 で示すような上半身の関節角度 $\Theta = \{\theta_i | i = 1, 2, ..., 8\}$ を用いている．また，クラスタリングのアルゴリズムとして k-means 法を利用している．このとき，姿勢クラスの数 M はクラスタ数 k と一致し，同一のクラスタに属する姿勢データは同一の姿勢クラスに分類されるものとする．

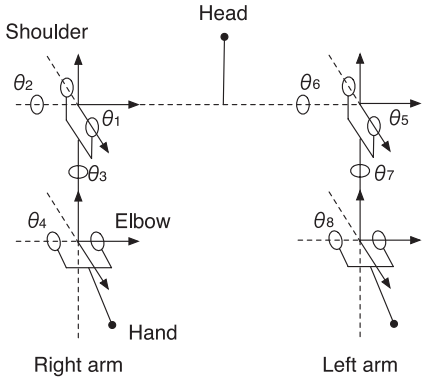


図 4　上半身の姿勢パラメータ

4.3 画像特徴量

本研究では，画像特徴量として，着衣や体格の違いによるシルエット画像の変化をある程度吸収できるメッシュ特徴量を用いる．メッシュ特徴量とは図 5 のように背景差分により得られた二値画像 ($X \times Y$ 画素) を $x \times y$ 画素 ($x \leq X, y \leq Y$) のメッシュに分割し，各メッシュの白い画素の含有率を特徴ベクトルの各要素とする．つまり，特徴ベクトル f は式 (11) の形で表され，特徴ベクトルの各要素は式 (12) で表される．

$$\mathrm{f} = (f_{a_{00}}, f_{a_{01}}, ..., f_{a_{ij}}, , ..., f_{a_{\frac{X}{x}\frac{Y}{y}}}) \tag{11}$$

$$f_{a_{ij}} = \text{number of white pixels in mesh}(i, j)/xy \tag{12}$$

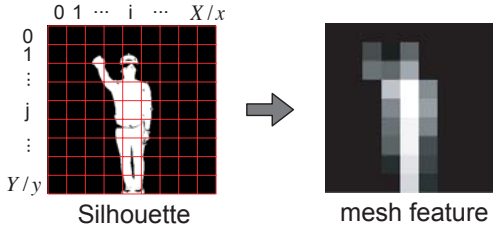


図 5　メッシュ特徴量

4.4 ハッシュテーブル

2 章の手法により設計したハッシュ関数により，事例データをハッシュテーブルに登録する（図 6 参照）．ハッシュテーブルの各バケットには，同一のハッシュ値を持つ事例データがリストとして保存されている．本研究では，同一のハッシュ値を持つ事例データは同じ画像クラスに所属すると定義する．なお，一部でリストを持たないバケットが存在するが，用意されたバケットの範囲に対し，事例データの取りうるハッシュ値の範囲が限られているためである．例えば，ハッシュ値が $[0, 0, ..., 0]$ のようにすべての要素が 0 となるような事例データは存在しない．

仮に，認識時に入力画像から得たハッシュ値に対応するバケットに事例データが存在しなかった場合，本研究では一時刻前のハッシュ値を参照することで対応する．

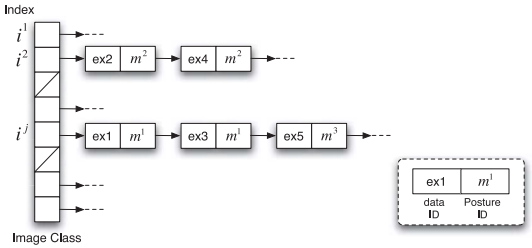


図 6　ハッシュテーブル

©2010 Information Processing Society of Japan

4.5 ハッシュテーブルを用いた画像出力確率の学習

本研究ではハッシュテーブルにより分類された事例データを用いて画像出力確率の学習を行う．画像クラス i^j ごとの姿勢クラス m^i の確率密度分布は，ハッシュテーブルのインデックス i^j に対応するリスト L_{ij} を用いることで式 (13) のように表される．

$$p(m^i|i^j) = \frac{\sum_{\mathbf{m}_l \in L_{ij}} \delta(\mathbf{m}_l = m^i)}{card(L_{ij})} \tag{13}$$

ここで $p(m^i|i^j)$ はリスト内の姿勢クラス m^i の頻度を表している．このとき，画像空間から姿勢空間への写像を $p(m^i|i^j)$ を各要素とする行列 $p(\mathbf{M}|\mathbf{I}) \in \mathbf{R}^{N\times M}$ によって表現する．よって，逆写像（姿勢空間から画像空間への写像）$p(\mathbf{I}|\mathbf{M}) \in \mathbf{R}^{M\times N}$ はベイズルールにより式 (14) のように表される．

$$p(\mathbf{I}|\mathbf{M}) = \frac{p(\mathbf{M}|\mathbf{I})p(\mathbf{I})}{p(\mathbf{M})} \tag{14}$$

なお，$p(\mathbf{I})$ は画像クラスの事前分布で，$p(\mathbf{M})$ は姿勢クラスの事前分布を表している．姿勢遷移モデルでは，任意の時刻において画像出力確率は同一のパラメータ $\boldsymbol{\phi}$ により制御されていることを前提としているため，式 (15) が成立する．

$$p(\mathbf{i}_t|\mathbf{m}_t, \boldsymbol{\phi}) = p(\mathbf{I}|\mathbf{M}) \tag{15}$$

5. 行動認識実験

5.1 実験環境

提案システムを用いた行動認識実験について説明する．実験では図 7 に示すように，上半身の動作を 16 種類用意し，それぞれが正しく識別されるかという内容の実験を行った．6 名の被験者が各動作を 15 回ずつ試行した際に取得した画像データと姿勢データを基に，事例データを作成した（表 1 参照）．

表 1 被験者データ

被験者	性別	身長	事例データ数
A	M	172	7267
B	M	169	7318
C	M	168	7277
D	M	167	7268
E	F	167	7358
F	F	158	7209

画像データのサイズは 360 × 360 [pixeles] で，身体各部位の姿勢パラメータは Nac 社の MAC3D System を用いて取得した．実験に使用した計算機は，Mac Pro (CPU:2 × 2.8Hz Quad-Core Intel Xeon , Memory:4GB 800MHz DDR2 FB-DIMM) である．また，メッシュサイズは 90 × 90 [pixels] で姿勢クラス数は $M = 18$ とした．



図 7 16 種類の行動

5.2 実　　験

認識精度

認識精度を評価するために，学習と認識に同一人物のデータを用いる場合と異なる人物を用いる場合の二種類の実験を行った．まず，学習と認識に同一人物のデータを用いる場合の実験では，6 人の被験者それぞれに対し事例データを学習用パターン 10 本，テスト用パターン 5 本に分離し，テスト用パターンを入力した際の認識率を導出した．その結果，平均認識率 94.2% という高い認識率が確認された．次に，学習と認識に異なる人物を用いる場合の実験では，1 人をテスト，残りの 5 人を学習に用いて，テストに用いる人物を入れ替えながら認識率を導出する交差検定 (6-fold cross validation) を行った．その結果，平均認識率

ⓒ2010 Information Processing Society of Japan

は 84.4% であった．また，それぞれの動作の識別結果を混同行列として図 8 に示す．

	act1	act2	act3	act4	act5	act6	act7	act8	act9	act10	act11	act12	act13	act14	act15	act16
act1	0.778	0	0.111	0	0	0	0.022	0	0	0	0	0	0.089	0	0	0
act2	0.033	0.911	0	0	0	0	0	0	0	0	0	0	0	0.044	0	0.011
act3	0.011	0	0.933	0	0	0	0	0	0.022	0	0	0	0.033	0	0	0
act4	0	0	0	0.922	0	0	0.011	0	0	0.067	0	0	0	0	0	0
act5	0.133	0	0	0	0.756	0	0.011	0	0.067	0.033	0	0	0	0	0	0
act6	0	0	0	0	0	0.622	0	0	0.033	0.333	0	0	0.011	0	0	0
act7	0	0	0	0	0	0	0.978	0	0.022	0	0	0	0	0	0	0
act8	0	0.022	0	0	0	0.011	0.011	0.844	0	0.111	0	0	0	0	0	0
act9	0	0	0	0	0	0	0.056	0	0.933	0.011	0	0	0	0	0	0
act10	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
act11	0.089	0	0	0	0	0	0	0	0	0	0.867	0.011	0	0	0.033	0
act12	0	0.033	0	0	0	0	0	0	0	0	0.067	0.767	0.044	0	0	0.089
act13	0.022	0	0	0.011	0	0	0.056	0	0	0	0	0	0.744	0	0.167	0
act14	0	0.078	0	0	0	0	0	0.056	0	0	0	0	0	0.644	0	0.222
act15	0	0	0	0	0	0	0	0	0	0	0.011	0	0.133	0	0.856	0
act16	0	0	0	0	0	0	0.011	0.011	0	0	0	0.022	0	0.011	0	0.944

図 8　混同行列

一部の動作で認識率が低いものがあるが，これは画像特徴空間上での軌跡の近い動作と混同したためである．その例として act1 を act13 と認識したときの結果を図 9 に示す．今回実験で使用したメッシュ特徴は画像を 4×4 等分するものであり，姿勢を詳細に区別をすることは難しい．しかし，提案システムにおいて，メッシュのサイズを小さくすることで，認識率が上昇する訳ではない．メッシュのサイズを小さくすると，特徴ベクトルの次元が増えるため，ハッシュ関数に用いる decision stump の候補の数 $\frac{XY}{xy}$ が増加する．本研究では，この候補からリスク関数 $\mathrm{R}(T_l)$ の値を基準に認識に使用する特徴ベクトルの要素 $a_{(l)}$ を決定しているが，ここで上位に選出される要素同士は相関の高いものである可能性が高い．そのため，画像のクラス分類に必要な要素の数（decision stump の数 L の値）を増やす必要があり，結果として画像クラスも増加する．このとき，画像クラスごとの事例データの数は減少し，姿勢の分布を正確に表現することは難しくなる．以上の問題の対処方法としては，事例データを統計的妥当性が保証されるまで増やすか，L の値を出来る限り低く抑えることが挙げられる．今回，軸の選択基準としてリスク関数の値を用いたが，今後は主成分分析により特徴ベクトルの相関の低い要素を残す等の処理を施すことで L の値を低く抑える手段を試みる必要がある．また，学習に用いる人物の数（テストとは別の人）を増やしたときの認識率の変化を図 10 に示す．このグラフから，人数を増やすに従って認識率が上昇することがわかるが，5 人の段階ではまだ上昇の途中であることから認識率の上限まで到達していないと考えられる．よって，さらに多くの人物の事例データを生成することで認識率が向上すると考えられる．

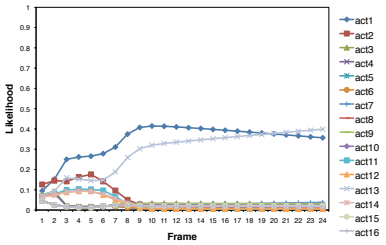


図 9　誤認識の例

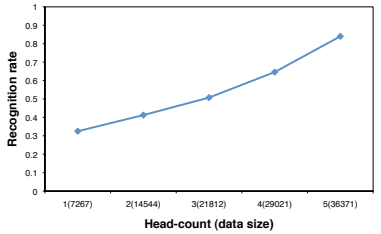


図 10　学習に用いる人物の数と認識率の関係

処 理 速 度

次に，認識に要する時間を測定した．シルエットが入力されてから最大尤度を持つモデルが決定されるまでの 1 フレーム当たりの計算時間を計測した結果，平均 3.86[ms] と高速であることが確認された．また，図 11 に示すように事例データ数によらず処理速度は一定であることが確認された．これは，ハッシュ関数の特性として探索が $O(1)$ であることから明らかである．

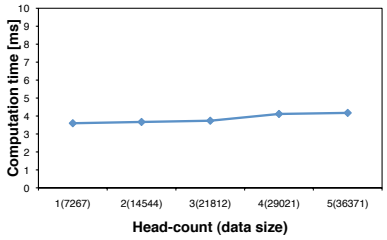


図 11　事例データ数と計算時間の関係

©2010 Information Processing Society of Japan

一連動作の認識

図 12 に複数の動作を連続で行ったときの認識結果（尤度の遷移）を示す．なお，一連動作は act2,act1,act4,act3 の順に行っている．グラフから，それぞれの動作が始まる時刻でそのカテゴリのモデルの尤度が上昇する様子が確認できる．

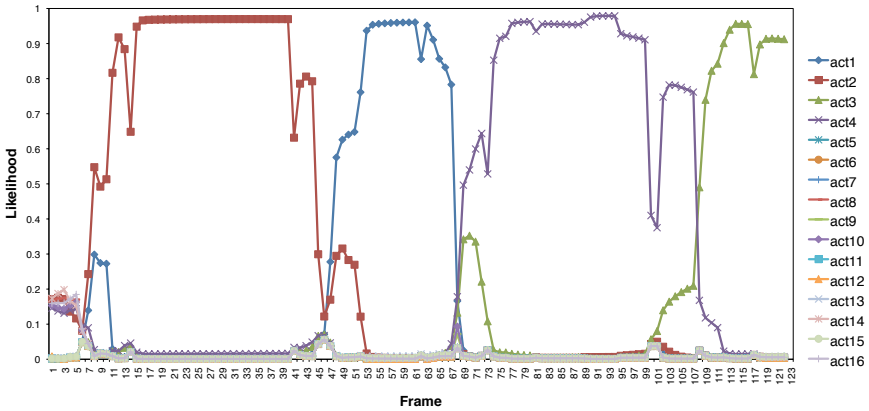


図 12 一連動作の認識

6. おわりに

本研究では画像処理による人物行動認識システムの構築を行った．画像と姿勢には一対多の関係があるが，ハッシュ関数を利用する事でその関係を学習した．さらに，画像変数と姿勢変数を用いたマルコフ連鎖モデルを構築することで，モデルパラメータの学習にハッシュ関数を利用できる方法を提案した．行動認識実験によりテストと学習に用いるデータが同一人物のときで平均 94.2%，6 名の被験者による交差検定を行った場合で平均 84.4% の認識率が確認された．また，シルエット入力から結果出力までの 1 フレーム当たりの平均処理時間は 3.85 [ms] と高速であり，事例データ数によらず一定であることが確認された．提案システムの今後の課題として，より複雑な行動の認識や被験者を増やした場合の評価が挙げられる．

参 考 文 献

1) 岡田奈美，岡村健史郎，浜本義彦，“画像認識による患者の行動理解の試み，” 電子情報通信学会技術研究報告, PRMU，Vol. 103，No. 738，pp. 31–36，2004.
2) 青木茂樹，岩井嘉男，大西正樹，小島篤博，“人物の位置 · 姿勢に注目した行動パターンの学習 · 認識と非日常状態検出への応用，” 電子情報通信学会論文誌 D-II，Vol. J87-D-II，No. 5，pp. 1083–1093，2004.
3) 樹木義道，岩井儀雄，谷内田正彦，“‘HMM を利用したジェスチャー認識，” 電子情報通信学会技術研究報告, PRMU，Vol. 100，No. 565，pp. 25–30，2001.
4) D. Weinland, E. Boyer, and R. Ronfard, “Action recognition from arbitrary views using 3d exemplars,” *ICCV*, pp. 1–7, 2007.
5) A. Elgammal, V. Shet, Y. Yacoob, and L.S. Davis, “Learning dynamics for exemplar-based gesture recognition,” *CVPR*, Vol. 1, pp. 571–578, 2003.
6) 畠直志，岩井儀雄，谷内田正彦，“動き情報と情報圧縮を用いたロバストなジェスチャ認識手法，” 電子情報通信学会論文誌 D，Vol. J81-D2，No. 9，pp. 1983–1992，1998.
7) 大町真一郎，大町方子，“多項式近似を用いた部分空間法による高速画像探索，” 電子情報通信学会論文誌 D，Vol. 191-D，No. 6，pp. 1561–1568，2008.
8) 青山一生，斉藤和巳，山田武士，上田修功，“グラフ索引構造を用いた高速類似探索，” 電子情報通信学会論文誌 D，Vol. 192，No. 9，pp. 1543–1554，2009.
9) S.Arya, D.Mount, R.Silverman, and A.Y.Wu, “An optimal algorithm for approximate nearest neighbor searching,” *ACM*, Vol. 45, No. 6, pp. 891–923, 1998.
10) A.Gionis, P.Indyk, and R.Motwani, “Similarity search in high dimensions via hashing,” *VLDB*, pp. 518–529, 1999.
11) M.Datar, N.Immorlica, P.Indyk, and V.Mirrokni, “Locality-sensitive hashing scheme based on p-stable distributions,” *ASCG*, pp. 253–262, 2004.
12) G.Shakhanarovich and T.Darrell P.Viola, “Fast pose estimation with parameter sensitive hashing,” *ICCV*, Vol. 2, pp. 750–757, 2003.
13) Y.Sagawa，M. Shimosaka, and T.S. T.Mori, “A fast near neighbor search metric for online example-based human pose estimation,” 電子情報通信学会技術研究報告, PRMU, Vol. 2007, No. 87, pp. 231–238, 2007.
14) P. Jain, B. Kulis, and K. Grauman, “Fast image search for learned metrics,” *CVPR*, pp. 1–8, 2008.

©2010 Information Processing Society of Japan