

準教師有りクラスタリングによる行動素抽出

植浦 総一郎[†], 岩井 儀雄[†], 谷内田 正彦[†]

[†]大阪大学大学院基礎工学研究科システム創成専攻

〒560-8531 大阪府豊中市待兼山町1-3

ueura@yachi-lab.sys.es.osaka-u.ac.jp, iwai@sys.es.osaka-u.ac.jp,

yachida@sys.es.osaka-u.ac.jp

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、動画像中の人物像の行動を認識するシステムの構築を目指している。システムの入力として画像データが与えられたとき、人物行動のパラメータを推定し行動認識を行うシステムを構築する。システム構築を実現するにあたり、事例データベース内の身体各部位の三次元位置データをクラスタリングすることにより、身体動作の基本要素である行動素の抽出を行うことを本研究の第一段階の目標とする。行動素を取り出すことができればそれらの遷移パターンにより行動モデルが構築でき、複雑な動作でも高速に計算できるようになる。本論文では提案システムの概要を示し、行動素抽出の手法について述べ、行動素抽出実験の結果を記す。

Extracting Action Primitives by Semi-supervised Clustering

Soichiro UEURA[†], Yoshio IWAI[†], Masahiko YACHIDA[†]

[†]Graduate School of Engineering Science, Osaka University

1-3 Machikaneyama, Toyonaka, Osaka 560-8531 Japan

ueura@yachi-lab.sys.es.osaka-u.ac.jp, iwai@sys.es.osaka-u.ac.jp,

yachida@sys.es.osaka-u.ac.jp

We have been constructing the system that recognizes human actions in image sequence. We collect image and 3D data of body parts by using a motion capture system. For the first step of constructing the system, we try to extract action primitives by clustering 3D data of parts of body. When action primitives can be extracted, human actions can be modeled by transitions of action primitives and any complex actions will be recognized fast. In this paper, we present a concept of action recognition system and a method for extracting action primitives and show experimental results.

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, 4, 5] など様々な応用が考えられている。そこで、本研究では非接触で動画像から人物行動を認識するシステムの構築を目指す。

動画像からの人物行動認識の従来手法には、隠れマルコフモデル (HMM) を利用したもの [2, 6] が多く提案されている。HMM は確率的に理論が展開しやすく、入力特徴量の時間伸縮にロバストであることが利点として知られている。しかし、欠点として、これまでの研究では一般に行動パターンと HMM が一対一に対応していたため、行動パターン分の HMM を用意する必要があった。このため、実環境でこれらの手法を用いる場合、記憶容量や計算コストが大きくなるという問題が生じていた。一方、音声認識の研究では、言語の音声を構成する最小単位である音素を用いた研究が行われている。音素を用いることで期待される効果として、必要学習データ数の減少、語彙拡張の簡単化、人物に依存しない認識など

が挙げられる [5].

近年, この効果を期待し, ジェスチャ認識の研究において, ジェスチャをさらに小さな単位 (これを行動素とする) に分け, 認識を行う研究が行われている [3, 4, 5]. 樹木らは, 行動素としてビデオから手作業で取り出したジェスチャをシンボル化したものを用いる手法を提案した [3]. 新宅は, 既存の行動素以外の行動素を観測した場合, それをデータベースに追加する手法を提案した [4]. Bauer らは, k-means アルゴリズムを用いて, 複数のジェスチャ内に存在する動作に対してクラスタリングを行い, 各クラスタを行動素として抽出する手法を提案した [5]. しかし, 樹木らや新宅の手法は個人差や試行差の出にくいジェスチャを扱うことを前提としており, 全身運動 (歩く, 座る, 等) を認識対象とする場合, 直接適用することはできない. また, Bauer らの手法では身体動作の事前知識無しでクラスタリングを行うため, 必ずしも行動を認識するのに都合の良い行動素が抽出されるとは限らない.

そこで本研究では, 学習データの一部に, 似た行動と異なる行動についての事前知識を教師情報として与え, 一連行動をクラスタリングするという準教師有りクラスタリング [7] の手法を用いて行動素抽出を行う. ただし, 本研究における行動素とは準教師有りクラスタリングで生成されたクラスタのことである. 行動素が抽出されれば, 行動モデルはそれらの遷移モデルとして表現することができる [9] ので, HMM を用いた従来手法の問題点を解消することができる.

以降, 第 2 章では本研究で提案する非接触行動認識システムの概要について述べ, 第 3 章では行動素抽出システムについて述べる. 第 4 章では行動素抽出実験とその結果について述べ, 実験結果をもとに, 得られた行動素の検証を行う. また, 第 5 章では結論と今後の課題について述べる.

2 非接触行動認識システム概要

本章では本研究で提案する非接触行動認識システムの概要について説明する. 提案システム概念図を図 1 に示す. 本研究では事例データベースにより一連行動の画像センサ入力と身体各部位の推定パラメータを結び付けて高速に行動認識を行う手法を提案する.

2.1 節では人物行動の学習手法について述べ, 2.2 節では人物行動の認識手法について述べる.

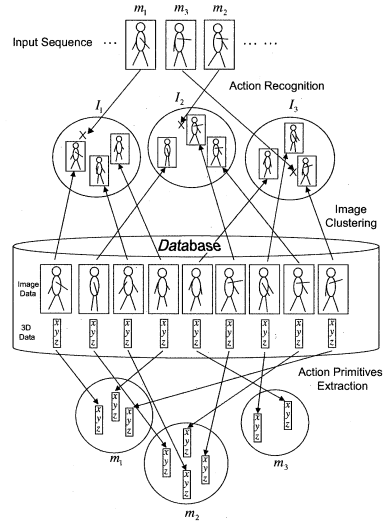


図 1: 提案システム概念図

2.1 行動学習

学習段階において人物行動の画像データとともに, モーションキャプチャシステムを利用することにより身体各部位の三次元位置データを収集し, これらを教師有り事例データとして事例データベースを生成する. 得られた事例データベース内の身体各部位の三次元位置データをクラスタリングすることにより, 身体動作の基本要素である行動素の抽出を行う. ただし, 本研究における行動素とは, 生成されたそれぞれのクラスタのことである. なお, 図 1 において各 m_i は得られた行動素を表している. クラスタリング手法としては, 人間が事例データの一部に似た動作の画像シーケンス, 異なる動作の画像シーケンスについての事前知識を教師情報として与え, その情報をもとにクラスタリングを行うという準教師有りクラスタリング [7] の手法をとる. 行動素の抽出手法の詳細については 3 章で述べる.

行動素を取り出すことができれば, それらの遷移により図 2 のような行動モデルを構築することができる. さらに認識段階で必要となる行動素間の遷移確率についても学習する. また, 行動認識のために, 事例データベース内の画像データについても同様に画像特徴量空間上で準教師有りクラスタリングを行う. なお, 図 1 において各 I_i は画像特徴量空間上で得られたクラスタを表している.

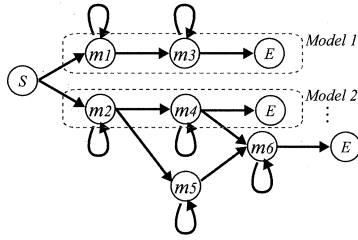


図 2: 行動モデル

2.2 行動認識

認識段階において入力画像データが与えられたとき、各画像クラスタ I_i との距離を比較し、距離が最小となるクラスタ I を決定する。クラスタ I の中から入力画像データと距離の近い教師画像データを数枚選択する。それら教師画像データに対応する行動素を調べ、その中の多数派の行動素を選択し、それを入力画像に対応する行動素とする。この方法は、入力画像データの近傍探索において、所属するクラスタ内の事例データのみを探索対象とするため、全事例データとの距離を計算する場合に比べ探索時間を短縮できる。

この手法を入力画像シーケンスに順次適用することで入力画像データ列が行動素列に変換される。そして、学習されている行動素間の遷移確率から事後確率が最大となるような行動モデルを選択することで行動認識を行う。

3 行動素抽出システム

本章では、行動素抽出システムについて説明する。学習対象となる身体各部位の三次元位置データをモーションキャプチャシステムを利用することにより取得し、得られた三次元位置データについて前処理、特徴量抽出、準教師有りクラスタリングを行うことで行動素抽出を試みる。

3.1 節、3.2 節、3.3 節でそれぞれ前処理、特徴量抽出、準教師有りクラスタリングについて述べる。

3.1 前処理

前処理では各部位の三次元位置データを補正することですべての行動において同条件で特徴量抽出ができる状態に変換する。処理内容として、代表点抽出

出及び欠損データの補間、座標系の変換、体の向きの一掃を順次行う。

3.1.1 代表点抽出及び欠損データの補間

ここでは、モーションキャプチャシステムにより得られた三次元位置データから各部位の代表点を抽出処理を行う。また、欠損データがある場合は補間処理を行う。それぞれの処理を以下に示す。

代表点抽出

各時刻について、そのときに有効なマーカー群から身体各部位の平均位置を計算する。ここで得られた時刻 t での各部位「頭」「右肩」「左肩」「右肘」「左肘」「右手」「左手」に対応するの三次元位置データをそれぞれ図 3 のように、 $\mathbf{x}_{he}(t)$ 、 $\mathbf{x}_{rs}(t)$ 、 $\mathbf{x}_{ls}(t)$ 、 $\mathbf{x}_{re}(t)$ 、 $\mathbf{x}_{le}(t)$ 、 $\mathbf{x}_{rh}(t)$ 、 $\mathbf{x}_{lh}(t)$ とする。

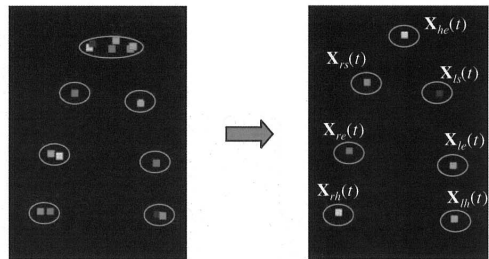


図 3: 代表点抽出

欠損データの補間

部位「頭」の平均位置が計算できなかった時刻について、前後の時刻から線形補間することにより位置を推定する。頭以外の部位で、平均位置が計算できなかった時刻については、頭-肩-肘-手の順で親子関係を作り、前後の時刻での親からのオフセットを球面線形補間して、現時刻での親からのオフセットを推定することで現在位置を推定する。

時刻 t において平均位置が計算できなかった頭以外の部位の三次元位置データを $\mathbf{x}_i(t)$ とする。その部位の親にあたる部位の三次元位置データを $\mathbf{x}_{i-1}(t)$ としたときの $\mathbf{x}_i(t)$ の補間処理の内容を以下に示し、補間処理の概要図を図 4 に示す。

時刻 t の前後の時刻で、部位 i の平均位置が計算できている時刻のうち t に最も近い時刻をそれぞれ t_{be} 、

t_{af} とする。このとき、前後の時刻での親からのオフセット $\mathbf{x}_{be}, \mathbf{x}_{af}$ を式 (1), 式 (2) により定義する。

$$\mathbf{x}_{be} = \mathbf{x}_i(t_{be}) - \mathbf{x}_{i-1}(t_{be}) \quad (1)$$

$$\mathbf{x}_{af} = \mathbf{x}_i(t_{af}) - \mathbf{x}_{i-1}(t_{af}) \quad (2)$$

\mathbf{x}_{be} と \mathbf{x}_{af} のなす角 ϕ は式 (3) のように求められる。

$$\phi = \arccos \frac{\mathbf{x}_{be} \cdot \mathbf{x}_{af}}{\|\mathbf{x}_{be}\| \cdot \|\mathbf{x}_{af}\|} \quad (3)$$

ここで、時刻 t での部位 i の部位 $i-1$ からのオフセットと \mathbf{x}_{be} のなす角 ϕ' を式 (4) のように推定する。

$$\phi' = \frac{t - t_{be}}{t_{af} - t_{be}} \phi \quad (4)$$

このとき、時刻 t での部位 i の三次元位置データ $\mathbf{x}_i(t)$ は式 (5) のようにして推定される。

$$\mathbf{x}_i(t) = \mathbf{x}_{i-1}(t) + R\left(\frac{\mathbf{x}_{be} \times \mathbf{x}_{af}}{\|\mathbf{x}_{be} \times \mathbf{x}_{af}\|}, \phi'\right) \mathbf{x}_{be} \quad (5)$$

ただし $R(\mathbf{V}, \phi)$ は回転軸 \mathbf{V} , 回転角 ϕ の回転行列である。

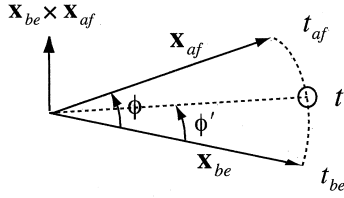


図 4: 欠損データの補間

3.1.2 座標系の変換

得られた各部位の三次元位置データは実験室内の一点を原点とする座標系により表現されており、部屋のどの場所にいるかという情報を含んでいる。しかし、行動素は人物の位置に依存せずに抽出されることが求められる。そこで時刻 t での座標系の原点 $\mathbf{x}_{org,t}$ を式 (6) のようにして求める。

$$\mathbf{x}_{org,t} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \mathbf{x}_{he}(t) \quad (6)$$

この処理を施すことで、変更後の座標系では、全身の上下運動を含む相対運動を観測することができ、人

物の位置によらずに行動素が抽出できる状態に変換される。

なお、時刻 t での身体各部位の三次元位置データ $\mathbf{x}_i(t)$ は式 (7) のように変更される。

$$\mathbf{x}'_i(t) = \mathbf{x}_i(t) - \mathbf{x}_{org,t} \quad (7)$$

3.1.3 体の向きの一

行動素は体の向きに依存せずに抽出されることが求められる。そこで、各時刻における両肩の位置から体の向きを推定し、すべての行動において体の向きを統一する。

実際の処理手順を以下に示す。

処理 1 時刻 t での肩の向き θ_t を両肩の x, y 座標値から推定する。(ただし、 $(x_{rs}(t), y_{rs}(t))$, $(x_{ls}(t), y_{ls}(t))$ はそれぞれ時刻 t での左右の肩の (x, y) 座標値である。)

$$\theta_t = \arctan \frac{y_{ls}(t) - y_{rs}(t)}{x_{ls}(t) - x_{rs}(t)} \quad (8)$$

処理 2 各部位 $\mathbf{x}'_i(t)$ の三次元位置データを z 軸周りに $-\theta_t$ 回転し、向きをそろえる。

$$\mathbf{x}''_i(t) = \begin{pmatrix} \cos \theta_t & \sin \theta_t & 0 \\ -\sin \theta_t & \cos \theta_t & 0 \\ 0 & 0 & 1 \end{pmatrix} \mathbf{x}'_i(t) \quad (9)$$

以降、簡単のため身体各部位の三次元位置データ $\mathbf{x}''_i(t)$ を $\mathbf{x}_i(t)$ とする。

3.2 特徴量抽出

前処理により得られた三次元位置データから、クラスタリングの入力となる各行動素の特徴量を抽出する。

3.2.1 一連行動から部分行動の切り出し

図 5 のように事例データベース内の一連行動のデータから部分行動のデータを切り出しを行う。ここでは、input frame f_i ($f_i = \{\mathbf{x}_i(t) | i = he, rs, \dots, lh\}$) は時刻 t における各部位の前処理済みの三次元位置データ集合を意味している。あらかじめ部分行動の固定フレーム長 N を与えて、クラスタリングの入力単位となる input block F_k ($F_k = \{f_{kN}, f_{kN+1}, \dots, f_{(k+1)N-1}\}$) を抜き出す。一連行動のフレーム長を L とすると、一連行動から生

成される部分行動集合は $F = \{F_k | k = 1, 2, \dots, \lfloor L/N \rfloor\}$ である。

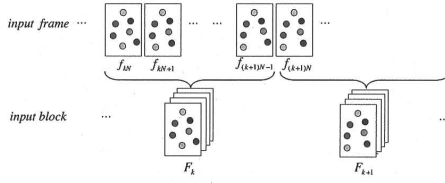


図 5: 部分行動の切り出し

3.2.2 特徴量抽出

これまでの処理によって得られた部分行動の三次元位置データから特徴量を抽出する。特徴量抽出のポイントとして、

ポイント 1 使用する特徴量の数は極力少なくする。

ポイント 2 特徴量同士相関があるものは避ける。

ポイント 3 個人差を吸収できる特徴量を抽出する。

が考えられる [11]。まず、ポイント 1 とポイント 2 を考慮し、図 6 のように特徴量に使用するフレーム数 n ($n \leq N$) をあらかじめ決め、input block から等間隔に抜き出す。次に、ポイント 3 を考慮し、各部位の速度を特徴量として用いる。速度を特徴量として用いる理由は、各部位の三次元位置よりも個人の体格の影響を受けにくいためであり、また動きを表現するパラメータとして適切だと考えられるためである。

ここで部分行動 F_k 内で特徴量として用いられる各部位の速度を式 (10) のように定義する。

$$\dot{\mathbf{x}}_i(j) = \frac{\mathbf{x}_i(kN + \frac{j(N-1)}{n-1}) - \mathbf{x}_i(kN + \frac{(j-1)(N-1)}{n-1})}{\frac{N-1}{n-1}} \cdot 30[\text{fps}] \quad (10)$$

$$j = 1, \dots, n-1$$

さらに、これらを用いて f_j を式 (11) のように定義する。

$$\hat{\mathbf{f}}_j = (\dot{\mathbf{x}}_{he}(j), \dot{\mathbf{x}}_{rs}(j), \dots, \dot{\mathbf{x}}_{lh}(j))^t \quad (11)$$

このとき、部分行動 F_k の特徴量ベクトル X_k は $\hat{\mathbf{f}}_j$ を用いて式 (12) のように表される。

$$X_k = (\hat{\mathbf{f}}_1, \dots, \hat{\mathbf{f}}_{n-1})^t \quad (12)$$

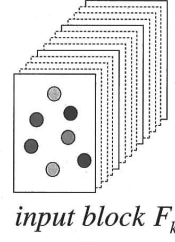


図 6: 特徴量に使用するフレームの抜き出し

3.3 準教師有りクラスタリング

複数の一連行動から得られた部分行動の特徴量を入力として与え、準教師有りクラスタリングによって行動素を抽出する。本研究では、あらかじめ人間がいくつかの入力データに対し、ある二つの部分行動が同一クラスタに存在するのか、別のクラスタに存在するのかという教師情報を与え、その制約を満たすようにクラスタリングを行う COP-kmeans[10] アルゴリズムを適用する。教師情報を入力の一部に与えるため、身体行動の事前知識をクラスタリング結果に反映させることができる点が準教師有りクラスタリングの利点である。

4 行動素抽出実験

京都大学 IMADE ルームのモーションキャプチャシステム PhaseSpace[8] を利用し行動素抽出実験を行った。4.1 節では実験手法について述べ、4.2 節では実験結果及び考察について述べる。

4.1 実験手法

実験では、机周りで起こる日常的な行動を設定し、その行動内から行動素抽出を行った。実験で行った一連行動内に含まれる行動パターンを以下に記す。

P1 湯のみを置いたお盆を持って歩く。

P2 お茶を出す。

P3 机の周りを歩く。

P4 イスに座る。

P5 新聞を読む(めくる)。

P6 イスに座った状態でゴミをゴミ箱に投げ捨てる。

P7 机の上のノートをノート1, ノート2, ノート3の順にまとめ, 指定した位置に置く。

P8 立っている状態から前屈し, ゴミをゴミ箱に入れる。

P9 机の上の雑巾がけを行う。

P10 イスから立ち上がる。

これら 10 個の行動パターンを用いて 3 つの一連行動を設定した。

「秘書行動」入室 → P1 → P2 → P3 → 退室

「事務作業」入室 → P3 → P4 → P5 → P6 → P30 → P3 → 退室

「テーブル掃除」入室 → P3 → P4 → P7 → P10 → P3 → P8 → P3 → P4 → P9 → P10 → P3 → 退室

これらの一連行動を体格, 性別の異なる被験者 A~H, 計 8 名が 4 回ずつ行った。これらの一連行動に対し 3 章の手法を用いて行動素抽出を行った。

4.2 実験結果及び考察

行動素数を 8 個~12 個で抽出したときの, 各行動パターンの行動素の遷移を表 1~表 5 に示す。ただし, m_i は行動素であり, i は行動素に割り当てられたラベルである。また, 行動素数ごとに行動素に割り当てられているラベルは異なる。

実験結果の考察

・行動素数による遷移パターンの変化

行動素数が 8 個, 9 個のときは P3 (歩く) に含まれる行動素と P6 (投げる) に含まれる行動素に同一のものが含まれているが, 行動素数が 10 個以上でそれぞれ別の行動素を含むようになっていく。さらに行動素数 10 個以上で P6 (投げる) に含まれる行動素は他の行動パターンに含まれていないことがわかる。このことから, 行動素数が増えるごとに学習する部分行動 F_k が細かく分類されていることがわかる。しか

表 1: 行動素数 8

行動	行動素遷移
P1	m_5
P2	$m_5m_8m_5$
P3	m_6m_1
P4	$m_8m_2m_3$
P5	$m_5m_7m_5$
P6	$m_5m_1m_5$
P7	$m_5m_8m_5m_7m_5m_7m_5m_7m_5$
P8	$m_5m_8m_2m_5$
P9	$m_5m_7m_5$
P10	$m_8m_4m_5$

表 2: 行動素数 9

行動	行動素遷移
P1	m_5
P2	$m_5m_7m_5$
P3	m_1m_4
P4	$m_7m_3m_8$
P5	$m_5m_9m_5$
P6	$m_5m_4m_5$
P7	$m_5m_8m_5m_9m_5m_9m_5m_9m_5$
P8	$m_5m_7m_3m_7m_5$
P9	$m_5m_9m_5$
P10	$m_7m_6m_5$

し, 行動素数が増えるごとに, 行動素の遷移パターンが複雑になり, 一つの行動パターンに対し複数の遷移パターンが存在する傾向が見られる。よって, 構築する行動認識システムの行動識別レベルに対して, できるだけ少ない行動素数を設定する必要があると考えられる。

・P1 について

各行動素数において P1 に含まれる行動素が他の行動パターンに多く含まれている。P1 はお盆を持って歩く動作であり, そこに含まれる行動素はほぼ上半身が動かない状態を示している。そのため他の行動パターンにおいて上半身の動きが少ない場合は, P1 内に含まれる行動素と同一のものとなっている。下半身の三次元位置データを取得できれば, P1 に他の行

表 3: 行動素数 10

行動	行動素遷移
P1	m_2
P2	$m_2m_6m_2$
P3	m_8m_5 or $m_8m_2m_5$
P4	$m_6m_4m_7$
P5	$m_2m_1m_2$
P6	$m_2m_{10}m_2$
P7	$m_2m_7m_2m_1m_2m_1m_2m_1m_2$
P8	$m_5m_6m_4m_6m_5$
P9	$m_2m_1m_2$
P10	$m_6m_9m_2$ or $m_6m_3m_2$

表 4: 行動素数 11

行動	行動素遷移
P1	m_{10}
P2	$m_{10}m_5m_{10}$
P3	m_6m_1 or $m_6m_{10}m_1$
P4	$m_5m_2m_8$ or $m_5m_9m_2m_8$
P5	$m_{10}m_7m_{10}$
P6	$m_{10}m_4m_{10}$
P7	$m_{10}m_{11}m_{10}m_7m_{10}m_7m_{10}m_7m_{10}$
P8	$m_5m_9m_2m_5m_{10}$
P9	$m_{10}m_7m_{10}$
P10	$m_{10}m_5m_3m_{10}$ or $m_{10}m_5m_3m_8m_{10}$

動パターンと異なる行動素が割当てられると考えられるが、現状で P1 に別の行動素を割り当てるには、座標系の変換前の部位「頭」の xy 平面における移動量を特徴量に含めることが有効だと考えられる。ただし、時刻 t での xy 平面における移動量は 3.1.2 節で述べた $\mathbf{x}_{org,t}$ を用いて、 $\|\mathbf{x}_{org,t} - \mathbf{x}_{org,t-1}\|$ で求められる。

教師情報による影響

各行動素数において P5 と P9 の行動素の遷移パターンが一致している。これは共にイスに座っている状態で右手を左右に動かせるという非常に似通った動作であることが原因であると考えられる。構築する認

表 5: 行動素数 12

行動	行動素遷移
P1	m_4
P2	$m_4m_{11}m_4$
P3	$m_{12}m_5$ or $m_{12}m_4m_5$
P4	$m_{11}m_{10}m_4$ or $m_{11}m_3m_{10}m_4$
P5	$m_4m_1m_4$ or $m_4m_6m_1m_4$
P6	$m_4m_2m_4$
P7	$m_4m_6m_7m_4m_6m_4m_1m_4m_1m_4$
P8	$m_4m_{11}m_3m_4$ or $m_4m_{11}m_3m_{11}m_4$
P9	$m_4m_1m_4$
P10	$m_4m_{11}m_8m_6m_4$ or $m_4m_{11}m_6m_8m_4$

識システムにより識別したい行動のレベルは異なるが、これら二つの行動パターンを別の行動として認識させる場合、それぞれの部分行動が異なるクラスタに入るという教師情報を与えることで対応することができる。P5 と P9 の部分行動に教師情報を与え、行動素数 12 個で準教師有りクラスタリングを行った結果を表 6 に示す。結果として P5 と P9 において異なる遷移パターンが得られた。しかし、P3 と P6 に再び同一の行動素が含まれている。これは、本来同一の行動素が抽出されるべき部分行動に対し異なるクラスタに入るという制約を与えたために、他のクラスタ間の距離関係が崩れたのではないかと考えられる。そこで、さらに行動素数を増やしてクラスタリングを試みたが、P3 と P6 に含まれる行動素は分離されず、行動素の遷移パターンがさらに複雑になり、一つの行動パターンに対し多くの行動素の遷移パターンが存在するという結果となった。これは、行動素数に対してクラスタリングの入力数（部分行動数）が少ないために、個人差、試行差により、部分行動が細かく分類されてしまうためだと考えられる。そこで、行動素数 12 個のまま、P3 と P6 の部分行動に異なるクラスタに入るという制約を与え準教師有りクラスタリングを行った結果を表 7 に示す。結果として、P3 と P6 に含まれる行動素は別のものとなり、P6 に含まれている行動素は他の行動パターンには含まれずに抽出されている。このことから、行動素抽出において、行動素数を増やさなくても、行動パターンごとに異なる行動素の遷移パターンを得ることができるという点で準教師有りクラスタリングは優れているといえる。

表 6: P5 と P9 の区別

行動	行動素遷移
P1	m_{12}
P2	$m_{12}m_6m_{12}$ or $m_{12}m_8m_6m_{12}$
P3	m_3m_9 or $m_3m_{12}m_9$
P4	$m_6m_1m_8m_{12}$
P5	$m_{12}m_{11}m_{12}$
P6	$m_{12}m_9m_{12}$
P7	$m_{12}m_8m_{12}m_4m_{12}m_4m_{12}m_4m_{12}$
P8	$m_{12}m_6m_7m_{12}$ or $m_{12}m_6m_7m_1m_{12}$
P9	$m_{12}m_4m_{12}$
P10	$m_{12}m_6m_{10}m_{12}$ or $m_{12}m_6m_2m_{10}m_{12}$

表 7: P5 と P9, P3 と P6 の区別

行動	行動素遷移
P1	m_3
P2	$m_3m_8m_3$ or $m_3m_{12}m_8m_3$
P3	m_5m_7 or $m_5m_3m_7$
P4	$m_8m_{10}m_1m_3$
P5	$m_3m_4m_3$
P6	$m_3m_6m_3$
P7	$m_3m_1m_3m_{12}m_3m_{12}m_3m_{12}m_3$
P8	$m_{12}m_8m_{10}m_8$ or $m_3m_8m_{10}m_8m_{10}$
P9	$m_3m_{12}m_3$
P10	$m_3m_8m_9m_3$ or $m_3m_8m_3m_9m_3$

5 おわりに

本研究は、動画像中の人物像の行動を認識するシステムの構築を目指している。人物行動の認識に必要な行動モデルの構築を行うために行動素抽出を行った。モーションキャプチャシステムを利用し、一連行動中の人物の身体各部位の三次元位置データを取得し、速度を特徴量としてクラスタリングすることにより行動素を抽出した。クラスタリングの手法として、学習データの一部の行動に似た行動と異なる行動についての事前知識を教師情報として与える準教師有りクラスタリングを適用した。行動素数を増やすことにより行動素の遷移パターンが複雑になり、一つの行動パターンに対し複数の行動素の遷移パターンが存在する傾向があるが、教師情報を適切に使用することで、行動素数を増やさなくても行動パター

ンごとに異なる行動素の遷移パターンが得られるという結果が得られた。行動素が安定して抽出されるようになれば、画像入力に対し行動認識を行うシステムの構築を目指す。

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 (20040312) 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 May 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, (20010111), pp.25-30, 2001.
- [4] 新宅聡子: “部分ジェスチャー抽出の試み”, 特別研究報告, 大阪大学基礎工学部, 2002.
- [5] B.Bauer. and H.Heinz: “Towards an Automatic Sign Language Recognition System Using Subunits,” in *International Gesture Workshop 2001*, I.Wachsmuth and T.sowa(eds.), LNAI 2298, pp.64-75, London, UK, 2002.
- [6] 大和淳司, 大谷淳, 石井健一郎: “隠れマルコフモデルを用いた動画像からの人物の行動認識,” 電子情報通信学会論文誌 D-II Vol. J76-D-II, No.12, pp.2556-2563, December 1993.
- [7] 神脇敏弘: “教師ありクラスタリングと絶対/相対クラスタリング,” 情報論的学習理論ワークショップ (IBIS2006), 2006.
- [8] PhaseSpace active motion capture system, < <http://www.nihonbinary.co.jp/Virtual/PhaseSpace/index.htm> >, 2008 年 2 月現在.
- [9] A. Matsumura, Y. Iwai and M. Yachida: “Tracking People and Actio Recognition from Omnidirectional Images,” in *System and Human Science*, Chapter.37, pp.501-514, Elsevier, 2005.
- [10] K.wagstaff, C. Cardie, S. Rogers, and S. Schroedl: “Constrained K-means Clustering with Background Knowledge,” in *ICML-2001*, pp. 577-584, 2001.
- [11] T.Hastie, R.Tibshirani, J.H.Friedman: “*The Elements of Statistical Learning: Data Mining, Inference, and Prediction*,” Springer Series in Statistics, 2002.