

Frequent Motion Pattern Extraction from Human Motion Data for Real-time Human Proxy

吉松, 寿人
九州大学システム情報科学研究院知能システム学部門

荒木, 裕
九州大学システム情報科学研究院知能システム学部門

有田, 大作
九州大学システム情報科学研究院知能システム学部門

谷口, 倫一郎
九州大学システム情報科学研究院知能システム学部門

<http://hdl.handle.net/2324/5977>

出版情報：画像の認識・理解シンポジウム, pp.1066-1073, 2005-07
バージョン：
権利関係：



リアルタイムヒューマンプロキシのための 動き情報からの頻出動作パターン抽出

吉松 寿人[†] 荒木 裕[†] 有田 大作[†] 谷口倫一郎[†]

[†]九州大学大学院システム情報科学府 〒 816-8580 福岡県春日市春日公園 6-1

E-mail: †{hisato,araki,arita,rin}@limu.is.kyushu-u.ac.jp

あらまし 近年, 仮想空間中のアバタを介して遠隔地のユーザ同士のコミュニケーションを支援する, 仮想環境コミュニケーションに関する研究が盛んに行われており, それを実現するひとつの方法としてリアルタイムヒューマンプロキシ (RHP) が提案されている. RHP によってコンピュータを介した自然なインタラクションを実現するためには, インタラクションに必要な動作を認識し, それを相手に伝達しなければならない. しかし, 人と人のインタラクションにおいて, どのような動作が重要なメッセージを伝達しているのかを判断することは難しい. そこで, 本稿では人体の動作解析を行うことで, 認識すべき動作を決定する手法を提案する. さらに, 人間の動作情報を獲得するシステムを構築し, 実験によってその有効性を示す.

キーワード リアルタイムヒューマンプロキシ, 人体動作解析, 動作認識, モチーフ, 仮想環境コミュニケーション

Frequent Motion Pattern Extraction from Human Motion Data for Real-time Human Proxy

Hisato YOSHIMATSU[†], Yutaka ARAKI[†], Daisaku ARITA[†], and Rin-ichiro TANIGUCHI[†]

[†] Department of Intelligent systems, Kyushu University

6-1 Kasuga-koen, Kasuga, Fukuoka, 816-8580 Japan

E-mail: †{hisato,araki,arita,rin}@limu.is.kyushu-u.ac.jp

Abstract Recently, there are a lot of researches on virtual environments for distant human communication. Real-time Human Proxy (RHP), which is a concept for such a virtual environment, has been proposed. For realizing natural communication by RHP, it is necessary to recognize human actions essential for human communication. However, it is difficult for system developers to decide which human actions should be recognized. For supporting the decision, we propose a human motion analysis method which automatically extracts frequent motion patterns as human action candidates. And, we show some experimental results for extracting human action.

Key words Real-time Human Proxy, human motion analysis, gesture recognition, motif, virtual environment communication

1. はじめに

1.1 研究の背景

バーチャルリアリティ技術の発展を背景として, 臨場感のある仮想的な環境を人工的に作り出し, 人間の代理として投影されたアバタを介してコミュニケーションを行う新たなコミュニケーションメディア (図 1) に注目が集まっている. 仮想環境をコミュニケーションの場として構築するため, 各参加者間の空間的な位置関係を仮想的に作りだすことができ, 動作の対象という情報を直感的に理解することが可能となる. また, 実世界のコミュニケーションと同様に, 参加者の人数が制限されるこ

とはない. したがって, 参加者に関するさまざまな情報をアバタに反映させ, 人間らしく振舞わせることで, face-to-face と同等な質のコミュニケーションが期待できる.

実際に, 仮想環境を用いたコミュニケーションシステムとして, NICE(Narrative Immersive Constructionist/ Collaborative Environments) [1] や, Nessie World [2] といったシステムが開発されている. いずれも, 仮想環境を用いてコミュニケーションを行うシステムであるが, 入力に特別な機器を用いたり, 単純なアバタを用いることによる不自然さといった欠点を持っている. これらは, face-to-face と同等な質のコミュニケーションを目指す際の大きな障害となる.

そこで、これらの問題を解決するために、仮想環境コミュニケーションにおける新たな概念である Real-time Human Proxy (RHP)[3] が提案されている。RHP は、人と人がコンピュータの存在を意識することなく、音声・表情・身振りなどの本来備わっている豊富なコミュニケーションスキルを使った自然なコミュニケーションを実現するための、新しいコミュニケーションメディアの確立を目指したものである。

コンピュータを介した自然なインタラクションを行うためには、人間の動作をリアルタイムで獲得し、必要な動作を認識する必要がある。さらに、認識すべき動作、すなわち、重要なメッセージを伝達している動作を選択しなければならない。この問題を解決するために、本稿では、人体の動作解析によってインタラクションに重要な動作を発見する手法を提案し、動作情報の獲得から認識までの流れを確立することを目指す。



図 1 仮想講義室とアバタ

2. Real-time Human Proxy (RHP)

2.1 RHP

RHP とは、実世界に存在する人間を実時間で仮想化し、その仮想化した情報をネットワークを介して通信することにより、遠隔地であたかも実在の人間がその場に存在するかのように振舞わせるという概念である。人間の仮想化とは、人間に関する一般的な情報（手や足の動き、体格、人間の表情、発せられる音声、着用している衣服など）をシンボルへ変換することをいい、仮想化された情報を再現するとは、変換されたシンボルから、CG アバタという媒介を介して人間の表現をつくり出すことをいう。

2.2 動作情報獲得システム

人間の動き情報はモーションキャプチャシステムを利用して獲得する。モーションキャプチャシステムを利用することにより、人間の体の動きそのものを入力として用いることが可能となり、普段のしぐさやその人の癖、無意識のうちに行う動作なども入力として利用することが可能となる。

コミュニケーションの状況が特定されていたとしても、シンボルとして抽出しなければならない情報の決定は容易なことではない。そこで、効率よくコミュニケーションに必要な情報を選別し、それらの情報からシンボルを生成する枠組みが必要となる。どのような動作を認識すべきかという問題に対しては、

あらかじめコミュニケーションに必要な動作を列挙しておき、それらの動作を認識するというトップダウン的な手法が考えられる。しかし、このトップダウン的な手法では、コミュニケーションに必要な動作を列挙する必要があるが、人間が手作業で行うことを考慮すると、過不足なく全て列挙するのは無理がある。そこで、本研究では、人間が頻繁に行う動作には意味があるという考えのもと、モーションキャプチャシステムから得られた動きの時系列データから、頻出する動作を自動的に見つけ、その後それらがコミュニケーションに必要なかどうかを判定するボトムアップ的なアプローチを試みる。人間の動きデータ中に頻出する動作をモチーフと呼ぶ。また、動作の認識手法に関しては、時系列データのパターン認識に最も一般的に用いられる隠れマルコフモデル (HMM) を用いて動作を認識する。抽出されたモチーフを学習データとして利用することで、学習データ収集の自動化を実現することができる。図 2 に動作情報獲得システムの概要を示す。

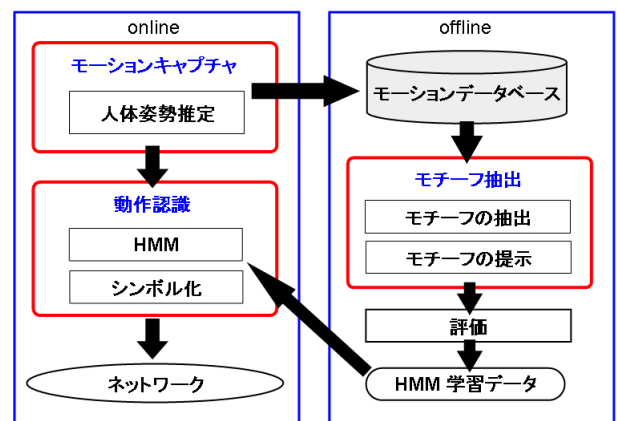


図 2 動作情報獲得システムの流れ

- モーションキャプチャ
実時間で、全身の各部位の 3 次元座標と、頭部の姿勢を推定する。
- モチーフ抽出
モーションキャプチャシステムを駆動させる際、モーションデータをデータベースに保存しておく。保存したモーションデータからユーザが頻繁に行う動作であるモチーフを抽出する。
- 動作認識
HMM を用いて、ユーザの動作認識を行い、シンボルに変換する。HMM の学習データは、モチーフ抽出器で抽出されたデータを用いることにより、自動的に収集される。

3. ビジョンベースモーションキャプチャシステム

本節では、仮想空間への入力装置として利用する実時間モーションキャプチャシステムについて述べる。このシステムを用いることにより、人物の姿勢を実時間で獲得することができる。また、画像情報のみを用いて人物の姿勢を推定するため、マーカーなどの機材を装着することを必要としないシステムである。

3.1 PC クラスタによる実時間多視点動画画像処理

利用するモーションキャプチャシステムでは、画像情報を人物の姿勢推定の入力として用いるため、複数のカメラを使用する。複数のカメラにより獲得された冗長な情報を互いに参照することで、単一のカメラのみによる観測で起こり得る隠蔽などの様々なエラーに対処することが可能となる。さらに、システムの実時間性を確保するために、高速ネットワークにより接続された複数のPCから構成されるPCクラスタを利用する。

3.2 カメラの構成

人物の全身が撮影可能なカメラ8台を、約45度間隔で円周上に配置して利用する。これらのカメラは固定されており、計測空間中に定義したワールド座標系とカメラ座標系間であらかじめキャリブレーションが行われているものとする。

3.3 画像からの特徴点抽出

センサーなどを用いない非接触での人物動作解析には、人体領域に相当する画像特徴として人物の肌色領域や輪郭線などの領域特徴が一般的に利用される。本システムでは、追跡が比較的安定に行える手先や足先の肌色領域と、人体の輪郭線情報の二つの異なる特徴量を同時に利用する。

3.4 人体の3次元姿勢推定

3.4.1 肌色領域の3次元位置推定

各カメラ画像から肌色追跡によって得られた肌色領域重心より、肌色領域の3次元位置を推定する。2次元画像中の1点は撮影中心を通る3次元空間中の直線に対応付けられる。2視点以上のカメラで同一の点が見えている場合、各視点からの拘束線の交点が密な点をとることでその点の3次元位置を推定することができる。

3.4.2 全身の姿勢推定

画像特徴量から推定されるのは頭部・両手・両足の5つの3次元位置である。これらの限られた情報を用いて人体の姿勢を推定するため、ロボティクスやコンピュータグラフィックスの分野で一般に用いられる逆運動学の数値解法を用いる。また、人体モデルの表現として、23の自由度を持つ多関節構造モデルを用いる。

姿勢推定は、下記の2段階の処理によって行う。

1. 逆運動学解法

人体モデルに基づき大まかな人間の姿勢を推定する。推定結果の肩・腰の位置はそのまま利用するが、肘・膝に関しては、次のステップで探索を行って推定する。

2. 肘の探索

肩・腰の位置と手・足先の位置から得られる肘・膝の存在領域と人間のシルエット画像との関係から肘・膝の位置を推定する。

図3に、モーションキャプチャシステムの動作結果を示す。

4. モチーフの抽出

RHPの概念をもとに、コンピュータを介した遠隔地同士のインタラクションを実現するためには、人と人のインタラクションにおいて、どのような動作が重要なメッセージを相手に

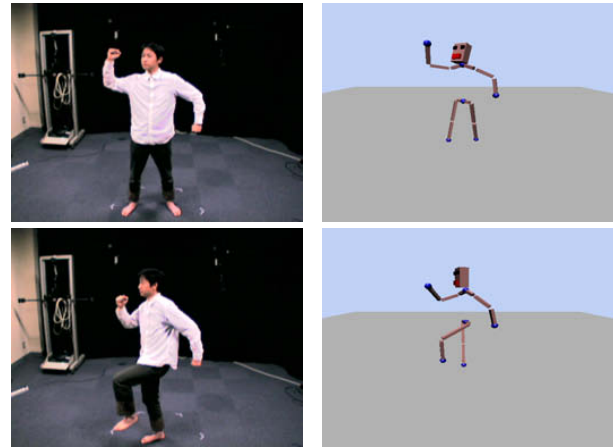


図3 入力画像とモーションキャプチャの結果。左:入力画像。右:CGアバタ

発しているのかという、コミュニケーションに必要な動作を考える必要がある。重要な動作を選出するために、2.2節で述べたボトムアップ的なアプローチを用いる。本節ではボトムアップ的なアプローチの主幹技術となるモチーフの抽出について述べる。モチーフとは動作データ中に頻出するパターン、つまり、与えられた時系列データの特徴を表し、頻出する類似の時間的変移を持つ部分を示すパターンと定義されている[7](図4)。

4.1 動作解析による頻出動作パターン(モチーフ)の抽出

時系列データの動作解析は、まず、時系列のデータを、意味のある最小単位の動作である基本動作に分割する。次に、それらの基本動作をクラスタリングし、基本動作にグループを表すラベルを付加する。これにより、時系列データをクラスタ番号を表す記号列で表現することができる。最後に、求められた記号列から頻出する部分記号列を見つけることによって、最終的に人間の動作データから重要な動きを検出する。この頻出する部分記号列がモチーフとなる。

3節で述べたモーションキャプチャシステムより得られる人間の動きデータは多次元であるため、音声などの一次元特徴量と比較して、基本動作ごとの分割には工夫を要する。さらに、同じ人が同じ内容の動作を行ったとしても、その動作に要する時間は微妙に異なってくる。したがって、クラスタリングにおいても、時間的、空間的な差異をある程度吸収するアルゴリズムが必要である。

次節以降では、まず、分割化がある程度容易に実現できる一次元特徴量からのモチーフ抽出手法について述べ、それをモーションキャプチャシステムから得られた多次元特徴量からのモチーフ抽出へと拡張する。

4.2 一次元特徴量からのモチーフ抽出

一次元時系列データに対するモチーフは、前節で述べたように以下の三段階の処理を経て抽出される。

- (1) セグメンテーション
- (2) クラスタリング
- (3) 頻出パターンの抽出

以下、各処理における詳細を述べる。

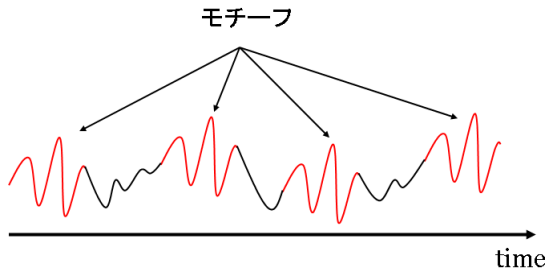


図 4 モチーフ

4.2.1 セグメンテーション: 基本動作への分割

人間は速度の変化から対象物体の動きを認識する。また、体の個々のパーツに注目すると、動作は必ず静止状態から速度を上げ、ある地点で速度が減少し静止するという過程をもっている。この状態の変化を利用して、一次元動作データを基本動作に分割する。つまり、速度が正から負、負から正、正から零、零から正、負から零、零から負に変化する部分を分割点と考える。しかしながら、単純に速度変化が生じる点で分割を行うと、モーションキャプチャシステムの精度を原因とするノイズや、人間の微妙な震えなどにより、時間的に極端に短い、または空間的にあまりに小さな動きをしている動作など、意味のないセグメントに分割されるという問題が生じる。そこで、時間的、空間的な評価により分割点を評価し、意味のない分割点を除去する操作を行う。時間的評価では、隣接する二つの分割点間の時間軸方向の距離が極端に短い部分において、それらの分割点の中点を新たな分割点として追加し、もとの二つの分割点は削除する。また、空間的評価では、隣接する二つの分割点での、特徴量の差が極端に小さい部分において、後の方の分割点を削除する。次に、隣接する三つの分割点候補に着目し、時間的に中間にある分割点での値がその前後の値を対角点とする長方形の中に包括される場合、その中間点を除去する操作を行うことにより、ノイズと思われる分割点を取り除く。図 5 にその様子を示す。まず、始めに ABC の三点に注目する。この三点の中間の点である B は A と C よりも値が大きくなっている。この場合、B 点は正しい分割点だと考える。次に、BCD の三点に注目する。C の高さは B と D の間に存在している。この場合、C 点はノイズとみなし分割点から除去する。以上の処理を最後の分割点候補まで繰り返し行うことにより、より正確な分割点を検出することができる。

図 6 に、モーションキャプチャシステムから得られた右手の y 座標 (y 軸は地面に垂直な向き) の時系列データに対してセグメンテーションを行った結果を示す。(a) は、入力データに対し、速度変化が生じる点において分割を行った結果であり、ノイズ除去を行う前のものである。図からもわかるように、モーションキャプチャシステムから得られる座標データには多くのノイズが含まれているため、手を静止させていたとしても、その座標値は安定せず、微妙な振動を生じる。その結果、多くの余分な分割点が検出されてしまっている。(b) 及び (c) は、(a) の分割点から、時間的、空間的な閾値によって、余分な分割点

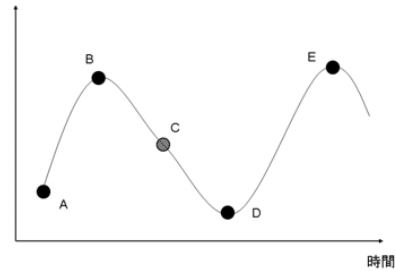
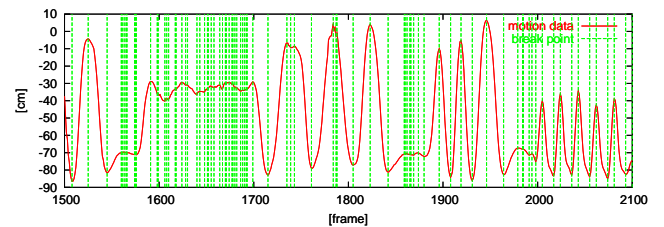
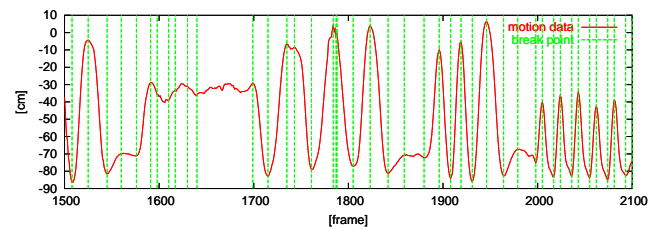


図 5 隣接 3 点間の関係を利用したノイズの除去

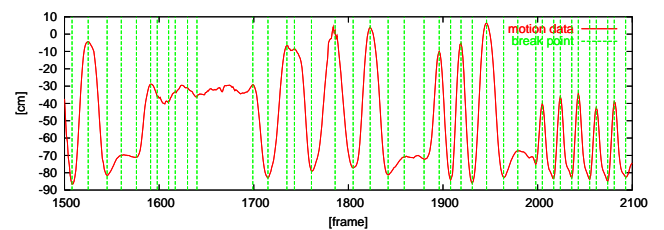
を取り除いた結果を表している。(d) は、(c) に示す分割候補点に対し、隣接 3 点間の関係を利用して、意味がないと考えられる分割点を除去した結果である。図より 1600 フレーム付近や、1750 フレーム付近でのノイズと思われる分割点が除去されていることが分かる。



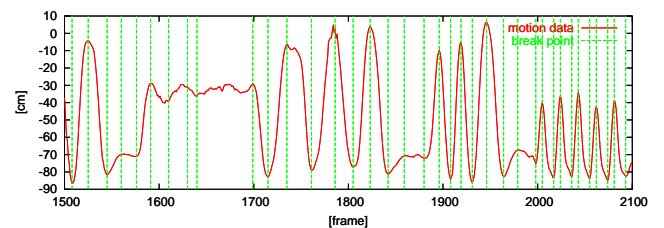
(a) 閾値処理前の分割点



(b) 時間的閾値によるノイズ除去後の分割点



(c) 空間的閾値によるノイズ除去後の分割点



(d) 隣接 3 点間の関係を用いたノイズ除去後の分割点

図 6 セグメンテーションの結果

4.2.2 クラスタリング：基本動作へのラベル付け

時系列動作データを基本動作に分割した後、その基本動作に対しクラスタリングを行い、似たような基本動作に同じ記号を与える。基本動作に記号を割り当てることにより、元の時系列動作データを記号列で表現することができる。

基本動作のクラスタリングは Nearest Neighbor (NN) 法を用いて行う。NN 法を用いたクラスタリングを行うためには、パターン間の距離、すなわち、分割された基本動作間の距離を求める必要がある。しかし、同じ人が同じ内容の動作を行ったとしても、その長さはそのつど変わり、しかも非線形に伸縮する。そこで、音声認識の分野で広く知られているパターンマッチング手法の一つである DP マッチングを用いて基本動作パターン間の距離を計算する。DP マッチングにより、基本動作の時系列長（基本動作パターンの時間的長さ）が異なる場合でも、動的計画法によってパターンを非線形に伸縮させることにより、信頼性の高い類似度を計算することが可能となる。DP マッチングにより導出された 2 パターン間の距離をデータの個数（フレーム数）で割った値を最終的な距離とする。

図 6(d) に示す基本動作に対してクラスタリングを行った結果を図 7 に示す。y 軸の値が同じであるフレームでの基本動作は、同一のクラスタに属していることを表している。

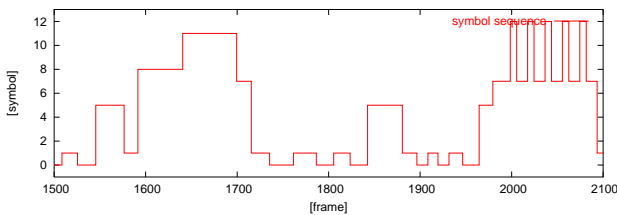


図 7 クラスタリングによって得られた記号列

4.2.3 頻出パターンの抽出

クラスタリングによって、時系列データを記号列として表現することができる。その記号列から、頻出パターンを抽出する。まず、探索窓を適当に決め、記号列から探索窓の大きさの部分記号列をすべて抽出する。その後、それぞれの部分記号列が、もとの記号列中に何回現れるか計算し、頻度の高い部分記号列をモチーフとして決定する。

図 8 に、これまで述べたアルゴリズムを用いて抽出したモチーフ例を示す。各モチーフは、上記の探索窓を 4 に設定して抽出したものであり、一つのデータは 4 つの基本動作から成り立っている。それぞれのモチーフは時間軸方向のずれを持っているが、DP マッチングによってデータの伸縮は吸収され、結果として一つの山と谷を持つ同一の動作として検出されている。

4.3 多次元特徴量からのモチーフ抽出

一次元特徴量からのモチーフ抽出手法を、身体の動作を表す多次元特徴量からのモチーフ抽出へと拡張する。4.2 節で述べた一次元データからのモチーフ抽出アルゴリズムと同様の手法で、全ての次元について、それぞれの次元でクラスタリングまでの処理を行う。記号列は、各次元によって分割点異なるために、そのままでは分析が困難である（図 9(a)）。そこで、分

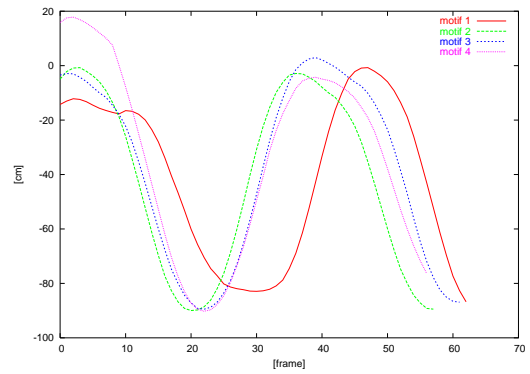


図 8 検出された記号数 4 のモチーフ

析を容易にするために、時間軸をあわせて同期をとることを考える。例として、二つの記号列間の同期をとる場合を説明する。まず、二つの記号列の全分割点において、図 9(b) に示すようにそれぞれの記号列を再分割する。その後、同期をとった二つの記号列から、図 9(c) のように新たな記号列を生成する。以上の処理を最終的に一つの記号列になるまで繰り返し行い、その記号列に対して頻出パターンを検索する。シンボルの同期をとる際、新たな記号列において、あまりにも時間的に短い動作を表す記号を削除することにより、無意味なモチーフの検出をある程度防ぐことができる。

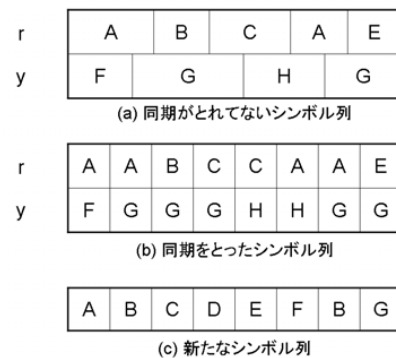


図 9 複数の記号列に対する同期処理

4.4 身体動作を表す特徴量

モーションキャプチャシステムから得られる特徴量は、人物の頭部、両手、両足、両肘、両肩、腰の両側、両膝および胴体の中心の 3 次元座標である。これらの各部位における 3 次元座標は、実世界に固定されたデカルト座標系を基準としており、同じ人物がまったく同じ動きをしたとしても、その人物が向いている方向や立ち位置が違っていれば、当然得られる特徴量も異なってくる。この性質は、動き情報解析によるモチーフ抽出を適用する際、非常に都合が悪い。そこで、モーションキャプチャシステムから得られる特徴量を、人物の向きや立ち位置に依存しない特徴量へと変換する必要がある。さらに、変換後の特徴量は、人物の動きを最もよく表現するものが好ましい。

そこで、上記の条件を満たす座標系として、図 10 に示すような頭部の位置姿勢を基準とした円柱座標系を新たな座標系として定義する。円柱座標系を用いることにより、たとえ人物の

向きや立ち位置が変わったとしても、まったく同じ動きをしていれば、同じ特徴量が得られる。図 10 では円柱座標系で表現した手先の特徴量を表している。ここで、手先の特徴量は以下の式で表される。

$$V = (r, \theta, y) \quad (1)$$

r は、地面から垂直に延ばした、頭部中心を通る直線と、手先までの距離を表しており、 y は、頭部中心と手先の高さの差を表している。また、 θ は人物の視線方向のベクトルと胴体から手へのベクトルとのなす角である。

頭部の特徴量に関しては、前述の円柱座標系で基準として用いていたので、そのまま適用することができない。頭部に関する情報として、モーションキャプチャシステムから、その 3 次元位置だけではなく、姿勢も推定することができる。したがって、頭部に関する特徴量を他の部位とは別に図 11 のように定義する。このときの頭部特徴量は以下の式で表される。

$$V = (\theta_1, \theta_2, \Delta\theta_3) \quad (2)$$

θ_1 は、視線方向ベクトル周りの角度、つまり顔の傾きを表しており、 θ_2 は視線方向ベクトルの仰角、 θ_3 は視線ベクトルの方位角である。 θ_3 に関しては、相対的値 $\Delta\theta_3$ を特徴量としているが、これは、顔の向きによらないモチーフ抽出を実現するためである。

この円柱座標系で表現された体の各部位の特徴量と、頭部の状態を表現する特徴量を用いて、動作解析を行うことにより、より信頼性の高いモチーフが抽出される。

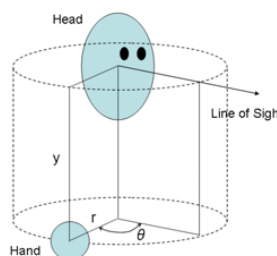


図 10 円柱座標系で表した手の特徴量

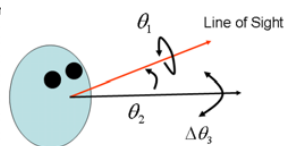


図 11 頭部を表現する特徴量

5. 動作認識

5.1 HMM による動作認識

モーションキャプチャシステムから得られた動作情報をシンボルへ変換するためには、その動作を認識しなければならない。本システムでは、連続型隠れマルコフモデル (Hidden Markov Model: HMM) を動作の確率モデルとして利用し、シンボル化すべき動作を認識する。HMM は確率モデルを用いたマッチング法で、音声認識、動作認識などのアプリケーションにおいて、特に不特定者を対象にした場合に有効な手法である。

5.2 HMM トポロジと学習データ

HMM のトポロジは、飛び越し無しの left-to-right 型を用いる。また、HMM の学習に用いるデータの特徴量は、4.4 節で説明した円柱座標系での特徴量を用いる。モチーフの抽出に用いる特徴量と同じものを用いることにより、モチーフ抽出の結果をそのまま学習データとして利用することができる。

6. 実験と考察

本節では、モチーフ抽出システムの再現性の評価として、意図的に行なった一連の動作から本当に想定したモチーフが検出可能か、実験とその考察について述べる。また、動作認識の評価として、先の実験で得られたモチーフの学習に加え、ルールベースの認識によるシンボル化に関する実験とその考察について述べる。

6.1 モチーフ抽出

モーションキャプチャシステムから得られた時系列モーションデータから、頻繁に行う動作をモチーフとして抽出することができるか確認するための実験を行った。

4 章で述べたモチーフ抽出アルゴリズムを用いて、12fps で獲得した約 40000 フレーム分のモーションデータを入力としてモチーフの抽出を行った。モチーフ抽出の対象部位は上半身に限定し、頭部・両手先・両肘の座標から求めた 15 次元特徴量を入力とした。実験に用いたモーションデータは、以下に示す意図的に行った動作がいくつか含まれている。

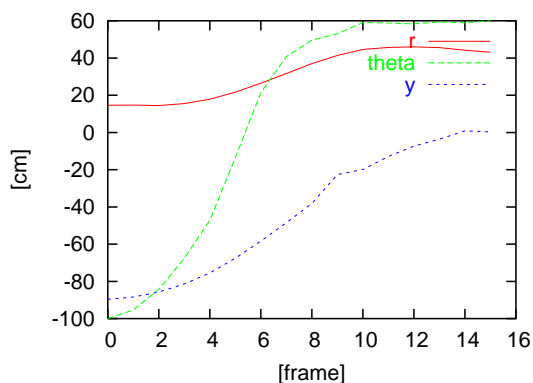
- 右手指差し
- 右手を頭の横で振る
- 左手でノックを行う
- 両手を頭の横で振る
- 右手をぐるぐる回す
- 両手で拍手を行う
- 両手で前ならえを行う
- 体の回りで両腕を回す
- 両手でばんざいをする
- 首を縦に振る
- 首を横に振る

実験により、意図的に行った動作がほぼモチーフとして確認できたが、首を縦に振る動作、横に振る動作は、期待される結果を得ることはできなかった。頭部はあまり動くことがなく、静止状態が長く続くことが多い。頭部が静止しているときも、モーションキャプチャシステムからのノイズにより、推定された頭部姿勢パラメータはわずかに振動している。この振動が、セグメンテーション処理における分割誤りを引き起こしてしまう。また、何回も行ったにも関わらず頻度が少ないモチーフとして抽出された動作も存在した。これは、セグメンテーション処理における分割誤りや、NN 法によるクラスタリングの際、閾値が小さすぎて似たような基本動作が同じクラスタに分類されなかったためだと考えられる。

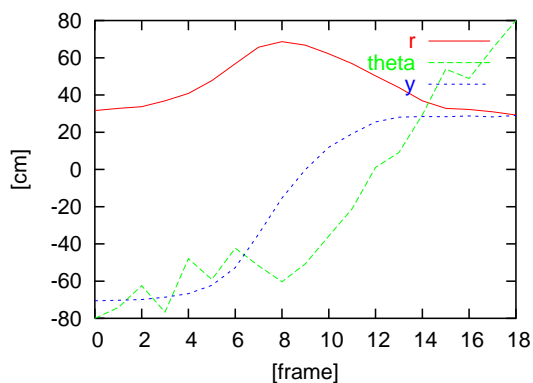
6.2 動作認識実験

実装した動作情報獲得システムを用いて、動作認識の実験を行った。RHP の一つの機能として実装された、動作に沿った動きを行うアバタ [4] を用意し、被験者が行う動作がアバタに正しく反映されるか確認した。動作情報獲得システムでシンボルに変換する動作を以下に示す。

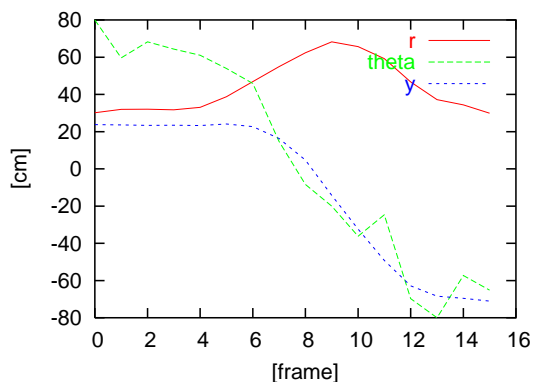
- { 右手 | 左手 } 指差し { 開始 | 変化 | 終了 }



(a) 指差し動作のモチーフ



(b) 拳手開始動作のモチーフ



(c) 拳手終了動作のモチーフ

図 12 HMM の学習データとして用いたモチーフ

- { 右手 | 左手 } 拳手 { 開始 | 終了 }
- 歩行 { 開始 | 変化 | 終了 }
- 頭部回転

これらの動作のうち、指差し開始、拳手開始、拳手終了の3つの動作に関しては、6.1節で抽出されたモチーフを学習データとして学習させたHMMを用いて認識を行った。学習データとして利用したモチーフの例を図12に示す。指差し開始、拳手開始、拳手終了以外の動作に関しては、ルールベースの手法によって認識を行う。例えば、歩行動作に関しては、どちらかの足の速度と胴体中心の速度がある閾値を超えたとき歩行開始とする。

図13に実験の結果を示す。それぞれの画像は、被験者と、アバターが表示されたディスプレイを同時に撮影したものである。実験により、動作が正しく認識されていることが確認できた。

また、モチーフとして抽出された動作の特徴量をHMMの学習データとして用いることで、モチーフと同じ動きを認識できることが確認できた。

7. おわりに

本稿では、仮想環境コミュニケーションのための新たな概念であるRHPについて述べた。また、コンピュータを介したコミュニケーションにおいて、どのような動作を認識し相手に伝えるべきであるかという問題に対し、人間の動作を解析し、頻繁に行う動作をモチーフとして抽出することにより、必要な動作の選択に一つの指標を与えることを提案した。また、人物の動作情報を獲得するシステムを実装し、その有効性を示した。

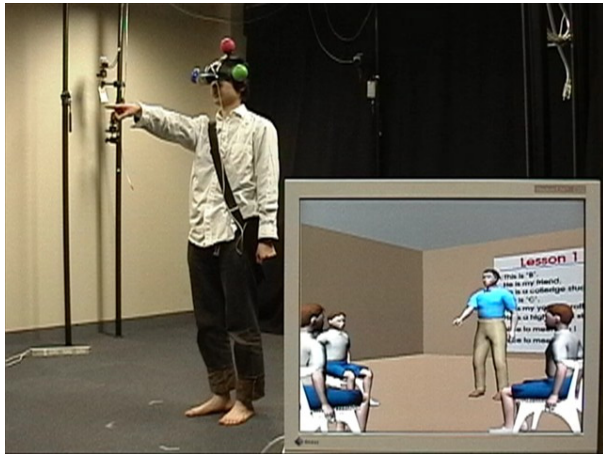
今後は、モチーフ抽出の精度向上を目指し、より多くの動作が必要な状況での行動から、どのようなモチーフが検出されるか、検証していく必要がある。また、それにより検出されたモチーフがコミュニケーションに必要な動作かどうかの評価方法や判断基準を定義していく必要がある。

謝 辞

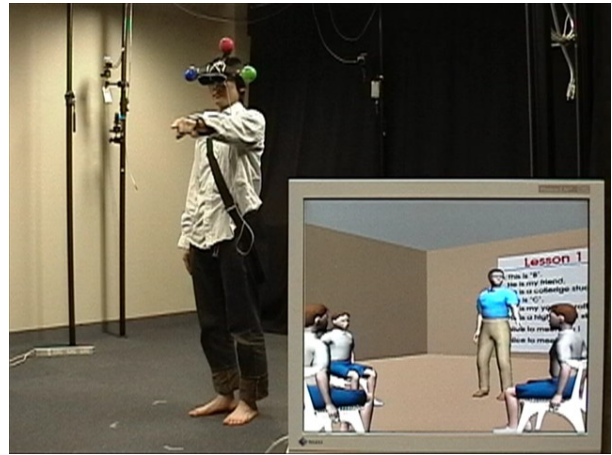
本研究の一部は、科学研究費補助金学術創成研究(2)「人間同士の自然なコミュニケーションを支援する知能メディア技術(13GS0003)」、および、科学研究費補助金若手研究(B)「リアルタイムヒューマンプロキシを用いたアバター遠隔地コミュニケーションの研究(16700108)」の補助を受けた。

文 献

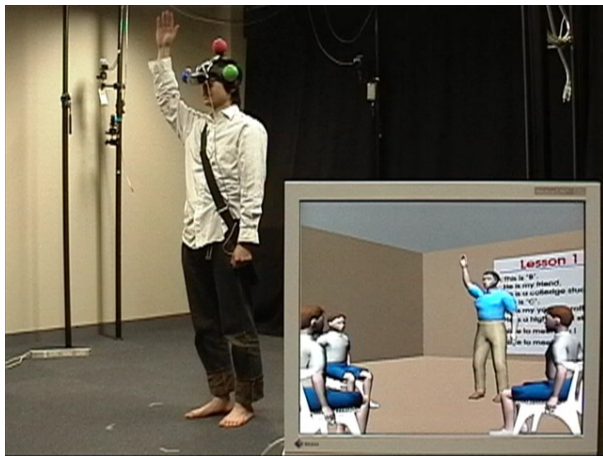
- [1] M. Roussos, A.E. Johnson, J. Leigh, C.A. Vasilakis, C.R. Barnes, T.G. Moher, "NICE: combining constructionism, narrative and collaboration in a virtual learning environment," ACM SIGGRAPH Computer Graphics, Volum 31, Issue 3, pp.62-63, 1997.
- [2] P. Jeffrey, A. McGrath, "Sharing serendipity in the workplace," Proc. of the third international conference on Collaborative virtual environments, pp.173-179, 2000.
- [3] D. Arita, H. Yoshimatsu, D. Hayama, M. Kunita, R. Taniguchi, "Real-time Human Proxy: An Avatar-based Interaction System," CD-ROM Proc. of International Conference on Multimedia and Expo, 2004.
- [4] D. Hayama, H. Yoshimatsu, H. Yoshimoto, D. Arita, R. Taniguchi, "Avatar generation for Real-time Human Proxy," Proc. of 10th International Conference on Virtual Systems and Multimedia, pp.386-395, 2004.
- [5] <http://www.cs.columbia.edu/hgs/rtp>
- [6] 岩澤, 海老原, 竹松, 坂口, 大谷, "「Shall We Dance?」の構築," 信学技報 PRMU98-114, pp.15-22, 1998.
- [7] J. Lin, E. Keogh, S. Lonardi, P. Patel, "Finding Motifs in Time Series," Proc. of the 2nd Workshop on Temporal Data Mining, pp.53-68, 2002.
- [8] 上原, 田中, 巻尾, "動作データからのモチーフの発見について," 第18回人工知能学会全国大会, 3D3-01, 2004.
- [9] T. Mori, K. Uehara, "Extraction of Primitive Motion and Discovery of Association Rules from Human Motion," Proc. of 10th IEEE International Workshop on Robot and Human Communication, pp.200-206, 2001.



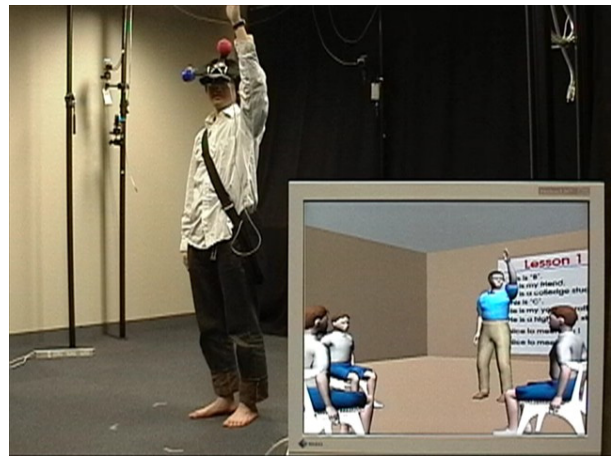
(a) 右手指差し



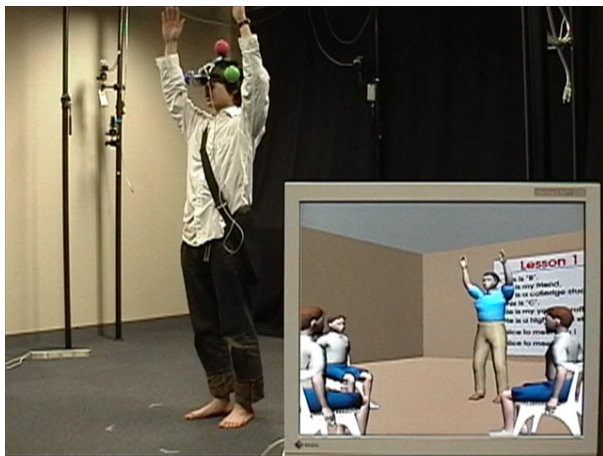
(b) 左手指差し



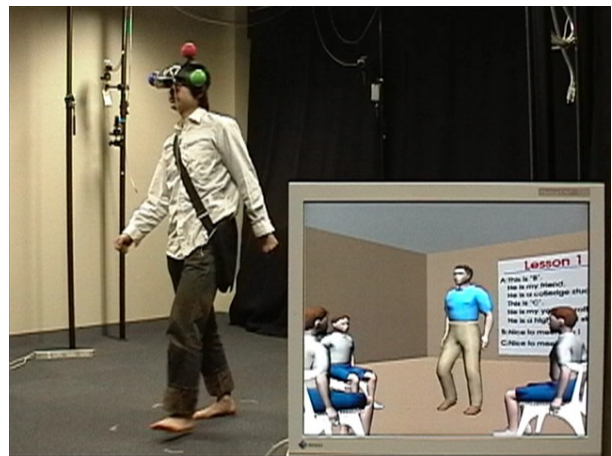
(c) 右手拳手



(d) 左手拳手



(e) 両手拳手



(f) 歩行

図 13 動作情報のシンボル化