Cell Broadband Engine™ を用いたマーカレス モーションキャプチャ

Markerless Motion Capture Using Cell Broadband Engine™

岡田　隆三　　近藤　伸宏
■ OKADA Ryuzo　　■ KONDOH Nobuhiro

モーションキャプチャは，人の動作をコンピュータに取り込む技術であり，コンピュータグラフィックス(CG)コンテンツ制作，動作認識による監視，コンピュータとのインタフェースなど，幅広い応用が考えられている。従来は，体の各部位にマーカやセンサを取り付ける必要があり，気軽に使用することはできなかった。

東芝が開発したマーカレス モーションキャプチャ技術は，1台のビデオカメラを用いて，マーカやセンサなどを体に取り付けなくても人の姿勢を取得することができる。また，Cell Broadband Engine™(注1)を使用することにより，リアルタイムに動作を取得することが可能になった。

Motion capture is a technique for capturing human motion and inputting the data into a computer. It has a variety of applications, including the production of computer graphic (CG) contents, surveillance by monitoring people's motions, and gesture input for computers. Conventional motion capture systems require markers and sensors attached to key parts of the body, which creates a certain amount of difficulty.

Toshiba has developed a markerless motion capture system that enables a single camera to capture human motion with no markers or sensors. This system utilizes the high computational performance of the Cell Broadband Engine™ to capture human motion in real time.

1　まえがき

モーションキャプチャには，CGコンテンツ制作，動作認識による監視，コンピュータやゲームのインタフェースなど幅広い応用があり，近年盛んに研究が行われている。

実用化されているモーションキャプチャ システムには，次のようなものがある。

(1)　光学式モーションキャプチャ　　体の様々な位置に目印(マーカ)を付けて，それを多数のカメラで撮影して各関節の角度を推定する。もっとも普及しており精度も高いが，装置が高価で，マーカの着用や専用のスタジオを必要とするなど制約が多い。

(2)　機械式，磁気式モーションキャプチャ　　体の各部位にポテンショメータやジャイロを装着して，関節角度を推定する。この場合も装置が高価で，ユーザーへの負担が大きい。

(3)　画像に基づくモーションキャプチャ　　初期及び終了姿勢を手動で指定し，その間の動きを画像から推定する。ソフトウェアで実現できるため安価であるが，姿勢を手動で与える必要がある。

これらは，CGクリエーターなど専門知識を持つプロフェッショナル向けで，一般のユーザーが手軽に使えるものではない。監視などマーカやセンサを体に取り付けることができない場合や，インタフェースへの応用などマーカが使用上障害となる場合は，画像による自動モーションキャプチャシステムが有力である。

東芝は，一般ユーザーがコンピュータを容易に操作することを目指して，1台のビデオの映像だけを用いて，マーカやセンサを取り付けることなくユーザーの動きを取得する，モーションキャプチャ システムを開発した[1]。画像から人の姿勢を効率的に認識する画像処理ソフトウェア技術と，Cell Broadband Engine™

図1. 開発したマーカレス モーションキャプチャ システム — ユーザーが右手を上げて立っている姿勢が認識され，ディスプレイ内に同じ姿勢のCGキャラクターが描画されている。
Real-time markerless motion capture system

(注1)　Cell Broadband Engineは，(株)ソニー・コンピュータエンタテインメントの商標。

(以下, Cell/B.E.と略記)の強力な演算能力により, リアルタイムにモーションキャプチャを行うことができる(**図1**)。ここでは, これらの画像認識技術とCell/B.E.への実装, 及び2006年10月に幕張メッセで開催されたCEATEC JAPAN 2006で展示したシステムについて概要を述べる。なお, このシステムは, 現在東芝科学館において常設展示されている。

2　画像による人の姿勢認識

マーカレス モーションキャプチャ システムの処理の概要を**図2**に示す。ビデオカメラで撮影されたユーザーの画像から, 背景差分(注2)によってシルエットを抽出する。システム内には, 3次元人体モデルと多数の姿勢サンプルから生成した, 様々な姿勢のシルエットで構成される姿勢辞書が記憶されている。この姿勢辞書内の多数のシルエットの中で, 画像から抽出されたユーザーのシルエットにもっとも近いものを, シルエット間の類似度(距離)に基づいて探索する。探索は, 様々な姿勢, 画像内の様々な位置, 大きさに対して行う。姿勢辞書内のシルエットは, 人体の全関節の角度を要素とする姿勢ベクトルと関連付けられており, 探索されたシルエットからユーザーの姿勢を得ることができる。

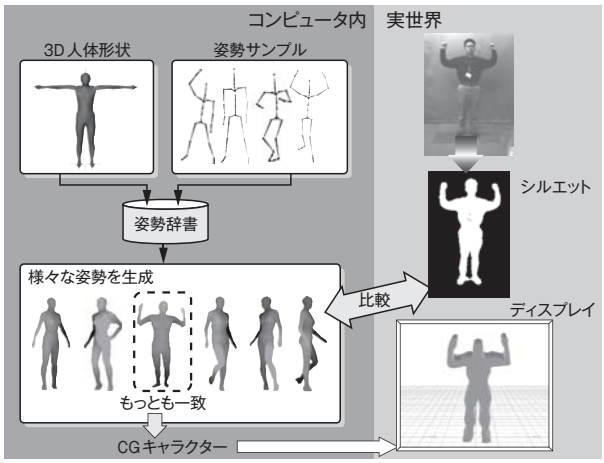

図2. 画像処理の概要 — コンピュータ内の姿勢辞書から生成した多数の姿勢候補の中で, 画像から抽出したシルエットにもっとも近い姿勢を探索することにより姿勢を認識する。認識した姿勢により, CGキャラクターを動作させる。

Overview of image processing

人体は, 非常に高い自由度を持つ多関節物体であるため, 様々な姿勢を認識するためには, 胴体や手足の関節角度が少しずつ異なる膨大な数の姿勢サンプルを, 姿勢辞書内に記憶しておく必要がある。二つの画像間の類似度を表すシルエット間の距離は計算コストが高く, 姿勢探索において, 姿勢辞書内のすべての姿勢について, リアルタイムでシルエット距離の計算を行うことは現実的ではない。そこで, あらかじめシルエット距離に基づいて構成された姿勢の木構造(**図3**)を用いて粗密探索を行うことで, 大幅に計算量を削減する。そのため, 姿勢辞書には, このようなシルエットから構成される姿勢の木構造も格納される。

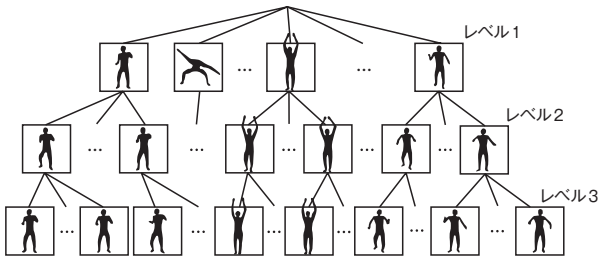

図3. 姿勢の木構造の構成例 — 木構造を用いた粗密探索で, 効率的な姿勢探索を行う。

Example of tree structure of poses

複数の大きく異なる姿勢に対して, ほぼ同じシルエットが得られる場合がある。例えば, 図2は前面から撮影された画像であるが, 同じ姿勢を背後から撮影した場合にもほぼ同じシルエットとなる。また, カメラの光軸方向の腕位置, すなわち腕と体の前後関係が多少変化しても, シルエットはほとんど変化しない。このようなシルエットのあいまい性のため, シルエットからだけでは姿勢推定を行うことは難しい。一つのカメラからの映像だけを用いてこの問題を解決するには, 姿勢の時系列情報を利用することが有効である。つまり, 過去の姿勢推定結果を用いて, あらかじめ定義しておいた人の運動モデルによって予測できる現在の姿勢と矛盾のない姿勢を推定結果として選択することにより, 妥当な姿勢を推定する。

このように, 粗密探索による効率的な姿勢探索と時系列情報を利用した姿勢推定手法をTree-based filteringと呼ぶ。

2.1　姿勢辞書

システム内に記憶されている姿勢辞書は, 3次元人体モデルと, 多数の姿勢サンプルから生成した様々な姿勢のシルエットで構成されている。シルエットは, ユーザーの体格によっても大きく変化するので, 様々な体格の3次元人体モデルを用いて, それぞれの体格についての姿勢辞書を用意する。使用した3次元人体モデルは, おとなに関しては洋服のJIS基準(JIS L4004及びL4005)に基づいて男性10体型, 女性14体型, 子どもについては身長3段階, 体重2段階の6体型(**図4**)である。これらの体型それぞれについて姿勢辞書が生成され, モーションキャプチャを行う際に, どの体型を使用するかを決定して, 対応する辞書をシステムに読み込む。

姿勢辞書内の多数の姿勢サンプルは, 市販のジャイロ式モー

(注2)　あらかじめ学習しておいた背景画像と現在の画像の差分を計算することにより, 変化部分を検出する処理。

一般論文

身長(cm) ＼ 体型	細	太
140	E	F
130	C	D
115	A	B

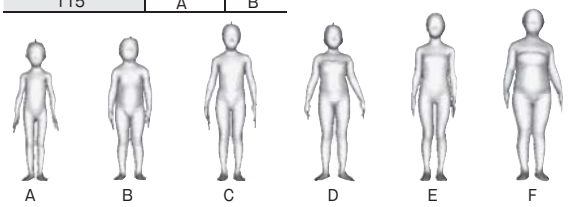

図4. 子供の3次元人体モデル — 身長3段階，体重2段階の合計6種類の人体モデルを使用する。

3D body models for children

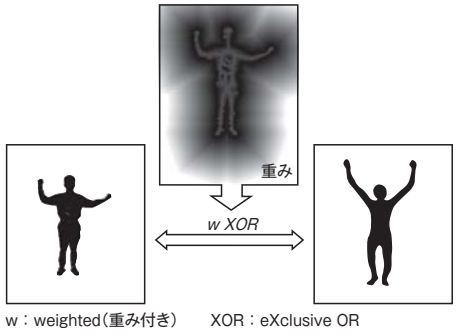

図5. シルエット距離 — 二つのシルエットを画素ごとの排他的論理和によって比較する。このとき，各画素に輪郭から遠いほど大きくなる重みを付けることにより，体型の違いや服装の違いに影響されにくいシルエットの比較を行うことができる。

Silhouette distance by weighted exclusive OR (wXOR)

ションキャプチャ システム(注3)を使用して収集した。取得した姿勢は延べ100万を超えるが，類似した姿勢を削除し，約5万を姿勢サンプルとして図3のような木構造を生成する。

姿勢の木構造は，次の節で述べるシルエット間の距離に基づいて姿勢サンプルを階層的にクラスタリングし，各姿勢サンプル集合を木構造のノード(節点)とすることにより生成する。第1階層では，異なるノードに含まれる姿勢サンプル間のシルエット距離があるしきい値より大きくなるように，すべての姿勢サンプルを分類する。第2階層では，第1階層の各ノードに含まれる姿勢サンプルについて小さいしきい値で同様の処理を行い，更に細かく分類を行う。このように，姿勢サンプルは，木構造の下層ほど細かく分類される。

姿勢認識時には，粗く分割された第1階層で，現在の画像から得られたシルエットと類似しているノードを複数選択し，それらのノードに対してだけ下階層のより細かい探索を行う。このような枝切りにより，姿勢認識の計算量を大幅に削減することができる。

2.2 シルエット距離

このシステムでは，姿勢辞書内に記憶されているシルエットと，現在の画像から得られたシルエットの比較によって姿勢を認識するため，シルエット間の距離の定義は重要である。一方で，シルエット距離の計算は，姿勢辞書内の様々な姿勢について，様々な大きさ，位置で計算されるため，もっとも計算コストの高い処理である。リアルタイムで姿勢認識を行うためには，計算量の少ないシルエット距離を定義する必要がある。

シルエット画像の画素ごとの排他的論理和(XOR)(注4)は，計算量がもっとも少ないシルエット距離の一つであるが，衣服や体型の違いによるシルエット輪郭付近の変化に影響を受けやすく，姿勢認識の安定性に問題がある。また，腕などの細い部位は，画像上では面積が小さくなるため，このような部位の姿勢推定の安定性にも問題がある。そこで，シルエットの輪郭から遠いほど大きく中心線で均等になるような重みを定義し，重み付きの排他的論理和を計算することにより，姿勢認識の安定性を向上させる(**図5**)。

2.3 時系列情報の利用

ほぼ同じシルエットを持つ複数の異なる姿勢の中から，妥当な姿勢を選択するため，時系列情報を用いる。あらかじめ人体の運動モデルを定義し，この運動モデルにもっとも一致する運動をしている姿勢を現在の姿勢として選択する。このシステムでは，前フレームの姿勢推定の結果を，そのまま現在の姿勢の予測値とする運動モデルとして使用する。この運動モデルは，単純だが様々な運動に対応することができる。

3 リアルタイム モーションキャプチャ システム

2章で述べた画像処理による姿勢認識手法をリアルタイムで動作させるため，Cell/B.E.を搭載したCellリファレンスセット(注5)にこの手法を実装した。CellリファレンスセットのPCI(Peripheral Component Interconnect)バスにIEEE1394(米国電気電子技術者協会規格1394) インタフェース ボードを追加し，IEEE1394カメラを1台接続して30フレーム毎秒の速度で640×480画素のカラー画像を取得する(注6)。この画像をCell/B.E.でフレーム処理してユーザーの姿勢を認識する。

Cell/B.E.は3.2 GHzで動作し，管理用の一つのPower Processor Element(PPE)と高い演算性能を持つ七つのSynergistic Processor Element(SPE)からなるマルチコア プロセッサ

(注3) 体の各部位にジャイロセンサを取り付け，特定の初期姿勢からの角度差を計測することにより，各関節の角度を取得する。

(注4) 理論演算の一つで，二つのシルエット画素を重ねたとき，シルエットが重なっている画素では0，重なっていない画素では1となる演算。

(注5) CELLリファレンスセットとは，当社製作のCELL BEを用いた試作コンピュータシステムの名称。

(注6) IEEE1394カメラを接続して動作させるために，Cellリファレンスセットにドライバ，ライブラリなどを独自に追加している。

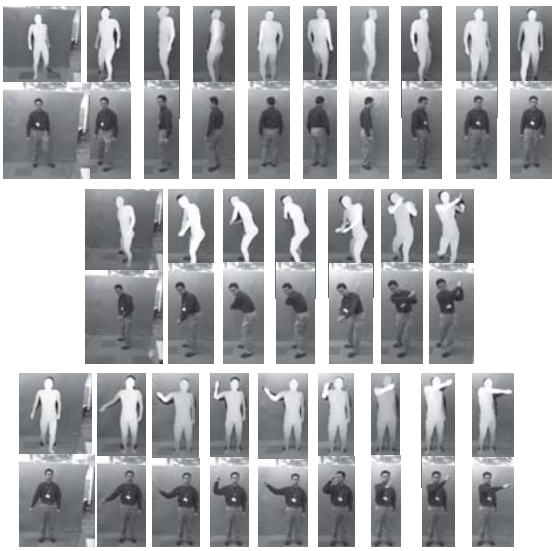

図6. 動作の認識例 — 取得した画像をフレーム処理し，ユーザーのターン，ゴルフのスイング，指差し動作を認識している例である。

Examples of pose recognition

である。背景差分によるシルエット抽出及びシルエット距離計算の各処理を七つのSPEで並列に行うことにより高速化を図り，1フレーム当たりの平均処理時間は86 msを達成している。ただし，処理がフレーム間隔(1/30 s)で終了しない場合には，フレームを飛ばして処理を行う。様々な動作を認識した結果を**図6**に示す。

4　コンピュータゲーム アプリケーション

Cell/B.E.を用いたリアルタイム モーションキャプチャ システムのアプリケーションとして，コンピュータゲームを試作した。**図7**に示すように，カメラの前で動作するプレーヤーの姿勢をCellリファレンスセットによってリアルタイム認識し，その結果をネットワーク経由でゲーム駆動用パソコン(PC)に転送する。ゲーム駆動用PCでは，敵キャラクターとプレーヤーキャラクターの駆動と描画を行う。プレーヤーキャラクターは，プレーヤーの姿勢認識結果に従って動作するので，プレーヤーと同じ動きをする。また，ゲーム駆動用PCは，姿勢の変化を解析してプレーヤーの特定の動作を認識し，敵への攻撃などのイベントを駆動してゲームを進行する(**図8**)。

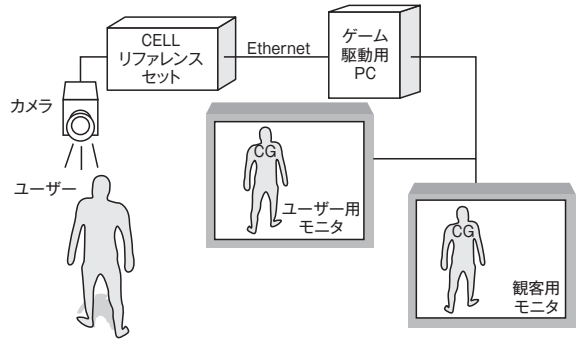

図7. コンピュータゲーム アプリケーションの構成 — 姿勢認識結果をネットワーク経由でゲーム駆動用PCに転送し，ゲームを行う。

Configuration of computer game application with motion capture system

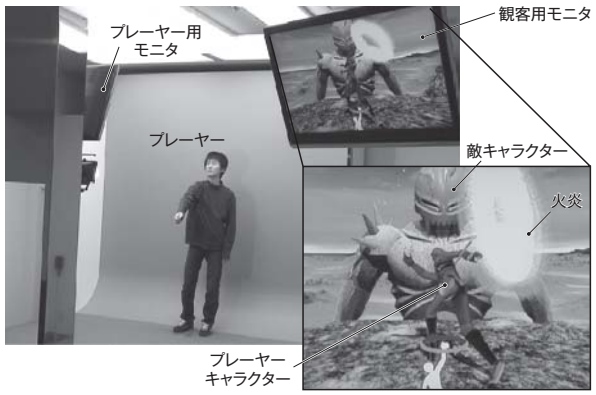

図8. ゲームの実行例 — 右腕を上から振り下ろす動作により，火炎を投げて敵キャラクターを攻撃している。プレーヤーの動作がプレーヤーキャラクターの動きに直ちに反映され，コントローラなどの機材なしで直接プレーヤーキャラクターを操作することができる。

Example of computer game with the motion capture system for controlling player's character

5　あとがき

画像処理による人の姿勢認識技術と，Cell/B.E.の高い演算性能により，リアルタイム マーカレス モーションキャプチャ システムを実現した。今後は，システムの動作条件を緩和するためのいっそうの技術開発と，応用分野の開拓を行っていく。

文　献

(1)　岡田隆三，ほか．“シルエットを用いたTree-Based Filteringによる人体の姿勢推定”．画像の認識・理解シンポジウム予稿集．仙台，2006－07，電子情報通信学会 情報システムソサイエティ パターン認識・メディア理解研究専門委員会(PRMU)．p.63－69.

岡田　隆三　OKADA Ryuzo, ph. D.

研究開発センター マルチメディアラボラトリ研究主務，工博。画像処理の研究開発に従事。電子情報通信学会会員。

Multimedia Lab.

近藤　伸宏　KONDOH Nobuhiro

セミコンダクター社 システムLSI事業部 ブロードバンドシステムLSI応用技術部主務。

画像処理アプリケーション開発業務に従事。

System LSI Div.

一般論文