

モーションデータを用いた人体動作の意味理解

Identification of Human Motion using Motion Capture System

西尾 孝治†
Koji Nishio

中村 徳裕†
Norihiro Nakamura

手島 裕詞‡
Yuji Teshima

金谷 孝之††
Takayuki Kanaya

小堀 研一†
Ken-ichi Kobori

1. まえがき

近年、コンピュータグラフィックスは技術の向上によってより写実的な表現が可能になりつつある。これにともなう、映画などの映像産業でも実写と合成して利用されるようになってきている。また、映像の全編をコンピュータグラフィックスで制作した例も多く見られる。これらの映像には人物などのキャラクターが登場するが、その動きの生成技術が注目されている。

従来は、ソフトウェアを用いてオペレータが動作を入力していたが、非常に手間がかかるという問題があった。そこで、現在ではより手軽に人体動作の入力ができるモーションキャプチャシステムが利用されるようになってきている。モーションキャプチャシステムは、人体の骨格を表すモデルを定義し、センサーで取得した人体の関節位置とモデルを対応付けることで、人体動作を取得するシステムである。

また、モーションキャプチャシステムを利用した無形文化財のデジタルアーカイブに関する研究も注目されている。

しかし、モーションキャプチャシステムは、比較的大掛かりな装置であるため導入時の初期コストが高く、一般に普及するにはいたっていない。このため、すでに取得済みのモーションデータを利用して所望の新しい動作を生成する研究が盛んに行われている。

このように、今後モーションキャプチャ装置はさらに広い分野で利用され、扱われるモーションデータの量も膨大になることが予想される。

このようななか、一般に市販されているビデオカメラを使用して、動画像からモーションデータを取得するシステム [1,2] も登場しており、導入時の初期コストが低くなりつつある。

一般的なモーションキャプチャ装置では、記録対象である人体の表面にマーカと呼ばれる目印を装着する必要がある。しかし、このシステムでは、マーカを設ける必要がなく、より手軽に人体の動作の記録を行うことができるようになってきている。このようなシステムを使用して人体動作の取得ができれば、不特定多数の人体動作の自動取得が可能になる。それらのデータから取得した動作の分類と意味理解ができれば防犯システムへの応用も期待することができる。

そこで、本稿ではモーションキャプチャシステムによって取得したモーションデータの意味理解について検討を行ったので、その手法の提案と結果を報告する。

2. モーションデータ

†大阪工業大学

‡静岡理科大学

††広島国際大学

本手法では、図1に示す17個の関節から構成される骨格をもつモーションデータを使用した。また、ファイル形式には、BVHを用いた。BVHファイルには、次の2つの情報が記述されている。

(1) 階層構造

腰部をルートノードとして、そこに接続する関節を木構造で表現したものを記述する。同時に、各関節間の骨格の長さも記述する。

(2) モーション

各フレームにおける、腰部の位置と向き、および各関節の回転角を記述する。

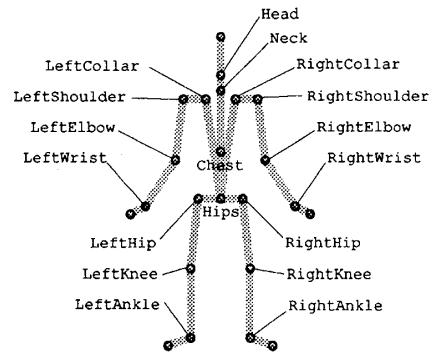


図1

3. 人体動作の意味理解

本手法の処理の流れを図2に示す。

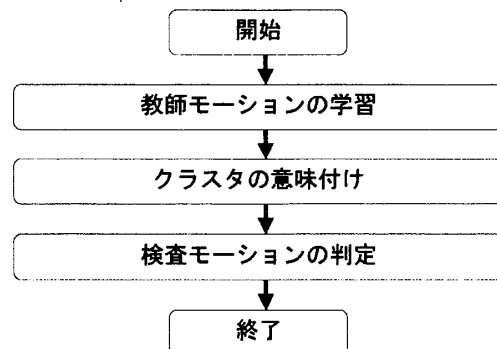


図2

はじめに、複数のモーションデータを教師として、SOMの学習を行う。これにより、動作の種類をあらわすクラスタが形成される。

次に、SOMに形成されているクラスタがもつモーションデータと判別したい基本的なモーションデータを比較して、クラスタに動作の意味付けを行う。

最後に、ノードに意味付けされた SOM を用いて、検査モーションから動作の意味理解を行う。

3.1 教師モーションの学習

動作の意味理解には SOM を利用する。SOM の学習には、N フレーム分の各関節の回転角を用いる。ただし、ワールド座標系に対する腰部の位置と向きは学習の対象としない。このため、SOM の各ノードには、3 軸×17 関節×N フレーム=51×N 次元のベクトルを学習させることとした。

SOM には、意味理解の対象となる基本動作に加えて複数のモーションデータを学習させる。なお、これらのモーションデータも 51×N 次元のベクトルで表し、教師ベクトルと呼ぶことにする。

教師モーションデータの学習は、次の手順で行う。

- (1) SOM の各ノードに格納されているベクトルと教師ベクトルの内積を求め、その値が最も小さくなるノードを探索し勝者ノードとする。
- (2) 勝者ノードを中心として、教師ベクトルを勝者ノードに近いノードほど強く、離れるほど弱く学習させる。

これにより、SOM 上で近いノードは似たベクトルを持つことになり、クラスタが形成される。

3.2 クラスタの意味付け

学習を終えた SOM にはクラスタが形成されているが、このままでは、各ノードに動作を表すベクトルが保存されているだけで、自然言語で表現された動作の意味そのものは保存されていない。そこで、判定結果として出力したい基本動作を用意し、次の手順で各ノードに自然言語による意味付けを行う。ここで、基本動作もベクトルで表すものとし、基本動作ベクトルと呼ぶことにする。

- (1) 基本動作を目視で確認し、ユーザが自然言語による意味付けを行う。
- (2) SOM のノードに格納されているベクトルと各基本動作ベクトルの内積を求め、その値が最も小さくなる基本動作を選び、対応する自然言語で表された意味をノードに格納する。

以上の処理を SOM の各ノードに対して行うことで、SOM に形成されたクラスタに意味付けを行う。

3.3 検査モーションの判定

意味付けされた SOM を使用してモーションデータの意味理解を行う。ここで、意味理解の対象となるモーションデータを検査モーションと呼ぶことにする。この処理では、フレーム単位で自然言語による動作の判定を行う。検査モーションの t フレーム目の意味理解は、次の手順で行う。なお、検査モーションの t フレーム目から t+N フレーム目までのモーションデータをベクトルで表現したものを検査ベクトルと呼ぶことにする。

- (1) SOM の各ノードに格納されているベクトルと検査ベクトルの内積を求め、その値が最も小さくなるノードを探索し類似ノードとする。
- (2) 類似ノードに格納されている意味を参照し、検査モーションの t フレーム目の意味付けを行う。

4. 実験

本手法の有効性を検証するために、モーションデータの意味を求める実験を行った。実験には、5 つの基本動作を含む、* つの動作を使用した。動作の学習、および判定には、縦横 20×20 のノードをもつ SOM を使用した。また、教師モーションデータのフレーム数 N を 20 フレームとし、40 回の学習を行った。

まず、SOM の学習状況を図 3 に示す。図中の文字は対応する基本動作を表し、実際には自然言語による動作の名称(意味)が保存されている。

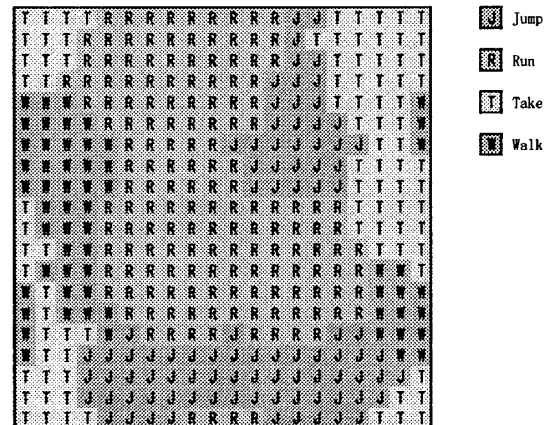


図 3

この結果を用いて、基本動作と似た動作を検査モーションとして 5 フレームおきに意味理解を行った。この結果を表 1 に示す。

表 1 検査モーションの意味理解

	フレーム番号								
	0	5	10	15	20	25	30	35	
検査モーションの種類	Jump	J	J	J	J	J	J	J	J
	Run	R	R	R	R	R	R	R	R
	Take	T	T	T	T	T	T	T	T
	Walk	W	W	W	W	W	W	W	W

表 1 からわかるように、検査モーションデータに対応する基本動作の意味が、ほぼ間違いなく対応付けられていることがわかる。

しかし、より多くの種類の動作の判別を行うには、さらに多くの教師モーションデータが必要になると考えられる。また、同じ種類の動作でも速さが変化すると正しい意味理解が得られないといった問題がある。

周期動作の周期を検出することや、特徴的なフレームを検出することで、意味単位で動作を切り出そうと考えている。これにより、切り出された動作のフレーム数を正規化することで、異なる速さの動作についても正しい意味理解ができるようになると考えている。

文 献

- [1] <http://www.gsport.co.jp/products/krops.html>
- [2] <http://www.webtech.co.jp/pvs/index.htm>