

# 画像に基づいた身体骨格データの推定による ダンス初心者の習得支援

相川 瑞華<sup>†</sup>

<sup>†</sup> 東京農工大工学府産業技術専攻

堀田政二<sup>††</sup>

<sup>††</sup> 東京農工大工学府工学研究院

## 1. はじめに

2008 年度の中学校学習指導要領改訂により、2012 年度から中学 1・2 年生の体育の授業にて、現代的なリズムのダンスやフォークダンスの学習が義務づけられた[1]。しかし 2013 年度の中学校全体における教員 1 人あたりの生徒数は、14.0 人であり、1 学級あたりの生徒数は 29.0 人である[2]。よって、少数の生徒に指導がおよばない可能性がある。そこで本稿では、動画から被写体の動作を推定することで、休日に家庭でダンス初心者の中学生在が、授業で覚えられなかった振り付けを習得できるようなシステムを提案する。

## 2. 提案システム概要

### 2.1 身体骨格データの推定

ユーザのダンスを動画で撮影し、本システムに入力する。コンピュータは、入力動画を画像に切り分ける。切り分けられた画像を入力画像 A とおく。その後、入力画像 A の HOG(Histograms of Oriented Gradients)特徴量をキーにデータベース内の類似画像を検索する。そして見つかった類似画像 A' に対応づけられている身体骨格データを、入力画像 A に新たに対応づける。この作業を動画長分繰り返し、入力動画の動作を推定する。

### 2.2 動作の比較とフィードバック

手本となるダンスをカメラで撮影し、本システムに入力する。コンピュータは、2.1 で推定されたユーザの動作と手本の動作を比較する。骨格部位ごとの座標差を指標に、ダンスが間違っている箇所を検出する。最後に、修正すべき部位と内容を画面に出力する。



図 2. 画像(左)と身体骨格データ(右)の対応例

## 3.2 HOG 特徴量を用いた類似画像検索

64×48ピクセルのグレースケール画像の HOG 特徴量を用いて画像検索を行う。入力画像を 16ピクセル×16ピクセルのセルに分割し、セル内の輝度から勾配強度と勾配方向を計算し、セルごとに特徴ベクトルを求める。勾配方向は、0~180° を 45° で分割した 4 方向を用いた。次に 3 セル×3 セルのブロックごとに特徴ベクトルを集計する。そして全ブロックの特徴ベクトルを連結し、72 次元の HOG 特徴量を得る。この HOG 特徴量  $u$  に対し、式(1)で記述されるマンハッタン距離  $d$  が最小となる HOG 特徴量  $v$  をもつ画像を類似画像とする。

$$d = \sum_{k=1}^{n=72} |u_k - v_k| \quad (1)$$

## 4. 類似画像データの検索結果

実装には C++, OpenCV2.0 を使い、また実験には 10 人分の walking の動作を使用した。ある人の動作 25 枚の画像を入力し、その 25 枚を除いた 9 人の 210 枚の画像から類似画像を検索した。その結果を図 3 に示す。



図 3. 入力画像(左)と検出された類似画像(右)

## 5. まとめ

HOG 特徴量を用いた類似画像検索を行った。今後は身体骨格データの推定や動作の比較を実装する。

## 参考文献

[1]「中学校学習指導要領解説 保健体育編 平成 20 年7月 文部科学省」  
[http://www.mext.go.jp/component/a\\_menu/education/micro\\_detail/\\_icsFiles/afildfile/2011/01/21/1234912\\_009.pdf](http://www.mext.go.jp/component/a_menu/education/micro_detail/_icsFiles/afildfile/2011/01/21/1234912_009.pdf)  
 [2]「文部科学統計要覧(平成 25 年版):文部科学省」  
[http://www.mext.go.jp/b\\_menu/toukei/002/002b/1337986.htm](http://www.mext.go.jp/b_menu/toukei/002/002b/1337986.htm)  
 [3]「MSR Action Recognition Datasets and Codes」  
<http://research.microsoft.com/en-us/um/people/zliu/actionrecorsrc/>

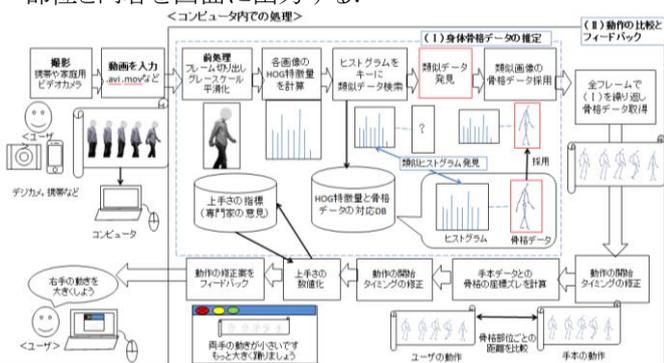


図1. 提案システムの概要図

## 3.身体骨格データの推定手法

### 3.1 動画と身体骨格データの対応データベース

Microsoft が公開しているデータセット[3]を使用した。Walking, stand up など 16 種の動作の動画, 身体骨格データ, 深度データが 10 人分収録されている。これはモーションセンサ Kinect を用いて計測されている。この動画をフレームごとに分けて、背景を削除した画像を作成する。そして、背景を削除した画像と身体骨格データを対応づける。