

# 社交ダンスの動作特性を考慮した マルチモーダルセンサによるダンスフィガー認識

松山 仁<sup>1,a)</sup> 浦野 健太<sup>1</sup> 廣井 慧<sup>1,†1</sup> 梶 克彦<sup>2</sup> 米澤 拓郎<sup>1</sup> 河口 信夫<sup>1</sup>

受付日 2019年12月23日, 採録日 2020年7月7日

**概要:** 本研究では社交ダンスを対象とし、加速度・角速度、視覚のマルチモーダルセンサを用いたダンスフィガー分類手法を提案する。社交ダンスにおける基本技術であるフィガーは、その種類の多さや複雑さゆえ初心者や中級者にとって練習が困難である。フィガーの自動認識によりユーザは自身の踊りの客観的な把握が可能となるため、ダンスフィガーの学習支援が可能となると考えられる。一方で、社交ダンスは2人1組で多彩な動きを行うため、既存の行動認識手法をそのまま適用することは難しい。本稿では、社交ダンスの動作特性を考慮し、フィガーの複雑さや遮蔽などの課題を解決したダンスフィガー認識手法を実現した。本研究では一般的な行動認識手法をベースライン手法として実装したうえで、社交ダンスの姿勢や動作特性を考慮した特徴量を設計・利用した手法を提案・実現し、両者の評価を行った。結果、提案手法の認識精度はF値0.97となり、ベースライン手法を全体で0.06上回った。特にフィガー別の分類精度では、最大で0.6の精度向上を達成した。さらに提案手法が遮蔽物に対しても頑健さを有することを示した。

**キーワード:** 行動認識, 機械学習, データ収集, 社交ダンス, ユビキタスコンピューティング

## Ballroom Dance Figure Recognition Using Multi-modal Sensors Considering Dance Motion Characteristics

HITOSHI MATSUYAMA<sup>1,a)</sup> KENTA URANO<sup>1</sup> KEI HIROI<sup>1,†1</sup> KATSUHIKO KAJI<sup>2</sup> TAKURO YONEZAWA<sup>1</sup>  
NOBUO KAWAGUCHI<sup>1</sup>

Received: December 23, 2019, Accepted: July 7, 2020

**Abstract:** The paper presents a ballroom dance figure classification method by extracting motion characteristics of dances using vision sensor, accelerometer, and gyroscope. It is difficult for intermediate students to practice figures, as they have tremendous types and movements. Thus enabling automatic recognition of the figure will help their exercises by letting dancers grasp themselves objectively. On the other hand, it is difficult to apply the existing activity recognition method as ballroom dance performance has a variety of movements in pairs of dancers. In this paper, we developed the dance figure recognition method robust to complex movement and occlusion by considering the characteristics of the dance. In this study, we first implement a general activity recognition method as a baseline method, then we proposed and realized a method of feature extraction utilizes the posture and motion characteristics of ballroom dance and evaluated both methods. As a result, we showed that the proposed method is 0.06 better than the baseline method in the F1 score and has robustness against occlusions.

**Keywords:** activity recognition, machine learning, data collection, ballroom dance, ubiquitous computing

<sup>1</sup> 名古屋大学大学院工学研究科情報・通信工学専攻  
Graduate School of Engineering, Nagoya University, Nagoya,  
Aichi 464-0814, Japan

<sup>2</sup> 愛知工業大学情報科学部情報科学科  
Department of Information Science, Aichi Institute of Tech-  
nology, Toyota, Aichi 470-0392, Japan

<sup>†1</sup> 現在, 京都大学防災研究所  
Presently with Disaster Prevention Research Institute,  
Kyoto University

<sup>a)</sup> hitoshi@ucl.nuee.nagoya-u.ac.jp

### 1. はじめに

社交ダンスは年齢や性別を超えて広く楽しまれているスポーツである。レッスンやデモンストレーション、パフォーマンスの完成度を競う競技会を通じて多くの人に楽しまれ [1], 高齢者の認識機能の保持効果も認められている [2]。ダンスパフォーマンスの評価は動きの完成度や美し

さ、音楽性により決められ、ほとんどの人がこれらの技術向上を目標に練習を行っている。近年、ダンスの練習支援を行うため情報技術を駆使したシステムの考案がなされている [3], [4], [5]。これらの研究ではダンス教師の踊り方を受講者に視覚的、触覚的に伝達する、またはシステム自体が教師の役割を果たすことでダンス練習支援を行っている。

しかしこれらの研究では、パフォーマンスを構成する「ダンスフィガー」に重点を置いていない。ダンスフィガーは2~3歩程度の足運びをひとまとめにした振付で、それらの組合せによって一連のパフォーマンスが行われる。各フィガーには踊り方の指針が存在する [6], [7], [8] ことから、社交ダンスの学習支援システムを開発するうえで、フィガー情報をシステムに与えることは重要である。しかし、フィガーの並べ方・構成は踊り手ごとに異なり、レッスンのたびに変更されることも多い。そのため、ユーザがそれぞれ自分のフィガー構成の情報をシステムに逐一与えることは非効率的である。そこで我々は行動認識の技術を適用してダンスフィガーを自動識別し、フィガーごとの踊り方の改善を行うことでパフォーマンス全体を向上させる社交ダンス練習支援システムの構築を目指す。本システムにより、新しくルーティンを振付けられた場合のみでなく、レッスンでのフィガー構成の変更にもフィガー情報をシステムに与えることが不要となり、さらに種目・内容を指定せずに自由にパフォーマンスを練習する際にもフィガー単位での動作を確認することが可能となる。

社交ダンスは2人1組で多彩な動きを行うため、一般的なダンス認識に用いられる視覚センサを用いた手法では微細な特徴の差異や人による遮蔽に対応しきれない。また、社交ダンスは一般的な行動と比較して行動クラスが多い一方、データ収集に協力可能な人数が限られていることから、大規模データの収集難易度と必要な期間が長くなる。そこで、データ量が十分でない場合にも活用できる特徴量設計を加えた手法が必要である。これらの要件から、本研究ではマルチモーダルセンサを用い、さらに社交ダンスの動作特性を考慮した特徴量設計によるダンスフィガー分類手法を提案する。

本研究でははじめに、7人のダンス経験者が踊る13種類のフィガーデータを合計1,820演技分収集した。本来社交ダンスは2人で踊るスポーツだが、個人練習は1人で行われるものが最も基礎的であることから、まずは単1人物でのフィガー認識を行った。また動画に加え、ウェアラブルセンサを用いて加速度・角速度を取得し、遮蔽物や微細な動きの違いなど2次元視覚情報の不足を補った。動画データはOpenPose [9] を用いて骨格位置情報に変換し、これらのデータを基にまずは加速度と角速度、動画、および両者を併用した場合について、基本の統計量である平均値と分散値を用いたフィガー分類と評価を行った。ウェアラブルセンサの装着位置の候補は両肘、両腰、両足首の6カ所用

意し、部位ごとの精度変化および視覚情報と併用した場合の精度を評価した。次に、最も精度の高かった位置のウェアラブルセンサと視覚情報それぞれに対して社交ダンスの動きの特性が現れやすい肘の動きや開き具合の角度変化、音楽アクセントの表現、フィガーの連結部分の変化特性などを表現する特徴量設計を行った。そして、それらの特徴量を用いて再度フィガー分類を行い精度を評価した。さらに遮蔽物に対する頑健性を確かめるため、体の一部のデータを欠損させて遮蔽の起きた状態を仮想的に再現して評価を行った。結果、提案手法の認識精度はF値0.97となり、ベースライン手法を全体で0.06上回った。特にフィガー別の分類精度では、最大で0.6の精度向上を達成した。また、遮蔽の起きた状態でもマルチモーダルセンサのハイブリッド手法によりF値0.91を、さらに特徴量設計を加えたことでF値0.97という精度を得た。

本研究による貢献は次の3つである。

- (1) 複数のセンサを対象とし、一般的な行動認識手法に基づいたベースラインとなるダンスフィガー認識手法を実装し、評価したこと
- (2) 社交ダンスの動作特性を考慮した特徴量を設計・利用したダンスフィガー認識手法を提案・実現したこと
- (3) ベースライン手法と比較し、精度や頑強さにおいて提案手法の有効性を評価したこと

本稿は以下のように構成される。2章で本研究の目的を述べ、関連研究を整理したうえで目的達成のための課題を明らかにする。3章ではデータ収集やラベリング、前処理を行った後にベースライン手法の設計および評価を行う。4章ではベースライン手法における課題点を整理し、それらを克服する社交ダンスの特性を考慮した特徴設計を提案する。そして設計した特徴量を5章で実装・評価して提案手法の有効性を示し、最後に6章で本稿をまとめる。

## 2. 社交ダンスにおける行動認識

### 2.1 社交ダンス支援におけるフィガー認識の重要性

本研究の最終的な目標は、ダンスフィガーの自動認識を用いた社交ダンス教示システムを構築することである。ダンスコーチングをサポートする最も一般的な方法は視覚センサを用いるものであり、主に教師の動きの情報を生徒に伝えることで生徒のパフォーマンスの修正を図る手法 [3], [4] と、システム自体が教師の役割を果たし、自動的に生徒のパフォーマンスを改善する手法 [5] とが存在する。これらは、個々の技、社交ダンスでいうフィガーの認識は行わず、手足の動作タイミングを一致させたり、パフォーマンスの一部または全体の姿勢改善を行っている。これらの手法は、動作の種類が少ない種目や、ポージング時間の長い種目での学習支援で特に有用であるといえる。

一方、社交ダンスのパフォーマンス改善において、フィガーの情報は重要である。社交ダンスは種目ごとに数十種

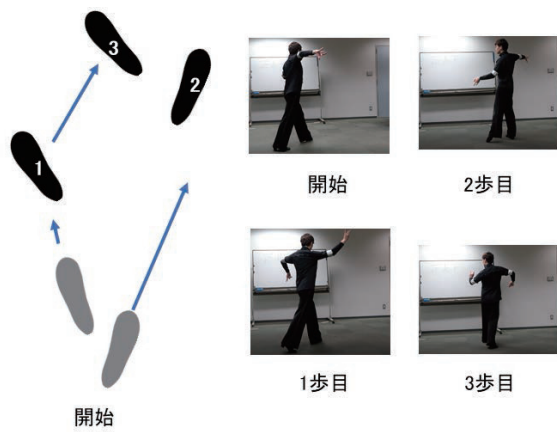


図 1 フィガーの一例 (Fan)

Fig. 1 An example of dance figures (Fan).

類またはそれ以上のフィガーを持ち、それぞれ踊り方が異なり、またフィガーごとに適した表現が求められる [6]. 中でも、ベーシックフィガーと呼ばれる基本的なフィガーは、ダンス教本で踊り方の指針が提示されている. 一例として Fan というベーシックフィガー (図 1) を示す. このフィガーは 1 歩目で左足を前進し、2 歩目で右足を左足に対して斜めに動かし、3 歩目で左足を右足に交差させるように動かしている. こういったベーシックフィガーは、他のあらゆる応用的なフィガーの土台となっており、ベーシックフィガーの完成度を向上させることはダンスパフォーマンス全体の向上に直結する. そのため、初心者から上級者まで、すべての競技者にとって重要な要素である. しかし、既存のダンス学習支援に関する研究は、個々の技に注目した支援を行っていないため、既存手法をフィガー学習に適用することは困難である.

ここで、社交ダンスレッスンとフィガーの関係とを段階に分けて考えると、(1) フィガーの踊り方と、一連のパフォーマンス (ルーティン) に各フィガーが登場する順番を覚える段階、(2) パフォーマンスを磨く段階の 2 段階が存在する. 本研究は、フィガー識別手法を活用したシステムにより、これら両方の段階の支援を可能にする. 具体的な利用シナリオの例を以下に示す.

- フィガー学習動画の生成：主に (1) の段階の初心者を対象に、指導者の模範的なダンスシーケンスを撮影し、フィガー名や注意点などをアノテーションした復習用動画を自動で作成
- フィガー検索機能：(2) の段階にて、フィガーの名前や踊り方の詳細を忘れたユーザが、踊りを記録することでフィガー名・踊り方を自動で調べられる
- 自由練習でのフィガー改善：(2) の段階にて、中・上級者のユーザが自由に様々なフィガーの練習を行い、各フィガーの動作を自動で識別しフィガーごとに改善  
本研究ではこれらのフィガー学習支援を目標に、まずはフィガーの自動認識手法を開発する. 社交ダンスは複雑な

動作を多く含み、また 2 人 1 組で踊られるため、次節で述べるように視覚・慣性のマルチモーダルセンサによる認識を行う. また、社交ダンスは一般的な行動と比較して行動クラスが多い一方、データ収集に協力可能な人数が限られていることから、大規模データの収集難易度が高くなる. そこで本研究では、社交ダンス動作特性を考慮した特徴量設計を用い、データ量が十分でない場合でも有用な、機械学習による分類手法を提案する.

## 2.2 関連研究

一般的にジェスチャ動作の認識では波形マッチング処理により動作検出を行う [10] が、社交ダンスの特にラテンアメリカンは動作の自由度・複雑度が非常に高く、さらに個人間で動作にばらつきがあることも多い. そのため、波形マッチングによりフィガー種別を検出するには多くの工夫が必要である. そこで我々は、ラテンアメリカンにおけるクアラチャアクションという連続的動作に注目した. このアクションはすべてのフィガーを通じて、骨盤周辺を動作起点として連続的に続けられており、手や足の動作もすべてこのアクションの延長でなされている [8]. このアクションは周期的・連続的であり、一方でフィガーごとの進行方向や回転にともない動作特性が変化する. これは静止や歩行などの日常生活における行動の特徴と共通している. そこで、本節では行動認識の主だった手法について整理し、適切な手法の検討を行う. 行動認識の主要なアプローチは (1) ウェアラブルセンサやスマートフォンの慣性センサなどのセンサデータを活用するもの (2) RGB カメラや深度カメラなどの視覚センサを活用するもの (3) 慣性センサと視覚センサを組み合わせたマルチモーダルセンサによるものに大別できる.

### 2.2.1 慣性センサを用いた手法

もっとも一般的な行動認識手法の 1 つが、スマートフォンやウェアラブルデバイスなどに内蔵された加速度計・ジャイロスコープを用いるものである [11], [12]. 収集されたセンサデータの処理により行動認識が行われ、行動認識技術を用いたヘルスケア [13] や、位置情報のトラッキング [14] などに利用される.

センサデータによる行動分類もさかに行われており、日常生活の行動クラス分類 [15] に加え、執筆や食事など細かな認識を行う研究 [16] も行われている. これらの手法は、視覚センサの手法と比べてデータ取得場所の風景や撮影方向、環境光による影響を受けず、身体に直接装着する場合には身体の細かな動作の情報を得られる利点がある. 一方で、センサは装着されている部分の信号を取得するのみであり、動きをとらえたい部位が増えるほど、装着するセンサ数も増加するという欠点が存在する.

### 2.2.2 視覚センサを用いた手法

RGB 画像や深度カメラなどの視覚センサによる手法も、行動認識の主要な方法である。特に近年さかんなのが、画像から人物の骨格位置を取得する、姿勢推定技術 [9], [17] を用いた行動認識である。

姿勢推定による人物の骨格位置、器官点位置情報を用いた行動分類も活発に行われ、歩行やジョギングなどの静止画を分類するタスクや、音楽活動・ウィンタースポーツなど多岐にわたる行動クラス分類も行われている [18]。これらの方法の長所は、マーカレスで人物の全身の情報が取得できることである。一方で欠点として、撮影場所やカメラの角度によって同じ人物の行動であってもデータが変化すること、遮蔽物や光量の影響を受けること、信号処理と比較して画像処理にかかる計算コストが大きいことがあげられる。

### 2.2.3 マルチモーダルセンサを用いた手法

慣性データ、または視覚データのみを活用する手法に加え、近年はこれら2つの手法を組み合わせた、マルチモーダルセンサによる行動認識の研究も増加している [19]。行動分類においては、慣性センサ・視覚データそれぞれの単一情報を用いた行動分類を行ったうえで、何らかのルールを設けて最終的な行動クラスを決定する手法と、最初からそれぞれのセンサの特徴量をあわせて入力し、重要度の高い特徴を選択して行動分類を行う手法とが存在する。両センサの利点を合わせることによる行動認識の精度の向上は認められており [20], [21]、さらに深層学習を用いたより頑健な行動認識も行われている [22], [23]。こういった手法は単一モーダリティでの行動認識と比較して高精度な結果が期待できる一方、視覚センサの長所である、マーカレスでの認識が行えなくなる。そのため、装着するウェアラブルセンサの数は最小限にし、使用者の負担を低減させることが望まれる。

社交ダンスは2人1組で位置を入れ替え続けながら踊るため他の踊り手が遮蔽物となりやすく、マルチモーダルな手法が必要だと考えられる。しかしこれまで社交ダンスを対象としたフィガー識別の研究は行われておらず、実際にどの程度の精度で認識が行われ、またどの程度マルチモーダル手法により遮蔽物に対する頑健さが得られるかは明らかでない。

### 2.2.4 既存研究をふまえた目的

本稿ではフィガー認識を行うため、まず既存研究をふまえた一般的なアプローチでどの程度フィガー認識が可能であるか明らかにする。そのうえで社交ダンスの特徴を考慮した手法として改良し、認識精度がどの程度向上するかを明らかにする。さらに社交ダンスが2人で踊られた場合の相互の視覚的遮蔽を再現し、視覚データの欠損に対する頑健さを検証する。

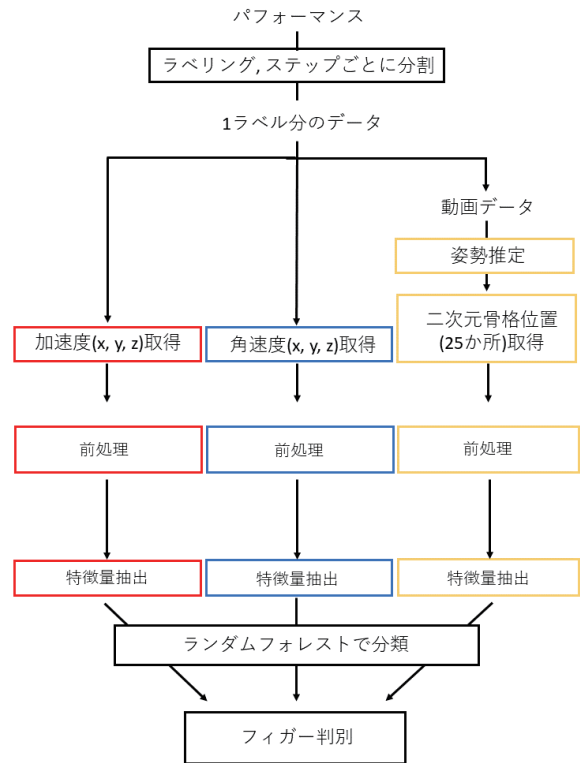


図 2 ベースライン手法の処理手順

Fig. 2 Data processing overview of the baseline method.

## 3. ベースライン手法の設計と実装

本章ではダンスフィガー分類の評価のベースラインとして、図 2 の手順で各センサデータから基本的な統計量を抽出して分類を行う。データ収集やラベリング、前処理についても本章で記述する。

### 3.1 認識対象とするダンスフィガー

本研究で対象とするのは、社交ダンスで最も基本的な種目の1つである、ラテンアメリカンのルンバである。社交ダンスの各種目にはベーシックフィガーと呼ばれる基礎的な振付が少なくとも数十種類存在し、初心者からプロまですべての踊り手が練習する。今回はそのなかから、類似した動きや左右対称な動きを含む13種類を選択する。実際のパフォーマンスでも類似したフィガーや左右対称なフィガーは頻繁に登場するため、これらを識別することが本研究の1つの課題となる。それぞれのダンスフィガーの様子は、図 3 に示される。各フィガーの区切り方には複数の考え方があがるが、3の拍から各フィガーの予備動作が始まっている [7], [8] ことから、今回は3の拍で開始、2の拍の最後で終了と定義する。図 3 における OpenBasic, FootChange などはフィガー名で、今回選択した物はすべて3歩で構成される。それぞれ観察すると、HockyStick と Alemana は1, 2の拍での体の向き、左手の動きを除き同

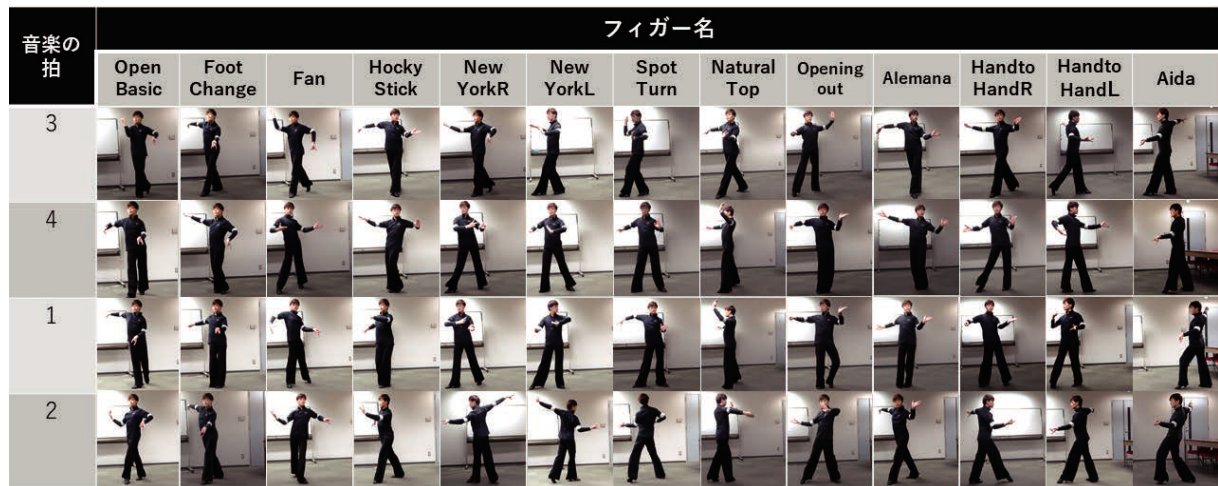


図 3 各ベーシックフィガーの名前と音楽に従った動作推移

Fig. 3 Dance figures and movements with respect to music counts.

じ動きをしていること、R・Lがついているフィガーはそれぞれ左右対称な動きをしていることなどが分かる。

### 3.2 分類の手順

ベースライン手法による分類では、動画データから骨格位置推定を行い、加速度・角速度データとともに行動認識で一般的な統計量である平均値・分散値を特徴量とし、ランダムフォレストを用いて図 2 の手順で分類を行う。はじめにフィガーの分類を行うために、各センサのデータに対して音楽的規則性を基に自動的にデータ分割してラベリングを行う。

次に特徴抽出について、一般的な行動分類であげられるのは、加速度や角速度の信号データに対して平均・分散の統計量を特徴抽出して分類モデルへ入力することである [11], [12]。また動画データについては、骨格位置推定手法を利用した分類手法が知られており [18]、そのうちの OpenPose [9], [24] を利用する。骨格位置推定を利用する場合、静止画であれば特徴量は骨格位置の座標そのものが使われるが、今回は時間変化のある動画像を用いるので、座標そのものの代わりに加速度・角速度と同様に平均・分散の統計量を入力とする。特徴抽出にあたり、骨格位置については画面内のダンサーの位置や移動、また身長やカメラとの距離の影響を除去するため前処理を施す。

### 3.3 データ収集

マルチモーダルなデータを含むデータセットはいくつか存在する [20], [25] もの、社交ダンスに焦点を当てたデータセットが存在しないため、はじめにデータ収集を行った。社交ダンスの経験者 7 人を集め、あらかじめ準備された 13 種類、合計 19 個（フィガーの接続の関係で、13 種類をそのままつなげることは困難）のベーシックフィガーのみで構成された一連の演技を踊ってもらい、そのデータを収集

表 1 踊り手の情報

Table 1 Height and experience of dancers.

	身長	経験年数
踊り手 1	173	17
踊り手 2	176	5
踊り手 3	182	1
踊り手 4	160	1
踊り手 5	171	4
踊り手 6	175	3
踊り手 7	177	5

した。

7 人の踊り手の身長、経験年数をまとめて表 1 に示す。本来社交ダンスは 2 人 1 組で踊られるものだが、個人練習は 1 人で行われるものが最も基礎的（シャドー練習）なため、まずは単独でのフィガー認識が重要となる。よって本研究ではまず、単独ダンサーによるフィガー認識の評価を行う。また同じダンスフィガーであっても男性と女性では踊り方が異なるため、今回は男性のフィガーのみを収集する。本研究の目的は、ユーザのダンスパフォーマンスにおけるフィガーを自動識別し、練習支援を行うことである。そのため、フィガー認識においては一流の上手なフィガーだけでなく、様々な完成度のフィガーを識別する必要がある。そこで今回は、ベーシックフィガーが人目で認識できれば習熟度は問わず、少なくとも 1 年以上の経験のある学生ダンサーを対象にダンスフィガーデータの収集を行った。

それぞれの踊り手の両腕、両腰、両足首にはウェアラブルセンサ (ATR-Promotions, TSND151) が装着されており、加速度と角速度の信号をサンプリングレート 120 Hz で取得する。利用しやすい練習支援システム実現するにはウェアラブルセンサの装着数は少ないほうが利用者の負担が少ないことから、評価においてはこれらのうち 1 カ所を選択して精度を確認し、最も高精度にフィガー分類を行え

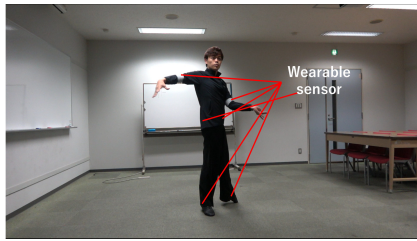


図 4 撮影方向その 1 (正面から)  
Fig. 4 Shooting direction 1.

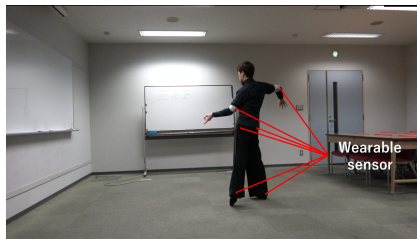


図 5 撮影方向その 2 (反対方向から)  
Fig. 5 Shooting direction 2.

る部位を見つけ出す。また同時に、120 fps で動画を記録できるビデオカメラ (SONY, FDR-AX60) を用意し、2 種類の撮影方向 (図 4 および図 5) からダンスの動画を記録する。踊り手には 13 種類、19 フィガーが連結された演技をそれぞれ 20 回ずつ踊ってもらい、それぞれの加速度・角速度・動画データを取得する。ただし、動画データについては、最初の 10 回を正面方向から、残りの 10 回を反対方向から撮影する。また、それぞれの動画には踊り手 1 人のみが写っており、その後の人物抽出において異なる人物の誤検知が起きないように留意した。なお本研究で収集したデータは、行動認識分野の発展に寄与するため、研究者向けに公開済みである [26]。

### 3.4 実装

本節では、フィガー分類に先立ち人物の移動影響の除去やスケーリングを行い、ダンス経験者により目視でフィガー名をラベリングした後に前節で述べた手法の実装を行う。

#### 3.4.1 骨格位置の前処理

ビデオデータからダンスフィガーの様子を抽出するため、はじめに OpenPose [9] を用いて各フレームの画像における人物の 25 カ所の骨格位置推定を行う。抽出された骨格位置は、図 6 に示す。この骨格位置情報はフレームごとに入手され、動画データのすべてのフレームについて骨格位置を取得し、各踊り手の体のパーツの位置の時間変化の様子が取得できる。ただし、入手された骨格位置情報は、体の向きや撮影方向の影響などにより、しばしば欠損値を含むことがある。欠損値については、python における pandas ライブラリ\*1 の interpolate 関数を用いて補間する。

\*1 <https://pandas.pydata.org/>

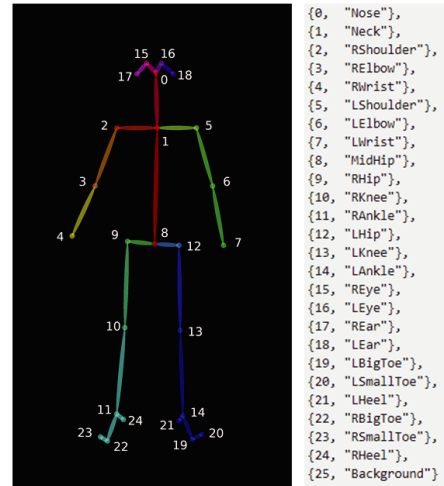


図 6 取得される骨格位置  
Fig. 6 Skeletal positions acquired.

#### Algorithm 1 移動影響の除去

```

1: for each keypoint_x,y of the 25 keypoints in each frame
   do
2:   keypoint_x -= Neck_x
3:   keypoint_y -= Neck_y
4: end for
    
```

#### Algorithm 2 身長・距離の正規化

```

1: for each keypoint_x,y of the 25 keypoints in each frame
   do
2:   keypoint_x = keypoint_x ×  $\frac{\text{distance}(\text{Neck}, \text{keypoint})}{\text{Neck}, \text{MidHip}}$ 
3:   keypoint_y = keypoint_y ×  $\frac{\text{distance}(\text{Neck}, \text{keypoint})}{\text{Neck}, \text{MidHip}}$ 
4: end for
    
```

骨格位置データは、画像内における各関節点の座標である。これを入力とすると、撮影方向や動作開始位置によって画面内で位置が変化したり、踊る位置が移動し、座標が変化してしまう。この影響で、同じダンスフィガーを踊っていても、モデル学習の際に意図しない画面内移動による影響を受ける恐れがある。そのため、前処理の段階で、Algorithm1 を用いて、体の各部位の座標を、首の座標からの相対座標で表現し直すことで、画面内移動による影響の除去を行う。

また、同じダンスフィガーを踊っていても、踊り手の体の大きさが変化すると、体の各部位の位置関係にも違いが生まれる。これを補正する手段として、Algorithm2 を用いてすべての踊り手の体の部位の相対位置のスケーリングの統一を行った。

#### 3.4.2 特徴量抽出

ここではベースライン手法として、フィガー長ごとに平均値・分散値を取得する。一般的にジェスチャ動作の認識では波形マッチング処理により動作検出を行う [10] が、社交ダンスの特にラテンアメリカンは動作の自由度・複雑度が非常に高く、さらに個人間で動作にばらつきがあること

も多い。そのため、波形マッチングによりフィガー種別を検出するためには多くの工夫が必要になると考えられる。そこで我々は、ラテンアメリカンにおけるクカラチャアクションという連続的動作に注目した。このアクションはすべてのフィガーを通じて、骨盤周辺を動作起点として連続的に続けられており、手や足の動作もすべてこのアクションの延長でなされている [8]。このアクションは周期的・連続的であり、一方でフィガーごとの進行方向や回転にともしない動作特性が変化することから、静止や歩行などの動作検出と同様に平均・分散を抽出することでフィガー検出が可能であると予想される。まず、前節で手動または自動でラベリングされた各フィガーについて、フィガー単位でデータを分割する。そして、各フィガー単位に対して、加速度、角速度、25カ所の体の部位の2次元座標の時間変化について、それぞれ平均値・分散値を計算し、それらの値をモデルへの入力とする。姿勢推定による骨格位置を用いた分類問題を解く場合、静止画であれば文献 [18] のようにデータ自体を入力にすることが考えられるが、今回は時間変化のあるデータを扱うため、加速度、角速度を用いた行動分類と同様に平均値と分散値を利用する。計算される特徴量を以下にまとめる。

- 加速度および角速度：それぞれの装着部位の各軸の信号の各1フィガー分の時間変化について、平均値，分散値（合計72種類）
- 骨格位置：各部位の2次元座標それぞれの1フィガー分の時間変化について、平均値，分散値（合計100種類）

### 3.5 ベースライン手法によるフィガー認識実験

加速度・角速度信号を用いた特徴抽出による分類について評価する。ウェアラブルセンサの装着位置候補として存在する6カ所それぞれについて精度 (F 値) を比較したものを表 2 にまとめる。各精度はランダムフォレストを用い、フィガーベースの5分割交差検証により算出されたものである。精度を比較すると、腰に装着した場合の精度を除き大きな精度の差は確認されなかった。次に、骨格位置による分類精度を同表にまとめたいうで、ウェアラブルセンサの各部位と骨格位置との組合せによる精度について表 2 右列にまとめる。ウェアラブルセンサの精度と比較すると、骨格位置による分類精度は 0.74 と低いことが確認された。

表 2 ベースライン手法による分類精度 (F 値)

Table 2 Classification accuracy with the baseline method.

センサ部位別精度	骨格位置	ハイブリッド
左足首	0.87	0.92
左肘	0.86	0.90
左腰	0.80	0.88
右足首	0.83	0.74
右肘	0.85	0.91
右腰	0.79	0.88

両者のハイブリッドによる精度では左の足首のウェアラブルセンサとの組合せが最も高い精度を示した。これは社交ダンスの特にラテンアメリカンにおいて重要な足さばきをウェアラブルセンサがとらえ、他の全身の動きを骨格位置がとらえられたことで両者の長所を生かした分類器を作成できたと考えられる。一方で、ウェアラブルセンサ・骨格位置それぞれを用いた分類精度はいずれも 0.9 を下回っており、改良が必要である。

## 4. 社交ダンスの動作特性を考慮した手法改良

ベースライン手法にて分類が困難であったフィガーの多くは、大まかな動作特性が同じで肘の動きや上半身の回転量に違いがあるものや、前後のフィガーとの連続による動作特性変化により前後のフィガーと誤認識されたものなどであった。そこで本章では社交ダンス、特にラテンアメリカンのベーシックフィガーの特性を考慮しこれらの課題を克服可能な特徴量を設計する。

### 4.1 社交ダンスのための特徴設計

注目する点は大きく2種類で、1つ目が「連続フィガーにおける各フィガーの動作変化傾向」、もう1つが「体や腕の動作特徴」である。

#### 4.1.1 連続フィガーにおける各フィガーの動作変化傾向

社交ダンスにおける各フィガーはそれぞれ固有の足捌きや骨盤、上半身の動きを持っている。これらは単独のフィガーとして踊られることもあるが、多くの場合は前後に異なるフィガーが配置されて複数のフィガーを連続して踊られる。この場合、前後のフィガーの種類に合わせてフィガーのつなぎ目にあたる部分の動作特性に変化が生まれる。このつなぎ目にあたる部分は2フィガーにわたって存在するため、図 7 に示すように各フィガーについてその前後1拍分の長さのデータを含めた6拍分のデータを割り当て、それを「前のフィガーからの遷移部分」、「現在のフィガー固有部分」、「後のフィガーへの遷移部分」の3つに分割して特徴抽出を行う。ただし加速度・角速度を用いた分類では連続したフィガー間の誤判定があまり見られなかったため、この処理は骨格位置情報に対して適用する。

#### 4.1.2 体や腕の動作特徴

1つ目の動作特徴は、フットアクションの行われるタイ

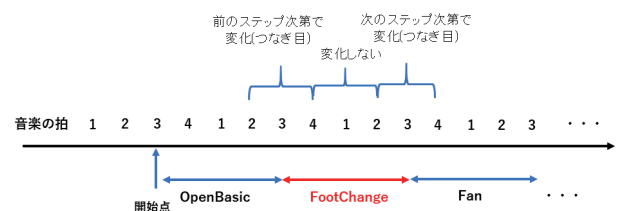


図 7 連続フィガーの様子

Fig. 7 An example of how continuous figures are connected.

ミングである。社交ダンスのフィガーには、左右対称の鏡像の関係にあるものが多く存在するため、どちらの足をどのタイミングで接地させたかを判定し左右の動きを区別する。手法としてはそれぞれのフィガー開始後のどのタイミングで最も加速度が強く変化したかを取得する。

フットアクション以外で社交ダンスの動きの特性を抽出できる要素に、腕やひじの作る角度の情報がある。これは、社交ダンスのラテンアメリカンにおけるリードアームとフリーアーム [6] の特性によるものである。リードアームとは男性が女性を導く手であり、フィガーごとに女性を導く向きや方法が異なる。フリーアームとはリードアームでないほうの手であり、動かし方に明確な決まりはないものの、観察すると一定の規則性が見出される。これらの特徴を抽出するために、(1) 左右手首・肘・肩の形成する角度（ひじの角度）、(2) 首と左右肩・肘の作る角度（肩のラインからみた肘の上がり具合）の値変化、および(3) 手首の高さが最大となるタイミングの値を取得する。

3つ目の要素は上体の回転である。フィガーの指南書 [6], [7], [8] を見ると、それぞれのフィガーについて体の回転量やそのタイミング、向きが指定されていることが分かる。ここでは、左右肩の2次元座標を利用して肩間のユークリッド距離を算出し、その値変化から体の回転の度合いを取得する。この値は、踊り手が撮影方向に対して垂直方向を向いている場合に最小、平行方向を向いている場合に最大となることから、体の回転量を表す指標として活用することができる。

## 4.2 実装

本節では特徴量の計算方法、およびダンスシーケンスへのフィガー情報の自動ラベリング方法について述べる。

### 4.2.1 特徴量の計算

はじめに、連続フィガーにおける各フィガーの動作変化傾向をとらえた特徴抽出を行う。注目しているフィガーに対して図7に示されるような「前のフィガーからの遷移部分」、「現在のフィガー固有部分」、「後のフィガーへの遷移部分」それぞれの部分を取り出し、各骨格位置の時間変化（x座標またはy座標を縦軸、時間を横軸とする）に対して平均値と分散値を計算する。

続いてフットアクションの抽出を行う。タイミングについては左足首に装着したウェアラブルセンサの3軸加速度からノルムを計算し、各フィガーの288サンプル中でその値が最大となる位置（インデックス、1から288番目のいずれか）を取得する。動作方向と変化の様子については各軸の信号の前半144サンプル（動き出しの動作傾向）をそれぞれ線形近似し、係数を取得する。

次に腕や肘の作る角度の抽出を行う。はじめに肩からみた肘の作る角度は骨格位置情報の肩、肘の座標をそれぞれ

$$\overrightarrow{Shoulder}(x, y) = \overrightarrow{Sh}(x, y), \overrightarrow{Elbow}(x, y) = \overrightarrow{El}(x, y)$$

として式(1)により計算する。

$$elbow\_angle = \frac{(\overrightarrow{Sh}(x, y) \cdot \overrightarrow{El}(x, y))}{(\|\overrightarrow{Sh}(x, y)\| \times \|\overrightarrow{El}(x, y)\|)} \quad (1)$$

ここでShoulder, Elbowはそれぞれ図6のようにOpenPoseにより取得されるR(L)Shoulder, R(L)Elbowを指す。フィガー内の各フレームについてこの角度を入手し、角度の時間変化に対して分散値を計算する。

また肘の開き具合の角度は骨格位置情報の肩、肘、手首の座標をそれぞれ

$$\overrightarrow{Shoulder}(x, y) = \overrightarrow{Sh}(x, y), \overrightarrow{Elbow}(x, y) = \overrightarrow{El}(x, y), \\ \overrightarrow{Wrist}(x, y) = \overrightarrow{Wr}(x, y)$$

として、式(2)により計算する。

$$arm\_angle \\ = \frac{((\overrightarrow{Sh}(x, y) - \overrightarrow{El}(x, y)) \cdot (\overrightarrow{Sh}(x, y) - \overrightarrow{Wr}(x, y)))}{(\|\overrightarrow{Sh}(x, y) - \overrightarrow{El}(x, y)\| \times \|\overrightarrow{Sh}(x, y) - \overrightarrow{Wr}(x, y)\|)} \quad (2)$$

ここでWristはOpenPoseにより取得されるR(L)Wristを指す。これをフィガー内の各フレームについて求め、角度の時間変化を入手する。その後、入手した角度の時間変化の前半半分を線形近似し、その係数を取得する。

続いて手首の高さが最大となるタイミングを抽出するため、OpenPoseによるRWrist, LWristの1フィガー、288サンプル分のy座標の時間変化について値が最大となるインデックスを取得する。

最後に上体の回転に関する情報を取得する。骨格位置のうち、両肩の座標を

$$\overrightarrow{RShoulder}(x, y) = \overrightarrow{RSh}(x, y), \\ \overrightarrow{LShoulder}(x, y) = \overrightarrow{LSh}(x, y)$$

として、左右肩の距離を式(3)により計算する。

$$chest\_distance \\ = \sqrt{(\overrightarrow{RSh}_x - \overrightarrow{LSh}_x)^2 + (\overrightarrow{RSh}_y - \overrightarrow{LSh}_y)^2} \quad (3)$$

この肩間の距離をフィガー内の各フレームについて求め、距離の時間変化を入手する。そして得られた時間変化に対して、合計値と分散値を取得する。

### 4.2.2 自動ラベリング

今回収集されたデータは、複数のダンスフィガーが連続した状態で取得されている。これは、収集されるフィガーの種類や数が多く、すべてを単一で取得するのが困難であったためである。前章では各データを目視で確認してラベリングを行ったが、現時点でデータは2時間以上におよ



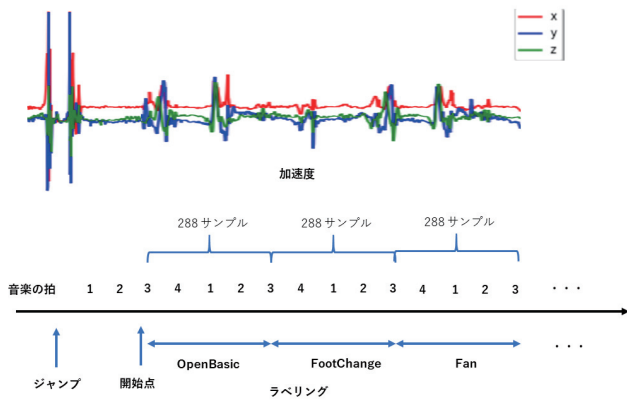


図 8 自動ラベリングの様子  
Fig. 8 Automatic figure labeling.

び、今後さらにデータを増量することを考えるとこの手法は非現実的である。そこでウェアラブルセンサのデータと動画データそれぞれの開始点をジャンプ位置により同期したうえで、連続したフィガーデータに対して音楽のBPMと各フィガーの持つ音楽リズム的な長さをを用いた自動ラベリング手法を以下の手順で行う (図 8 参照)。

- (1) 登場するフィガーの持つ拍の数を調べる (今回はすべて 4 拍で 1 フィガー)。
- (2) 使用する音楽の BPM を調べる (今回は BPM = 100)。
- (3) 1 分で 100 拍、つまり 25 回フィガーが登場するので、

$$60(\text{second}) \times 120(\text{samples per second}) \div 25(\text{figures}) = 288(\text{samples per figure})$$

を計算し、1 フィガー分のサンプル数を導出

- (4) 最初のフィガーが開始される点のみ与え、上で計算されるサンプル数ごとにフィガーを分割
- (5) フィガーが登場する順番を与えることで自動的にラベリングを行う。

この処理により各フィガーに 4 拍分、288 サンプルのデータが割り当てられる。

## 5. 評価

本章では初めに 3 章の手動ラベリングと 4 章の自動ラベリングによるベースライン手法を用いた分類精度の比較を行い、自動ラベリングによる分類精度の低下が許容可能であることを示す。一般的な統計量を用いたフィガー分類結果と、提案手法であるダンスの特性を考慮した特徴量を用いた分類結果を比較し評価する。さらに、社交ダンスが本来 2 人で踊られるものであることを考慮し、骨格位置情報の一部を欠損させた状態で分類精度を算出し評価する。

### 5.1 自動ラベリングの評価

手動ラベリングと自動ラベリングそれぞれを用いて、ベースラインの特徴量を用いた分類を行った結果を表 3 にまとめる。結果、骨格位置による分類精度には変化が生まれ

表 3 自動・手動ラベリングによるベースライン手法の精度比較  
Table 3 Comparison of the accuracies between manual and automatic labeling.

センサ	F 値	
	手動	自動
加速度・角速度 (左足首)	0.87	0.84
画像	0.74	0.74
ハイブリッド	0.92	0.91

ず、加速度・角速度センサとハイブリッドについてもそれぞれ最大 0.05 の F 値の低下にとどまった。精度が低下した原因は、手動ラベリングの場合は踊り手が音楽の拍からずれて踊っていた場合 (「2」の拍の終わり目で終わるべきフィガーが、次の「3」の拍にはみ出して踊られてしまうような場合) にも踊り手の動きに合わせて柔軟にラベルを振ったのに対し、自動ラベリングでは一律に拍でラベルを区切ったため、拍からはみ出した部分が次のフィガーラベルに含まれてしまったためだと考えられる。しかし劇的な精度低下には至らなかったため、本研究では今後はこの自動ラベリング手法を採用し効率的にラベリングを行い、さらに手動ラベリングにより取得できていた音楽の拍からずれた動きを補うような特徴設計も目指す。

### 5.2 設計した特徴量の評価

本研究の目的は、フィガー情報を用いた社交ダンス練習支援システムの構築である。そこでフィガー認識における要求仕様・精度として、ダンス教師によるフィガー認識精度を指標とする。競技会における 1 曲の長さは 1 分 20 秒前後であり、競技者は 1 曲でルーティンを 1.5 から 2 周程度することが多いことから、ダンスルーティンはおおよそ 45 秒から 1 分程度、フィガー数にして 18 から 24 個を保有すると仮定できる。個人差はあるが熟練した教師ならば 1 回のルーティンあたりのフィガーの誤認識は 1 回までに抑えられると考え、精度にして、17/18(0.944) から 23/24(0.958) 程度が必要である。そこで本評価では、F 値 0.958 を目標の指標とする。

#### 5.2.1 全体の精度評価

ベースライン分類精度と提案手法による精度を各フィガー別、および集計したものを図 9 にまとめる。表 6、表 7 の各 1 行目に示されるように、加速度・角速度による分類精度が 0.09、骨格位置を用いた分類精度が F 値で 0.1 上昇しており、提案手法により精度が大きく向上することが確認された。特に動きの類似した Hockystick と Alemana についてはベースライン手法では分類が困難であった一方、提案手法では細かな特徴をとらえて正確に分類できたことが分かる。また、提案した特徴量設計を用いたハイブリッド認識の精度は 0.97 であり、これはダンス教師による認識精度の指標として仮定した 0.958 を上回る結果となった。

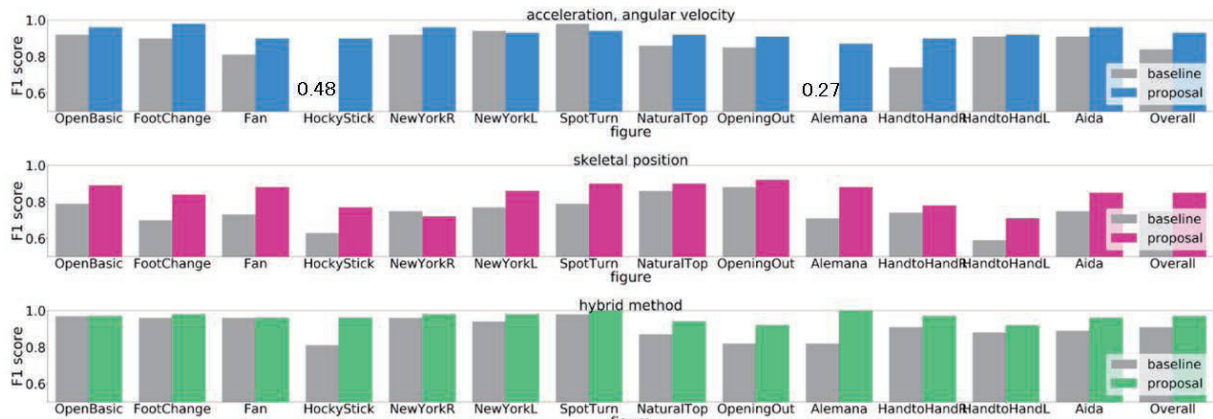


図 9 各センサにおけるベースライン・提案手法の比較

Fig. 9 A comparison of baseline and proposal method for each sensor modality.

5.2.2 加速度・角速度を用いた特徴量の評価

続いてそれぞれの特徴量の重要度について確認する。はじめに表 4 に表される加速度・角速度を用いたランダムフォレスト分類器の特徴重要度を確認すると、1位の y 軸角速度の平均値に続いて各軸加速度の最大値のインデックスがそれぞれ 2 から 4 位に現れた。これは、主に足が接地した際の足の動作方向とタイミングが抽出されたためだと考えられる。ラテンアメリカンの指南書 [8] によると足の進行方向は壁方向、中央方向など 8 方向が存在し、各フィガーのステップごとに方向が定められている。また、どのタイミングでどちらの足が動かされるかもフィガーごとに異なる。たとえば、ベースライン手法において分類が困難であった HockeyStick/Alemana の組における動作の様子 (図 10 参照) を見ると、左足を寄せる動作の終了時の足の向きが異なることが分かる。このような特徴について、各軸それぞれの加速度ノルムが最大となるタイミングを取得することで、「どちらの足が、いつ、どの方向に動いたか」をとらえることができたと考えられる。

5.2.3 骨格位置を用いた特徴量の評価

次に表 5 にまとめられた骨格位置情報を基にした特徴量について、最も分類に寄与したのは左手右手の y 座標、すなわち高さの最大値のインデックスであった。今回の骨格位置座標は一方のビデオデータを用いて取得されたため、体の中心に近い部位の細かな座標変化が隠れやすくなった一方で、フィガーごとの手の高さ変化は撮影方向に影響されづらく重要度が高くなったと考えられる。また、リーダー役 (主に男性) の左手はリードアームと呼ばれ、女性の動作を導き、また補助する役目を担っている。そのため、左手の y 座標が最大となるタイミングを抽出することで、女性が男性の腕の下をくぐる動作を含むフィガーにおいて「どのタイミングで女性が腕の下をくぐるように導かれたか」という情報を取得できたと考えられる。一方、右手はフリーアームと呼ばれ、動かし方の定義は明確にされておらず、音楽に合わせて自然に動かすことが好ましい

表 4 加速度・角速度特徴量の重要度

Table 4 Feature importance of acceleration and angular velocity.

重要度 (順位)	特徴量
1	y 軸角速度の平均値
2	x 軸加速度最大値のインデックス
3	y 軸加速度最大値のインデックス
4	z 軸加速度最大値のインデックス
5	x 軸角速度の分散値
6	y 軸角速度の 1 次近似の係数
7	y 軸加速度の 1 次近似の係数
8	y 軸角速度の平均値
9	z 軸角速度の平均値
10	y 軸角速度の 1 次近似の係数

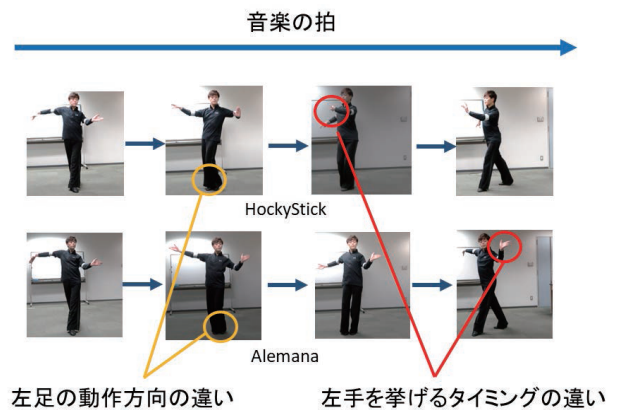


図 10 重要な特徴量の例

Fig. 10 Examples of the key feature values.

とされている [8]. 今回右手首の高さの最大値が認識に寄与した理由として、(1) 動作方法が明確に定義されていないフリーアーム動作において、各フィガーの潜在的動作特性を取得した (2) データ収集に参加した踊り手が同じダンス団体の出身であったため、動作傾向が類似していたの 2 種類が考えられるが、現状では (2) のデータ出自の偏りによる影響を否定することが難しい。右手の高さ変化の認識に

表 5 骨格位置の特徴量の重要度

Table 5 Feature importance of the body joints.

重要度 (順位)	特徴量
1	左手 y 座標の最大値のインデックス
2	右手 y 座標の最大値のインデックス
3	鼻 x 座標のフィガー固有部分の分散値
4	右膝 y 座標のフィガー固有部分の分散値
5	左右肩の距離の分散値
6	左足親指 x 座標のフィガー固有部分の平均値
7	肩から見た左肘の角度の分散値
8	左耳 y 座標の後フィガー遷移部分の平均値
9	右肩 x 座標のフィガー固有部分の平均値
10	右肩 y 座標の前フィガー遷移部分の平均値

おける有効性を確認するため、今後複数の団体に協力を打診しデータの多様性を充実させる必要がある。続いて鼻の x 軸方向の変化の分散が大きく寄与したのは、体の各部位が首からの相対座標で表されていることから顔の向き変化がフィガー分類において重要であったことを表している。

そのほか、全体的に足や顔の部位による特徴量が重要であった。手首の高さと同様に、RGB カメラにより 2 次元化されたデータにおいて体の末端や顔など比較的動きが現れやすい部位が分類に寄与したと考えられる。また、4.1.1 項で述べた連続フィガーにおける動作変化を抽出した特徴量では、「現在のフィガー固有部分」が重要度の高い特徴量として多くあらわれた一方、遷移部分の特徴量はあまり現れなかった。遷移部分と固有部分のそれぞれからフィガー特性を抽出できると予想していたが、遷移部分は分類に貢献せず、一方で遷移部分を排除されたフィガー固有部分が、本来のフィガーごとの特徴をよく表したと考えられる。

### 5.3 骨格位置の一部を欠損させた分類での評価

次に、骨格位置の一部を欠損させた場合の分類精度を表 6、表 7 にまとめる。欠損させる条件は、社交ダンスが男女 2 名で踊られることから縦方向に障害物（人物）が発生すると仮定し、体の右半分または左半分、および右 3/4 と左 3/4 を欠損させた状態で分類を行った。ここで、左半分とは各動画フレームにおける首の x 座標を基準に、それより負の側に位置している体の部位のことである。また左 3/4 とは、各フレームにおいて体の全部を x 座標により並べ替え、小さいほうから 18 カ所を欠損させたものである。

結果を表 6、表 7 より確認すると、ベースライン手法および提案手法のそれぞれについて、各部位を欠損させることで骨格位置による分類精度が大きく低下しているのに対して、加速度・角速度とのハイブリッドによる分類精度では低下の度合いが小さいことが確認される。また、体の半分が欠損している場合骨格位置情報がハイブリッド分類の精度に若干寄与した一方、体の 3/4 が欠損している場合に

表 6 ベースラインの欠損による比較

Table 6 Comparison of the baseline results with occlusion.

精度 (F 値)			
欠損	左足首の加速度・角速度	骨格位置	ハイブリッド
なし		0.74	0.91
右半身欠損		0.71	0.84
左半身欠損	0.84	0.72	0.84
右 3/4 欠損		0.64	0.84
左 3/4 欠損		0.61	0.84

表 7 提案手法の欠損による比較

Table 7 Comparison of the proposed method with occlusion.

精度 (F 値)			
欠損	左足首の加速度・角速度	骨格位置	ハイブリッド
なし		0.85	0.97
右半身欠損		0.78	0.94
左半身欠損	0.93	0.77	0.94
右 3/4 欠損		0.63	0.93
左 3/4 欠損		0.59	0.93

つについては骨格位置情報による特徴量にかかわらず分類精度が変化しないことも確認されたことから、現在の提案手法における骨格位置に基づく特徴量抽出のみでは、遮蔽物に対する頑健性を確保できないことが分かる。また、ベースライン手法と提案手法の各ハイブリッド認識精度を比べると、提案手法では各欠損状態における精度が 0.06 から 0.09 程度高く、マルチモーダルセンサによるハイブリッド手法だけでなく提案手法における特徴量設計、特に加速度・角速度に対する特徴量の工夫が遮蔽物に対する認識の頑健性に寄与できることが分かる。今後実際に 2 人 1 組で踊っているデータを収集し、体の部位の欠損がどの程度起こりうるのかを確認し、2 人 1 組のデータで活用するため特徴抽出をさらに改良する必要がある。

## 6. まとめ

本研究では社交ダンスにおける、加速度・角速度、骨格位置を用いたダンスフィガー分類手法を提案した。はじめに、ウェアラブルセンサの加速度・角速度センサの信号データ、および画像データから得た骨格位置座標の時間変化を入手し、基本的な統計量を用いた行動認識のベースライン手法によりダンスフィガーの分類を行った。その結果、分類精度として加速度・角速度を用いた分類で F 値 0.84、骨格位置を用いた分類で 0.74、両者のハイブリッドで 0.91 の精度を得た。次にフットワークタイミングや腕の角度、高さといったダンスフィガーを特徴づける動作特性を分析したうえで、特徴量を設計し性能を評価した。結果として加速度・角速度および骨格位置それぞれで F 値 0.93、0.85、さらにハイブリッド手法で 0.97 の精度を得、いずれもベースライン手法を上回った。また社交ダンスが 2 人で踊るスポーツであることを考慮し、視覚情報の一部を欠損させた

場合の精度も算出したところ F 値 0.93 の精度を確認し、マルチモーダルセンサを用いることで提案手法が遮蔽物に対して一定の頑健さを持つことを示した。

今回の検証では分類器にランダムフォレストを用いたが、他の分類器や深層学習を用いた手法などで同様の分類を行い、分類器の検証が必要である。また設計した特徴量の有効性については、今後データ量をさらに拡充して確認する必要がある。ただし、ラテンアメリカンだけでも 5 種目、それぞれに数十種類またはそれ以上のフィガーが存在し、各種目の各フィガーについて潤沢な量のデータを収集することは容易ではない。そのため、データ量の確保が困難な状態でも認識精度を向上させられる提案手法は十分に有用であると考えられる。今後はダンスフィガーの種類と踊り手の数、また各フィガーデータ自体を増量し、十分な量のデータに対しても提案手法が有効となるか、もしくはデータ量が限定された環境においてのみ提案手法が有効となるのか検証を行う必要がある。

音楽性に関して、本研究は現在、フィガーの指針に則った動作改善を目指している。今後、タイミングを外して踊る場合のように、ユーザの個性に対して頑強な認識を行うには、各フィガーにおいて行われることの多いアレンジを考慮する必要がある。音楽リズムについては、社交ダンスにおける音楽の刻み幅は最小で 1/4 拍までが使用されることが多いため、1/4 拍を最小のタイムウィンドウとし、フィガーの遷移状態に応じたセグメンテーション方法の開発が必要である。また、本研究は現在ルンバに着目した分析を行っているが、将来的には他のラテンアメリカン種目や、社交ダンスのもう 1 つのジャンルであるスタンダードにも応用が期待され、提案手法の応用性について検証が必要である。ラテンアメリカンでは、音楽リズムの特性やフィガーに含まれる音楽の拍数は種目により変化するが、フィガーの遷移に関する特性はある程度共通しており、提案した特徴量設計の方針はラテンアメリカンの他の種目にも応用可能だと考えられる。一方、スタンダードについては、動作の傾向や音楽的特性、フィガーの遷移にともなう動作変化などがラテンアメリカンと異なる部分も多く、分析・検討が必要である。

**謝辞** 本研究の一部は科研費 JP17H01762, および JST CREST JPMJCR1882 の援助を受けたものです。また、データ収集には、名古屋大学競技ダンス部の部員および卒部生の方々のご協力をいただきました。

## 参考文献

- [1] 海野 敏, 高橋あゆみ, 小山久美: 日本のバレエ学習人口とバレエ参加率に関する大規模社会調査の比較分析, 東洋大学社会学部紀要, Vol.50, No.1, pp.51–65 (2012).
- [2] Merom, D., Cumming, R., Mathieu, E., Anstey, K.J., Rissel, C., Simpson, J.M., Morton, R.L., Cerin, E., Sherrington, C. and Lord, S.R.: Can Social Dancing Prevent Falls in Older Adults? A Protocol of the Dance, Aging, Cognition, Economics (DAnCE) Fall Prevention Randomised Controlled Trial, *BMC Public Health*, Vol.13, No.1, p.477 (online), DOI: 10.1186/1471-2458-13-477 (2013).
- [3] Yamauchi, M., Shinomoto, R., Nishiwaki, E., Onozawa, R. and Kitahara, T.: Development of Dance Training Support System Using Kinect and Wireless Mouse, *The Symposium of Entertainment Computing*, Vol.2013, pp.332–338 (2013) (online), available from (<https://ci.nii.ac.jp/naid/170000078924/>).
- [4] Narazani, M., Seaborn, K., Hiyama, A. and Inami, M.: StepSync: Wearable Skill Transfer System for Real-time Foot-Based Interaction (2018).
- [5] 黄 宏軒, 上亟正樹, 関 優樹, 李 周浩, 川越恭二: 仮想社交ダンスインストラクターの構築, 人工知能学会論文誌, Vol.28, No.2, pp.187–196 (オンライン), DOI: 10.1527/tjsai.28.187 (2013).
- [6] BallroomGuide: BALLROOM GUIDE (online), available from (<http://www.ballroomguide.com/index.html>) (accessed 2019-12-21).
- [7] 日本ボールルームダンス連盟: ラテンアメリカンルンバ, 日本ボールルームダンス連盟 (1999).
- [8] Hearn, G.: *A Technique of Advanced Latin-American Figures*, Japan Ballroom Dance Federation (2013).
- [9] Cao, Z., Simon, T., Wei, S.-E. and Sheikh, Y.: Real-time Multi-Person 2D Pose Estimation using Part Affinity Fields, *IEEE Conference on Computer Vision and Pattern Recognition* (2017).
- [10] 村尾和哉, 寺田 努: 加速度センサの定常性判定による動作認識手法, 情報処理学会論文誌, Vol.52, No.6, pp.1968–1979 (2011).
- [11] Su, X., Tong, H. and Ji, P.: Activity recognition with smartphone sensors, *Tsinghua Science and Technology*, Vol.19, No.3, pp.235–249 (online), DOI: 10.1109/TST.2014.6838194 (2014).
- [12] Anjum, A. and Ilyas, M.U.: Activity recognition using smartphone sensors, *2013 IEEE 10th Consumer Communications and Networking Conference (CCNC)*, pp.914–919 (online), DOI: 10.1109/CCNC.2013.6488584 (2013).
- [13] Wang, Y., Cang, S. and Yu, H.: A survey on wearable sensor modality centred human activity recognition in health care, *Expert Systems with Applications*, Vol.137, pp.167–190 (online), DOI: <https://doi.org/10.1016/j.eswa.2019.04.057> (2019).
- [14] Maurer, U., Rowe, A., Smailagic, A. and Siewiorek, D.P.: eWatch: A wearable sensor and notification platform, *International Workshop on Wearable and Implantable Body Sensor Networks (BSN'06)*, pp.4–145 (online), DOI: 10.1109/BSN.2006.24 (2006).
- [15] Sun, L., Zhang, D., Li, B., Guo, B. and Li, S.: Activity Recognition on an Accelerometer Embedded Mobile Phone with Varying Positions and Orientations, *Ubiquitous Intelligence and Computing*, Yu, Z., Liscano, R., Chen, G., Zhang, D. and Zhou, X. (Eds.), pp.548–562, Springer Berlin Heidelberg (2010).
- [16] Shoaib, M., Bosch, S., Incel, O.D., Scholten, H. and Havinga, P.J.M.: Complex Human Activity Recognition Using Smartphone and Wrist-Worn Motion Sensors, *Sensors*, Vol.16, No.4 (online), DOI: 10.3390/s16040426 (2016).
- [17] Alp Guler, R., Neverova, N. and Kokkinos, I.: DensePose: Dense Human Pose Estimation in the Wild, *The IEEE Conference on Computer Vision and Pattern*

- Recognition (CVPR)* (2018).
- [18] Bearman, A.L. and Dong, C.: Human Pose Estimation and Activity Classification Using Convolutional Neural Networks (2015).
- [19] Ehatisham-Ul-Haq, M., Javed, A., Azam, M.A., Malik, H.M.A., Irtaza, A., Lee, I.H. and Mahmood, M.T.: Robust Human Activity Recognition Using Multimodal Feature-Level Fusion, *IEEE Access*, Vol.7, pp.60736-60751 (online), DOI: 10.1109/ACCESS.2019.2913393 (2019).
- [20] Chen, C., Jafari, R. and Kehtarnavaz, N.: UTD-MHAD: A multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor, *2015 IEEE International Conference on Image Processing (ICIP)*, pp.168-172 (online), DOI: 10.1109/ICIP.2015.7350781 (2015).
- [21] Chen, C., Jafari, R. and Kehtarnavaz, N.: A Real-Time Human Action Recognition System Using Depth and Inertial Sensor Fusion, *IEEE Sensors Journal*, Vol.16, No.3, pp.773-781 (online), DOI: 10.1109/JSEN.2015.2487358 (2016).
- [22] Hwang, I., Cha, G. and Oh, S.: Multi-modal human action recognition using deep neural networks fusing image and inertial sensor data, *2017 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, pp.278-283 (online), DOI: 10.1109/MFI.2017.8170441 (2017).
- [23] Dawar, N., Ostadabbas, S. and Kehtarnavaz, N.: Data Augmentation in Deep Learning-Based Fusion of Depth and Inertial Sensing for Action Recognition, *IEEE Sensors Letters*, Vol.3, No.1, pp.1-4 (online), DOI: 10.1109/LSENS.2018.2878572 (2019).
- [24] Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E. and Sheikh, Y.: OpenPose: Realtime Multi-person 2D Pose Estimation using Part Affinity Fields, arXiv preprint arXiv:1812.08008 (2018).
- [25] Sigal, L., Balan, A. and Black, M.J.: HumanEva: Synchronized Video and Motion Capture Dataset and Baseline Algorithm for Evaluation of Articulated Human Motion, *International Journal of Computer Vision*, Vol.87, No.1, pp.4-27 (2010).
- [26] Matsuyama, H., Hiroi, K., Kaji, K., Yonezawa, T. and Kawaguchi, N.: Ballroom Dance Step Type Recognition by Random Forest Using Video and Wearable Sensor, *International Workshop on Human Activity Sensing Corpus and Application* (2019).



松山 仁 (学生会員)

2018年名古屋大学工学部電気電子・情報工学科卒業。2020年同大学大学院修士課程修了。行動センシングの研究に従事。



浦野 健太 (学生会員)

2016年名古屋大学工学部電気電子・情報工学科卒業。2018年同大学大学院修士課程修了。IoT機器を用いた屋内位置推定に関する研究に従事。



廣井 慧 (正会員)

2004年東北大学工学部電子工学専攻卒業。同年東日本電信電話株式会社入社。2014年慶應義塾大学大学院メディアデザイン研究科博士(メディアデザイン学)。名古屋大学未来社会創造機構特任助教, 同大学工学研究科助教を経て, 2020年から京都大学防災研究所准教授。防災情報システム, 災害情報の時空間解析の研究に従事。



梶 克彦 (正会員)

2002年名古屋大学工学部電気電子工学科卒業。2007年同大学大学院情報科学研究科博士課程修了。博士(情報科学)。NTTコミュニケーション科学基礎研究所リサーチアソシエイト, 名古屋大学大学院工学研究科助教を経て, 2015年より愛知工業大学情報科学部准教授。日本ソフトウェア科学会会員。屋内位置推定, 遠隔コミュニケーション支援の研究に従事。



米澤 拓郎 (正会員)

2010年慶應義塾大学大学院政策・メディア研究科後期課程博士号取得後, 同大学院特任助教, 特任講師, 特任准教授を経て, 2019年より名古屋大学大学院工学研究科准教授。主に, ユビキタスコンピューティングシステム, ヒューマンコンピュータインタラクション, センサネットワーク等の研究に従事。ACM, 各会員。



河口 信夫 (正会員)

1990年名古屋大学工学部電気電子工学科卒業。1995年同大学大学院工学研究科情報工学専攻博士課程満了。同年同大学工学部助手，同大学講師，准教授を経て，2009年より同大学大学院研究科教授。NPO 位置推定サービス

研究機構 Lisra 代表理事。モバイルコミュニケーション，ユビキタスコンピューティング，行動センシングの研究に従事。博士（工学）。ACM, IEEE, 人工知能学会，日本ソフトウェア科学会，電子情報通信学会，日本音響学会各会員。本会シニア会員。