

AIST Dance Video Database: ダンス情報処理研究のためのストリートダンス動画データベース

土田 修平¹ 深山 覚¹ 濱崎 雅弘¹ 後藤 真孝¹

概要: 本稿では、研究目的で共通に利用できるストリートダンス動画データベース「AIST Dance Video Database (AIST Dance DB / AIST ダンス動画データベース)」について述べる。本研究用 DB には、新規に制作したダンス音楽 60 曲と、それに合わせて独自の振り付けでストリートダンスを踊っているダンサーを撮影した動画 13,940 本が収録されている。ダンスと音楽は密接な関わりを持ち、音楽情報処理分野においてもさまざまなダンス関連のテーマが研究されてきた。しかし、ダンス情報処理が持つ潜在的な可能性の大きさに対して、まだ研究事例が多いとはいえなかった。そこで我々は、ダンスのジャンルの一つとして重要なストリートダンスに焦点を当てて、ダンス情報処理研究の基盤となることを目指して本 DB を構築し、2019 年 11 月に一般公開を開始した。本稿ではさらに、本 DB によって可能になる研究テーマの事例を提案し、特にダンスモーションのジャンルを分類するタスクについては、4 つのベースライン手法を開発・評価した結果を報告する。

1. はじめに

音楽情報処理分野では、音楽の音響信号や楽譜、歌詞だけでなく、音楽に関連した動画や画像、ソーシャルメディアのデータなど、多様でマルチモーダルな情報が既に扱われてきている。音楽情報処理で対象とされてきたダンス楽曲は、元々ダンスのために制作されており、ダンスモーション（ダンスの動作）と密接に関連している。そうしたダンス楽曲やダンスモーション、ダンサーの属性等のダンスに関連した情報（以下、ダンス情報）は多様であり、それらを扱う研究は、音楽情報処理分野において今後ますます活発になっていくと考えられる。

そこで、そうしたダンス情報を対象とした多様な技術・研究テーマをカバーする研究分野名として、「**ダンス情報処理 (Dance Information Processing)**」[1] を提唱する。これまで音楽情報処理研究の一環として、ダンス楽曲に関連したさまざまな論文が発表されてきた。伝統的なダンス音楽 [2-5] やエレクトロニック・ダンス・ミュージック (EDM) [6-12] に関する研究、さらにはそうしたダンス楽曲を分類する研究 [13, 14] などがある。これらの研究もダンス情報処理に位置付けられる。一方、ダンスモーションも活発に研究されており、楽曲に基づいたダンスモーションの自動生成 [15-21]、ダンス動画上でのダンスモーショ



図 1 AIST Dance Video Database (AIST Dance DB) の収録ダンス動画例。複数のジャンル（ストリートダンスで主要な 10 種類のジャンル）や、複数のダンサー（35 名のプロダンサーによるソロダンスおよびグループダンス）、複数のカメラ（最大 9 台のビデオカメラをダンサーを囲むように設置）によるダンス動画を収録している点が特徴的である。

ンからのリズム推定 [22]、ダンスモーションによる楽曲検索 [23]、ダンスモーションによる楽曲テンポ制御 [24]、ダンスモーションの個人識別 [25]、ダンスモーションのジャ

¹ 産業技術総合研究所
National Institute of Advanced Industrial Science and Technology (AIST)

ンル分類 [26] などが挙げられる。さらに、ダンスモーションに関する研究は、音楽に関連した人間の身体の動きを扱うので、演奏中におこなわれる演奏者の動作に関する研究 [27–31] とも関連深い。しかし、こうしたダンス情報処理研究を進展させる上で、世界中の研究者が自由に利用可能な大規模なストリートダンス動画データベース (DB) は存在しなかった。

そこで我々は、ストリートダンスに焦点を当て、研究者が研究目的に使用する上で、共通利用の自由、学術利用の自由が確保された「AIST Dance Video Database (AIST Dance DB / AIST ダンス動画データベース)」(図 1) を構築した [1]。本 DB は、ストリートダンスで主要な 10 種類のジャンル (ブレイク、ポップ、ロック、ワック、ミドルヒップホップ、LA スタイルヒップホップ、ハウス、クランプ、ストリートジャズ、バレエジャズ) をカバーした 13,940 本のダンス動画と、それぞれのダンスジャンルに合わせて作曲された 60 曲のダンス楽曲によって構成されている。10 種類のそれぞれのダンスジャンルに対して、異なるテンポのダンス楽曲が 6 曲ずつ収録されており、5 年以上のダンス経験を有する 35 名のプロダンサー (男性 20 名、女性 15 名) がそれら楽曲に合わせて踊っている。各ダンスジャンルごとに 1,389 本の動画を収録した合計 13,888 本^{*1}の動画に加え、ダンサーのグループが一緒に踊る際の三つの典型的なシチュエーション (ショーケース、サイファー、バトル) の下でのダンス動画を 52 本収録している。収録動画の多様性を考慮して設計しており、それらをダンサーの人数で分類すると、計 13,940 本のうち、12,988 本はダンサーが 1 人で踊るソロダンスで、952 本は複数のダンサーと一緒に踊るグループダンスとなっている。振り付けの種類で分類すると、計 13,940 本のうち、10,798 本は基本的な振り付けで構成された基本動作で、3,142 本は各ダンサーが独自に振り付けをしたフリーダンスとなっている。また、最大 9 台のビデオカメラをダンサーを囲むように配置し、異なる方向から同時にダンスを撮影した。

我々は 2019 年 11 月に AIST Dance DB の一般公開を開始し [1]、学術研究目的であれば無償で <https://aistdancedb.ongaaccel.jp> からすべてのダンス動画とダンス楽曲をダウンロード可能となっている。これにより、ダンス情報処理の研究分野が立ち上がり、様々なテーマ・タスクに関する研究が促進されることを期待している。例えば、ダンスモーションのジャンル分類タスクや、ダンサー識別タスク、ダンステクニクの推定タスクなどに活用できる。そうしたダンス情報処理における基本タスクの一つとして、本稿では、ストリートダンスのためのダンスモーションジャンル分類タスクを提案する。我々は、10 種類のダンスジャンルからそれぞれ 3 名ずつ、合計 30 名の

ダンサーによって独自に振り付けされた 210 本のダンス動画を、本タスクの実験用のデータとした。それらのダンス動画から抽出したダンスモーションを入力を与えて、Long Short-Term Memory (LSTM) に基づくモデルと、Support Vector Machine (SVM) に基づくモデルを用いて構成した分類器を学習し、評価した。ダンスジャンルを分類する 4 つのベースライン手法を開発して比較評価した結果、32 秒の長さで切り出した動画を与えたときに、91.4% の精度でダンスジャンルを分類できることを確認した。しかし、その動画の長さを 0.67 秒に短くすると、精度は 56.6% まで低下した。この結果は、今後の同タスクのベースライン性能として他の研究者が比較評価に用いることができる。

2. 関連研究



学術研究における研究用 DB の重要性は、さまざまな研究分野で広く認識されてきた。ダンスに関連した研究においても同様で、DB の構築にダンス関連の研究者が大きな労力を費やしてきた [32]。例えば、「Martial Arts, Dancing and Sports Dataset (MADS)」[11] では、ヒップホップとジャズダンスのダンス動画がそれぞれ 6 本ずつ収録されている。3 台のカメラで 3 方向から撮影しており、動画データには深度カメラ (Kinect を利用) から得られた奥行き情報も付与されている。しかし、各ジャンルの動画は、ダンスモーションのジャンル分類や振り付け自動生成といったタスクに対して十分な本数ではなく、各ジャンルの特性を効果的に表現するための十分な量のダンスモーションを含んでいなかった。また、小規模なダンスデータセット [33–36] も構築されてきた。

ダンス動画だけでなく、異なる種類のセンサデータを提供する DB も公開されている。Stavrakis ら [37] は、ダンス動画に加えて高品質のモーションキャプチャデータを提供する「Dance Motion Capture Database」を公開している。この DB には、ギリシャ・キプロスの伝統的なダンスの他に、モダンダンス、フラメンコ、ベリーダンス、サルサ、ヒップホップといったダンスジャンルも含まれている。同様に Tang ら [38] は、モーションキャプチャシステムを用いて、チャチャ、タンゴ、ルンバ、ワルツの 4 つのダンスジャンルを含む DB を構築した。Essid ら [39] は、1 台の Kinect と 5 台のビデオカメラを用いて、15 名のサルサダンサーが 2~5 種類の事前に決められた振り付けで踊るダンスを撮影し、その動画データセットを公開している。さらに、ダンサーが身に付けた複数の慣性センサから取得されたデータ (加速度、角速度、磁気) が、すべての動画データに付属している点が特徴的である。上記のように従来さまざまな DB が公開されてきたが、ストリートダンスの多様なジャンルをカバーし、複数のカメラで撮影され、かつ、複数のダンサーが踊っているダンス動画を含む DB は存在しなかった。






^{*1} 1,389 本 x 10 ジャンル = 13,890 本となる設計だったが、撮影時の不備で 2 本欠損して 13,888 本となっている。

AIST Dance DB 10 ダンスジャンル : 1,389 本 (1,080+189+90+30) × 10 ジャンル

動画 : 13,940 本
ダンス : 1,618
楽曲数 : 60 曲
ダンスジャンル : 10
ダンサー : 35 名
男性 20 名
女性 15 名
カメラ台数 : 最大 9 台
総収録時間 : 118.1 時間

基本ダンス : 1,080 本 (3×10×4×9)		グループダンス : 90 本 (1×10×9)		バレエジャズ ストリートジャズ クランプ ハウス LAスタイルヒップホップ ミドルヒップホップ ワック ロック ポップ ブレイク		
	ダンサー	3			グループ (2名)	1
	振り付け	1人あたり 10 ダンス			振り付け	1人あたり 10 ダンス
	振り付けタイプ	4			カメラ	9
	カメラ	9			長さ	64 拍, 平均 52 秒
	長さ	16 拍, 平均 23 秒				
フリーダンス : 189 本 (3×7×9)		移動カメラ : 30 本 ((2×3+1×4)×3)				
	ダンサー	3		ダンサー	3	
	振り付け	1人あたり 7 ダンス		振り付け	1人あたり 3 or 4 ダンス	
	カメラ	9		カメラ	移動 1 & 固定 2	
	長さ	64 拍, 平均 52 秒		長さ	64 拍, 平均 54 秒	

シチュエーション : 52 本

ショーケース : 27 本 (1×3×9)		サイファー : 10 本 (1×2×5)		バトル : 15 本 (3×1×5)	
	グループ (10名)	1		グループ (10名)	1
	振り付け	1グループあたり 3 ダンス		セット	2
	カメラ	9		カメラ	5
	長さ	96 拍, 平均 75 秒		長さ	1 動画あたり 10 分
				グループ (2名)	3
				セット	1
				カメラ	5
				長さ	1 動画あたり 4 分

全 10 種類のダンスジャンルのそれぞれにおいて、以下の 4 つのカテゴリに分類できる 1,389 本の動画が収録されている。

- ・基本ダンス: 120 種類のダンスを 9 方向から撮影した動画 1,080 本。
ジャンルごとと基本的な動きを 4 つの異なる印象 (激しい, 緩い, 固い, 柔らかい) で表現。
 - ・フリーダンス: 各ダンサーが独自に振り付けした 21 種類のダンスを 9 方向から撮影した動画 189 本。
 - ・グループダンス: 3 名のダンサーが独自の振り付けを一緒に踊った 10 種類のグループダンスを 9 方向から撮影した動画 90 本。
 - ・移動カメラ: 10 種類のダンスをカメラを前後に移動しながら (ドリーイン, ドリーアウトして) 3 方向から撮影した動画 30 本。
- その他、以下の三つのシチュエーションの下で撮影された 52 本の動画が収録されている。
- ・ショーケース: 10 名のダンサーがステージでのダンスパフォーマンスを観客の前で披露することを想定して踊った 3 種類のグループダンスを 9 方向から撮影した動画 27 本。
 - ・サイファー: 10 名のダンサーが輪になって並び、ランダムな順番で踊り続ける二つのセットを 5 方向から撮影した動画 10 本。
 - ・バトル: 2 名のダンサーが向かい合って踊る三つのセットを 5 方向から撮影した動画 15 本。

図 2 AIST Dance DB の全体構成。

「YouTube-8M」[40] と「Music Video Dataset」[41] はダンス研究用に構築された DB ではないが、前者には 181,579 本のダンス動画、後者にはダンスに焦点を当てた動画を含む 1,600 本のミュージックビデオが含まれている。これらの DB は、ダンス情報処理の観点で整理された DB ではないため、ダンス情報処理の関連タスクに取り組む目的でそのまま利用することは難しい。それに対して AIST Dance DB では、ダンスジャンル名、基本的な振り付けの種類、シチュエーション名、ダンサーや楽曲の情報などの属性を持つように、体系的に記録・整理した動画が収録されている特長を持つ。

3. AIST Dance Video Database (AIST Dance DB)

3.1 設計指針

我々は、以下の三つを重点的に考慮して AIST Dance DB を設計した。

- ・独自に振り付けされたダンスの動画ファイルと新規に制作したダンス楽曲ファイル
ダンス楽曲と、その楽曲に合わせて踊ったダンスモー

ションとの関係を分析・調査できるような DB にすることが重要である。そこでまず、プロの音楽家が、それぞれのダンスジャンルの踊りに合ったダンス楽曲を新規に制作した。次に、その楽曲をプロのダンサーが聞きながら、それに合わせて踊っているところを撮影した。その結果、各動画には、ダンスモーションが映像に含まれているだけでなく、使用されているダンス楽曲も背景音楽 (BGM) として含まれている。また、各ダンス楽曲は個別のファイルとしても収録されている。

- ・ダンスジャンルや振り付けの多様性
ダンス情報処理の多様なテーマで活用できるようにするには、多様なダンスジャンルや振り付けを含むことが重要である。そこで、10 種類のダンスジャンル、35 名のダンサー、異なる人数のダンサーによるダンス (ソロダンスとグループダンス)、異なる振り付け、異なる振り付け難易度 (基本ダンスとフリーダンス) のような多様なダンスが収録されるようにした。
- ・さまざまな方向からの撮影
異なる視点と角度から同じダンスモーションを分析するために、ダンサーを囲むように複数のカメラを配置

し、最大で9種類の方向から同時に撮影した。これにより、たとえ正面のカメラからは捉えられない身体の動きがあっても、背面のカメラから捉えられる。

3.2 データベースの収録内容

3.1節の設計指針に基づいて構築した「AIST Dance DB」の全体構成を図2に示す。13,940本の動画には、1,618種類のストリートダンスが収録されている。そのうち、13,888本は10種類のストリートダンスジャンルの動画で、残りの52本は三つのシチュエーションの下で撮影された動画である。10種類のダンスジャンルは、ストリートダンスの専門家と相談して決定した。ダンスジャンルは、1970年頃から1990年頃までのダンススタイルであるオールドスクール系（ブレイク、ポップ、ロック、ワック）と、1990年頃以降のダンススタイルであるニュースクール系（ミドルヒップホップ、LAスタイルヒップホップ、ハウス、クランプ、ストリートジャズ、バレエジャズ）に分けられる。

ダンス楽曲60曲は、それぞれ6曲ずつ、10種類のダンスジャンルに対応している。各ジャンルにおいて、ハウスを除く6曲のテンポは、80, 90, 100, 110, 120, 130 BPM（1分間の四分音符の個数）に設定した。ハウスでは遅いテンポだとダンスの動作に合わないため、6曲のテンポは、110, 115, 120, 125, 130, 135 BPMに設定した。

一つのダンスジャンル内で多様な振り付けを確保するため、プロのダンサー35名（男性20名、女性15名）の踊りを撮影した。各ジャンルに対して、少なくとも3名のダンサーが踊った。すべてのダンサーは5年以上のダンス経験を有する。また、すべての動画はフルカラーで収録されたが、服装の選択をダンサーに委ねたところ、モノトーンの服を着ていたダンサーが多かった。

図2上部に示したように、ダンスジャンル一つあたり、1名のダンサーが基本的な振り付けで踊っている基本ダンス動画が1,080本、1名のダンサーが独自の振り付けで踊っているフリーダンス動画が189本、複数のダンサーが一緒に独自の振り付けを踊っているグループダンス動画が90本、1名のダンサーによるダンスをカメラを移動させながら撮影した移動カメラ動画が30本の、合計1,389本の動画が収録されている。移動カメラ動画以外のすべてのカメラ位置は固定されている。それらに加えて、図2下部のように、ダンサーの集団がダンスを楽しむ際の3種類のシチュエーションを想定して撮影した、ショーケース動画が27本、サイファー動画が10本、バトル動画が15本の合計52本の動画が収録されている。

ダンサーが振り付けを考えて踊る際には、ダンスジャンルに合った振り付けをすることとした。それぞれの振り付けに対して撮影は1回を原則としたが、明らかな間違いが起きたら再撮影をした。撮影では、ダンサーの全身が映像内に収まるように正面のカメラを慎重に配置した。正面以

外のカメラも、グループダンスでないときには、ダンサーの全身が極力映像内に収まるように配置した。

4. ダンスモーションジャンル分類

本章では、我々がダンス情報処理研究のタスクの一つとして提案するダンスモーションジャンル分類タスクについて述べる。楽曲ジャンルを分類するタスクは、音楽情報処理研究において広く知られたタスクの一つであるが、ダンスモーションジャンル分類タスクは、ダンスモーションのジャンルを分類する新しいタスクである。これは、ダンス動画推薦システムのようなアプリケーションを構築する際に有用であり、ダンス情報処理における標準問題になりえると考えている。

我々は、ダンスモーションジャンル分類タスクのベースラインとなる4つの手法を提案・開発した。ここでの研究上の問いは以下の4つである。

- RQ1: 動画から抽出するフレームのみを用いて、ダンスジャンル分類を行えるか？
- RQ2: ダンスジャンル分類モデルの学習には、何フレーム必要か？
- RQ3: ダンスジャンルによって分類のし易さは異なるか？
- RQ4: 拍位置はダンスジャンル分類の性能向上に役立つか？

4.1 実験条件

実験用データセットとして、AIST Dance DB内の「フリーダンス」カテゴリのダンス動画を用いてサブセットを作成した。正面カメラから撮影した210本のダンス動画から構成されており、10種類のダンスジャンルを網羅している。各ジャンル3名ずつ合計30名のダンサーが、それぞれ7つのフリーダンスを踊っているため、各ジャンルごとに異なる21本のダンス動画がある。各動画の楽曲は64拍（4拍×16小節）で構成されている。なお、1拍の長さは4分音符一つ分とする。

以上の210本のダンス動画を、トレーニングデータ（126本の動画）、バリデーションデータ（14本の動画）、およびテストデータ（70本の動画）に分割した。各ジャンルに対し、2名のダンサーによる14本のダンス動画をトレーニングデータとバリデーションデータに用い、残り1名のダンサーによる7本のダンス動画をテストデータとして用いた。したがってトレーニングデータとバリデーションデータに含まれるダンサーは、テストデータ中には一切含まれていない。

4.2 手法

4つのベースライン手法の概要を図3に示す。入力動画から抽出したフレームを用いて、10種類のダンスジャンル

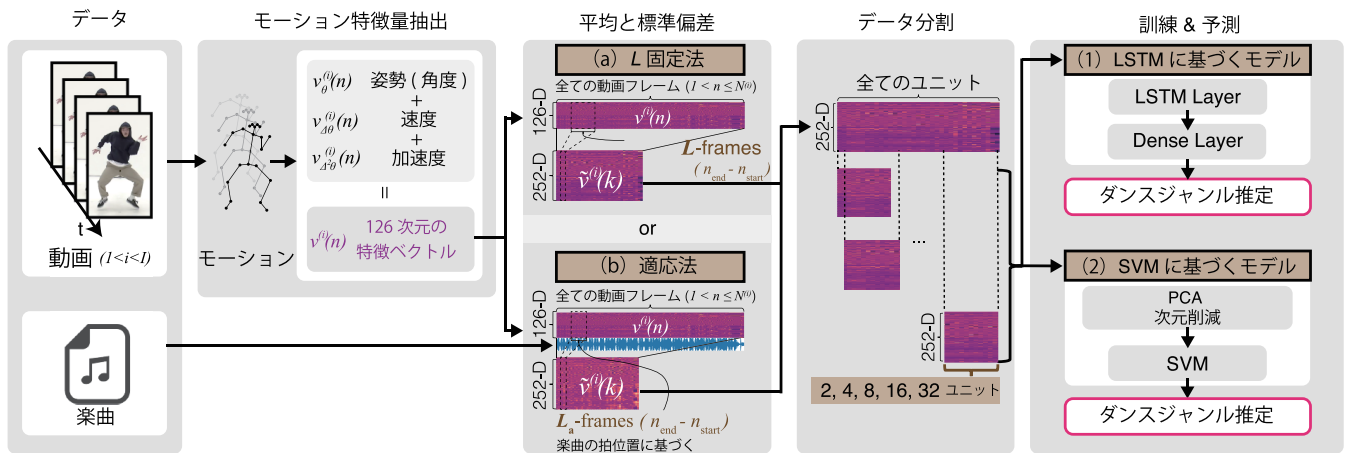


図3 ダンスモーションジャンル分類タスクのための4つのベースライン法の概要: (1-a) LSTMに基づくL固定法, (2-a) SVMに基づくL固定法, (1-b) LSTMに基づく適応法, (2-b) SVMに基づく適応法.

から一つのダンスジャンルに分類するよう各モデルを学習した。モーション特徴量抽出の第1段階では、すべてのフレーム(60 fps)でのダンサーの骨格情報(身体の姿勢と動き)をOpenPoseライブラリ[42]を用いて推定する。これにより、推定された身体の姿勢と動きには元のRGB情報が含まれていないため、我々の分類器のAIST Dance DBへの依存度を低減でき、他の動画にも適用できる可能性が高くなる。

ダンスを特徴付ける要素として、姿勢とモーション(動作)が重要である。そこで、OpenPoseによって推定された骨格情報から得られる21個の関節角度すべてを1フレームごとに計算する。各関節角度は、角度のコサインとサインを計算することで、 θ_x と θ_y の2次元の値へと変換して用いる。これにより、角度間の距離計算が容易になる。結果として、21次元の角度を表すベクトルを42次元の特徴ベクトルに変換する。 i 番目の動画の n 番目のフレームにおける関節角度を42次元の特徴ベクトル $v_{\theta}^{(i)}(n)$ で表す。DBのビデオ総数を I 、 i 番目の動画のフレーム数を $N^{(i)}$ とする($1 \leq i \leq I, 1 \leq n \leq N^{(i)}$)。骨格情報が検出されなかった関節角度は0を代入する。次に、フレーム間の関節角度の変化に焦点を当て、フレーム間の速度と加速度を計算することで身体の動作を表した。速度を $v_{\Delta\theta}^{(i)}(n)$ 、加速度を $v_{\Delta^2\theta}^{(i)}(n)$ と表し、これらを以下の式によって算出する。

$$v_{\Delta\theta}^{(i)}(n) = v_{\theta}^{(i)}(n) - v_{\theta}^{(i)}(n-1), \quad (1)$$

$$v_{\Delta^2\theta}^{(i)}(n) = v_{\Delta\theta}^{(i)}(n) - v_{\Delta\theta}^{(i)}(n-1). \quad (2)$$

以上三つの特徴ベクトル、 $v_{\theta}^{(i)}(n)$ および $v_{\Delta\theta}^{(i)}(n)$ 、 $v_{\Delta^2\theta}^{(i)}(n)$ を、126次元のベクトル $v^{(i)}(n)$ にまとめた。

第2段階として、身体の動作を表す126次元のベクトルを指定された時間間隔(ユニット)ごとに集約する。このとき、拍位置情報を使用する場合(適応法)と使わない場合(L固定法)の二つの手法を用いた。すべての拍位置は、

各楽曲のテンポによって自動的に決定される。以下に二つの手法の詳細を示す。

- 適応法: 拍位置情報を用いる場合、4種類のテンポ依存な可変長ユニット内でベクトルは集約される。1, 2, 3または4つの拍を一つのユニットとして扱った。拍に対応する長さ L_a は、テンポが80から135の範囲の場合、27から45フレームの範囲を表す。
- L固定法: さまざまな固定長ユニット内でベクトルが集約される。一つのユニット長 L は、20, 40, 60, ..., 500動画フレームを持つ。

各手法において、動画内のユニット毎に特徴ベクトルを計算する。各ユニットの n_{start} 番目から n_{end} 番目までの動画フレームの身体の動作を表すベクトルの平均と標準偏差(各126次元)を計算し、それらベクトルを k 番目の252次元のユニット特徴ベクトル $\tilde{v}^{(i)}(k)$ として連結する。

第3段階として、各手法においてウィンドウ特徴ベクトルを計算する。ウィンドウ特徴ベクトルとは、一定のウィンドウの幅における身体の動作の特徴を表すベクトルである。5つの異なるウィンドウ幅を用いて、ユニット特徴ベクトルをウィンドウ特徴ベクトルに集約し、ダンスジャンルを識別するのに必要な動画フレーム数を調べた。ウィンドウ特徴ベクトルは、すべてのユニット特徴ベクトル $\tilde{v}^{(i)}(k)$ を、2, 4, 8, 16, および32ユニットのウィンドウ幅で連結させることでそれぞれ得られる。例えばウィンドウ幅が4の場合、まず $\tilde{v}^{(i)}(k)$ から $\tilde{v}^{(i)}(k+3)$ のユニット特徴ベクトルを連結してウィンドウ特徴ベクトルを得る。次に、ウィンドウを1ユニット分時間軸方向にスライドさせて、 $\tilde{v}^{(i)}(k)$ から $\tilde{v}^{(i)}(k+3)$ のユニット特徴ベクトルを連結させて次のウィンドウ特徴ベクトルを得る。このようにして、5つの異なるウィンドウ幅によって集約されたウィンドウ特徴ベクトルを取得した。

第4段階として、適応法とL固定法の二つを、LSTMに

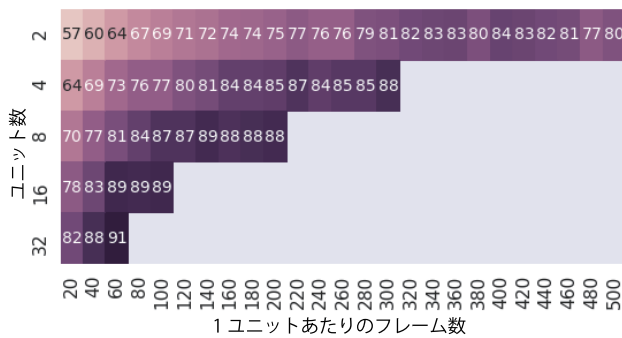


図4 1ユニットあたりのフレーム数とユニット数の異なる組み合わせに関する LSTM に基づく L 固定法を用いたジャンル分類精度の比較。空白は、フレームの合計サイズが動画の長さを超えて結果が得られないことを示す。

基づくモデルと SVM に基づくモデルの二つとそれぞれ組み合わせることで、4つのベースライン手法を準備した。各手法は、すべてのウィンドウ特徴ベクトルを10種類のダンスジャンルに分類する。LSTM に基づくモデルでは、1層の LSTM セル [43] を持つ双方向リカレントニューラルネットワーク (RNN) を用いた。ネットワークは、ダンスジャンルを示す10次元の one-hot ベクトルを出力する。活性化関数 ReLU (Rectified linear unit) を LSTM の出力に適用した。またバッチ正規化を、Dense Layer の出力層に適用した。損失関数にはクロスエントロピーを用い、バッチサイズは10とした。このモデルを100エポックまで 5×10^{-4} の学習率で訓練し、バリデーションロスが最小となったモデルを訓練済みモデルとした。このモデルは、PyTorch [44] で実装した。SVM に基づくモデルでは、最初に主成分分析を用いてトレーニングデータの次元を削減し、200次元のベクトルを得て、SVM モデルを訓練した。最後に、これら二つのモデルを用いて、動画中の各ウィンドウ特徴ベクトルでのダンスジャンルを分類した。

5. 結果

まず、3分割交差検証 (1分割ごとに、テストデータ作成のため、異なるダンサーを振り分ける) を用いて、ダンスモーションジャンル分類の性能を評価した。各ダンスジャンルごとに分類の正解率を求め、その平均値をジャンル分類精度とした。ジャンル分類精度が最も高かったのはフレーム数が60フレームでユニット数が32ユニットの LSTM に基づく L 固定法を用いた場合の91.4%であった。なお、SVM に基づく L 固定法の場合、ジャンル分類精度は最大でも84.0%に留まった。「RQ1: 動画から抽出するフレームのみを用いて、ダンスジャンル分類を行えるか?」に対しては、このデータセットにおいて、ダンスジャンルは比較的高い精度で分類できることがわかった。

次に、1ユニットあたりのフレーム数とユニット数を変えた場合のジャンル分類精度の変化を調べた。結果を図4

に示す。動画の40フレーム (20フレーム / ユニット \times 2ユニット) のみを用いて、56.6%の精度でダンスジャンルを分類できた。これはダンス動画のうちわずか0.67秒間を用いるだけで、これだけのダンスジャンル分類が可能であることを示している。これは我々の予想を超えて短い区間であった。「RQ2: ダンスジャンル分類モデルの学習には、何フレーム必要か?」に対しては、フレーム数が少ないほど分類精度は落ちるが、1秒未満のわずかなダンスモーションだけでも、ある程度のダンスジャンル分類が可能だとわかった。

さらに、LSTM に基づく L 固定法の混合行列を作成してダンスジャンルごとの違いについて分析した。クランプは分類が比較的容易で、ハウスは分類が比較的困難であることがわかった。また分類精度は、分類器への入力を計算するための1ユニットあたりのフレーム数とユニット数に依存することがわかった。フレームとユニットが少ない場合では、分類器がストリートジャズとバレエジャズを混同しやすく、ハウスの分類精度は低下していることがわかった。ストリートジャズとバレエジャズには類似する姿勢が含まれており、ハウスには単純な横移動など、他のダンスジャンルに共通する動きが多く含まれていることが原因と考えられる。「RQ3: ダンスジャンルによって分類のし易さは異なるか?」に対しては、ダンスジャンルによって分類のし易さは異なり、さらにそれはフレーム数とユニット数の影響も受けることがわかった。

最後に、楽曲の拍位置を用いた場合の分類精度について調べた。適応法を用いた場合の最も高いジャンル分類精度は83.4%であり、 L 固定法では91.4%であった。特に SVM に基づく適応法においては、最も高いジャンル分類精度は我々の予想に反して80.7%まで低下した。「RQ4: 拍位置はダンスジャンル分類の性能向上に役立つか?」に対しては、今回はそのような効果を見つけることはできなかった。楽曲の拍位置を用いる場合は改善の余地を残しており、今後の課題とする。

6. 議論

6.1 ダンス情報処理

ダンス情報処理では、未解決で魅力的な課題が多く、研究テーマの宝庫である。図5に示すように、ダンス情報処理は (a) ダンスモーション分析、(b) ダンスモーション生成、(c) ダンス楽曲分析、(d) ダンス楽曲生成の4つのカテゴリに分類できる*2。(a) ダンスモーション分析では、ダンスモーションのジャンル分類、ダンサー識別、ダンステクニック推定、振り付けの構造解析、ダンスのムード解析など、ダンスモーションのあらゆる側面を自動的に分析する研究が考えられる。(b) ダンスモーション生成の典型的な

*2 これらのカテゴリ以外のダンス情報処理もあり得るが、本稿では単純のために4つに分けて議論する。

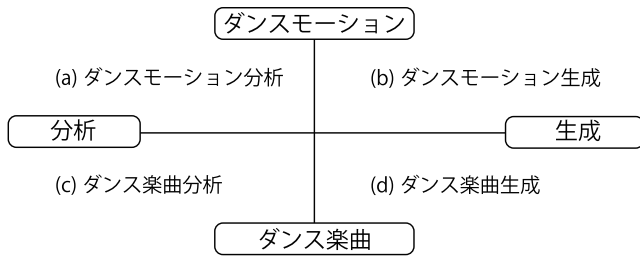


図5 ダンス情報処理の概要

研究課題は、ロボットダンサーやコンピュータグラフィックス(CG)ダンサーなどで活用できるような、人間の動きと区別できない自然なダンスモーションを自動的に生成することである。本稿の冒頭で述べたように、(c)ダンス楽曲分析や(d)ダンス楽曲生成を含むダンス楽曲に関する研究は、音楽情報処理分野で既にいろいろと取り組まれてきた。しかし、ダンス楽曲をより深く分析・分類したり、さまざまなスタイルや目的に合うダンス楽曲を生成したりする技術は、まだ発展できる余地が大きい。

さらに、これら4つのカテゴリを個別に発展させるだけでなく、複数にまたがる研究開発も魅力的である。例えば、(a)ダンスモーション分析と(c)ダンス楽曲分析を組み合わせることで、ダンス動画中の映像側のダンスモーションを分析した結果が、音楽側の構造推定に役立てられる可能性がある。逆に、音楽側の分析結果(拍位置等)は、映像側のダンスモーションの分析に有用である。この(a)と(c)の組み合わせは、ダンスモーションに適した楽曲を見つける際にも重要で、将来的には、ダンスフロアやダンスイベントの会場にいるさまざまなダンサーに適した楽曲を、自動で選曲・再生するDJシステムが実現可能になるかもしれない。ダンス動画中のダンスモーションと音楽との関係が明らかになれば、ダンス動画の制作・編集支援システム等を実現する上で有用である。他にも、(a)ダンスモーション分析と(b)ダンスモーション生成を組み合わせることで、インターネット上の大規模なダンス動画群から3次元のダンスモーションを抽出し、その踊りを機械学習することで、ダンスモーションの自動生成システムを構築できる可能性がある。ダンサーの踊りのスタイルを分析してモデル化できれば、そのスタイルでロボットダンサーやCGダンサーを踊らせることも可能となる。

ダンス情報処理では、マルチモーダルな情報処理が不可欠なテーマも多く、その観点からも魅力的である。例えば、ダンス練習支援[45-54]や、振り付け作成支援[55-57]、ダンスパフォーマンスの拡張[58-69]、グループダンスの支援・解析[70-74]、ダンスのアーカイブ[75-78]、ダンスパフォーマンスの対応付け[79,80]、ダンス動画編集[81-83]、ダンススタイル転写[84]など、既に多様なテーマが研究されてきた。このようにダンス情報処理の研究分野は、今後幅広い方向への発展が期待できる。

6.2 AIST Dance Video Database

AIST Dance DBは、こうしたダンス情報処理の幅広い研究テーマに貢献できると我々は考えている。特に、ダンスモーション分析の研究(図5(a))での有用性が高く、ダンスモーションジャンル分類タスク以外にも、図2の基本ダンス、フリーダンス、グループダンス、移動カメラのカテゴリにダンス動画を自動分類する研究テーマも考えられる。基本ダンス動画では、すべてのダンサーが同一の基本的な振り付けで踊っているため、ダンサー間の微妙なダンスモーションの差を分析することで、個人の癖の検出やダンスモーションの個人差を吸収したダンステクニック検出に有用である。フリーダンス動画では、さまざまなダンサーが同じ楽曲に合わせて独自の振り付けを踊っているため、ダンスモーションにおけるダンサーの個性の差を分析でき、その結果はダンサー識別システムの開発に貢献できる。最大9方向から撮影された動画群は、任意の方向からダンスモーションを認識できるシステムの開発に役立つ。4.2節で述べたようにOpenPoseなどの画像処理技術を用いれば、ダンスモーションの自動生成を含む多様な目的での機械学習に有用なダンスモーションが抽出できる。

音楽情報処理の観点からは、AIST Dance DBに収録されたダンス楽曲60曲が、10種類のダンスジャンルとダンスモーションに対応していることが価値を持つ。ダンス楽曲の分類技術をダンスモーション併用の有無に応じて研究したり、ダンスモーションの分類技術をダンス楽曲併用の有無に応じて研究したりすることができる。さらに、ダンス楽曲とダンスモーションとの相関構造を、マルチモーダルに分析することもできる。「RWC研究用音楽データベース」[85,86]等と同様に、本DBは共通利用の自由、学術利用の自由が確保された研究用DBであるため、技術を比較評価するための共通ベンチマークとしても活用できる。

7. おわりに

本稿の主な貢献は以下の三つである。

- (1) 新規に制作したダンス楽曲に合わせて踊る独自のストリートダンス動画を収録した世界初の大規模な研究用DBを構築した。
- (2) 新たな研究分野「ダンス情報処理」を提唱し、議論した。
- (3) ダンスモーションジャンル分類タスクを提案し、4つのベースライン手法を開発して評価した。

今後、AIST Dance DBが世界中の研究者によって利用されて、多様なダンス情報処理技術の研究開発に貢献し、学術・文化・社会的なインパクトを与えることを我々は期待している。

謝辞 動画の自動切り出し作業に協力していただいたFabrizio Pedersoli氏に感謝する。本研究の一部はJST ACCEL(JPMJAC1602)の支援を受けた。

参考文献

- [1] Tsuchida, S., Fukayama, S., Hamasaki, M. and Goto, M.: AIST Dance Video Database: Multi-genre, Multi-dancer, and Multi-camera Database for Dance Information Processing, *Proceedings of the 20th International Society for Music Information Retrieval Conference, ISMIR 2019*, pp. 501–510 (2019).
- [2] Duggan, B., O’Shea, B., Gainza, M. and Cunningham, P.: Machine Annotation of Sets of Traditional Irish Dance Tunes, *Proceedings of the 9th International Society for Music Information Retrieval Conference, ISMIR 2016*, pp. 401–406 (2008).
- [3] Risk, L., Mok, L., Hankinson, A. and Cumming, J.: Melodic Similarity in Traditional French-Canadian Instrumental Dance Tunes, *Proceedings of the 16th International Society for Music Information Retrieval Conference, ISMIR 2015*, pp. 93–99 (2015).
- [4] Beauguitte, P., Duggan, B. and Kelleher, J. D.: A Corpus of Annotated Irish Traditional Dance Music Recordings: Design and Benchmark Evaluations, *Proceedings of the 17th International Society for Music Information Retrieval Conference, ISMIR 2016*, pp. 53–59 (2016).
- [5] Holzapfel, A. and Benetos, E.: The Sousta Corpus: Beat-Informed Automatic Transcription of Traditional Dance Tunes, *Proceedings of the 17th International Society for Music Information Retrieval Conference, ISMIR 2016*, pp. 531–537 (2016).
- [6] Collins, N.: Influence in Early Electronic Dance Music: An Audio Content Analysis Investigation, *Proceedings of the 13th International Society for Music Information Retrieval Conference, ISMIR 2012*, pp. 1–6 (2012).
- [7] Kell, T. and Tzanetakis, G.: Empirical Analysis of Track Selection and Ordering in Electronic Dance Music using Audio Feature Extraction, *Proceedings of the 14th International Society for Music Information Retrieval Conference, ISMIR 2013*, pp. 505–510 (2013).
- [8] Yadati, K., Larson, M., Liem, C. C. S. and Hanjalic, A.: Detecting Drops in Electronic Dance Music: Content based approaches to a socially significant music event, *Proceedings of the 15th International Society for Music Information Retrieval Conference, ISMIR 2014*, pp. 143–148 (2014).
- [9] Panteli, M., Bogaards, N. and Honingh, A. K.: Modeling Rhythm Similarity for Electronic Dance Music, *Proceedings of the 15th International Society for Music Information Retrieval Conference, ISMIR 2014*, pp. 537–542 (2014).
- [10] Knees, P., Faraldo, Á., Herrera, P., Vogl, R., Böck, S., Hörschläger, F. and Goff, M. L.: Two Data Sets for Tempo Estimation and Key Detection in Electronic Dance Music Annotated from User Corrections, *Proceedings of the 16th International Society for Music Information Retrieval Conference, ISMIR 2015*, pp. 364–370 (2015).
- [11] Zhang, W., Liu, Z., Zhou, L., Leung, H. and Chan, A. B.: Martial Arts, Dancing and Sports Dataset, *Journal of Image Vision Computing*, Vol. 61, No. C, pp. 22–39 (2017).
- [12] Schreiber, H. and Müller, M.: A Crowdsourced Experiment for Tempo Estimation of Electronic Dance Music, *Proceedings of the 19th International Society for Music Information Retrieval Conference, ISMIR 2018*, pp. 409–415 (2018).
- [13] Dixon, S., Pampalk, E. and Widmer, G.: Classification of dance music by periodicity patterns, *Proceedings of the 4th International Conference on Music Information Retrieval, ISMIR 2003*, pp. 159–165 (2003).
- [14] Gouyon, F. and Dixon, S.: Dance music classification: A tempo-based approach, *Proceedings of the 5th International Conference on Music Information Retrieval, ISMIR 2004* (2004).
- [15] Goto, M. and Muraoka, Y.: A Beat Tracking System for Acoustic Signals of Music, *Proceedings of the ACM International Conference on Multimedia, MM 1994*, pp. 365–372 (1994).
- [16] Offi, F., Erzin, E., Yemez, Y. and Tekalp, A. M.: Multi-modal analysis of dance performances for music-driven choreography synthesis, *Proceedings of the IEEE Conference on Acoustics, Speech, and Signal Processing, ICASSP 2010*, pp. 2466–2469 (2010).
- [17] Offi, F., Erzin, E., Yemez, Y. and Tekalp, A. M.: Learn2Dance: Learning Statistical Music-to-Dance Mappings for Choreography Synthesis, *Journal of IEEE Transactions on Multimedia*, Vol. 14, No. 3-2, pp. 747–759 (2012).
- [18] Fukayama, S. and Goto, M.: Music Content Driven Automated Choreography With Beat-wise Motion Connectivity Constraints, *Proceedings of the 12th Sound and Music Computing Conference, SMC 2015*, pp. 177–183 (2015).
- [19] Alemi, O., Françoise, J. and Pasquier, P.: GrooveNet: Real-time music-driven dance movement generation using artificial neural networks, *Workshop on Machine Learning for Creativity, 23rd ACM SIGKDD Conference on Knowledge Discovery and Data Mining, SIGKDD 2017*, Vol. 8, p. 26 (2017).
- [20] Gkiokas, A. and Katsouros, V.: Convolutional Neural Networks for Real-Time Beat Tracking: A Dancing Robot Application, *Proceedings of the 18th International Society for Music Information Retrieval Conference, ISMIR 2017*, pp. 286–293 (2017).
- [21] Tang, T., Jia, J. and Mao, H.: Dance with Melody: An LSTM-autoencoder Approach to Music-oriented Dance Synthesis, *Proceedings of the 26th ACM International Conference on Multimedia, MM 2018*, pp. 1598–1606 (2018).
- [22] Chu, W. T. and Tsai, S. Y.: Rhythm of Motion Extraction and Rhythm-Based Cross-Media Alignment for Dance Videos, *Journal of IEEE Transactions on Multimedia*, Vol. 14, No. 1, pp. 129–141 (2012).
- [23] Tsuchida, S., Fukayama, S. and Goto, M.: Query-by-Dancing: A Dance Music Retrieval System Based on Body-Motion Similarity, *Proceedings of the 24th International Conference on MultiMedia Modeling, MMM 2019*, pp. 251–263 (2019).
- [24] Guedes, C.: Controlling Musical Tempo from Dance Movement in Real-Time: A Possible Approach, *Proceedings of the 29th International Computer Music Conference, ICMC 2003* (2003).
- [25] Raptis, M., Kirovski, D. and Hoppe, H.: Real-time Classification of Dance Gestures from Skeleton Animation, *Proceedings of the 10th ACM SIGGRAPH/Eurographics Symposium on Computer Animation, SCA 2011*, pp. 147–156 (2011).
- [26] Kim, D., Kim, D. H. and Kwak, K. C.: Classification of K-Pop dance movements based on skeleton information obtained by a kinect sensor, *Sensors*, Vol. 17, No. 6, p. 1261 (2017).
- [27] Godøy, R. I. and Jensenius, A. R.: Body Movement in Music Information Retrieval, *Proceedings of the 10th International Society for Music Information Retrieval Conference, ISMIR 2009*, pp. 45–50 (2009).
- [28] MacRitchie, J., Buck, B. and Bailey, N. J.: Visualising Musical Structure Through Performance Gesture, *Proceedings of the 10th International Society for Music Information Retrieval Conference, ISMIR 2009*, pp. 237–242 (2009).
- [29] Nymoen, K., Godøy, R. I., Jensenius, A. R. and Tørresen, J.: Analyzing correspondence between sound objects and body motion, *Journal of ACM Transactions on Applied Perception*,

- Vol. 10, No. 2, pp. 9:1–9:22 (2013).
- [30] Seger, R. A., Wanderley, M. M. and Koerich, A. L.: Automatic detection of musicians' ancillary gestures based on video analysis, *Journal of Expert Systems with Applications*, Vol. 41, No. 4, pp. 2098–2106 (2014).
- [31] Li, B., Maezawa, A. and Duan, Z.: Skeleton Plays Piano: Online Generation of Pianist Body Movements from MIDI Performance, *Proceedings of the 19th International Society for Music Information Retrieval Conference, ISMIR 2018*, pp. 218–224 (2018).
- [32] Rizzo, A., Raheb, K. E., Whatley, S., Cisneros, R. M., Zanoni, M., Camurri, A., Viro, V., Matos, J.-M., Piana, S. and Buccoli, M.: WhoLoDancE: Whole-body interaction learning for dance education, *Proceedings of the Workshop on Cultural Informatics, co-located with the EUROMED International Conference on Digital Heritage, EUROMED 2018*, pp. 41–50 (2018).
- [33] Wang, P. and Rehg, J. M.: A Modular Approach to the Analysis and Evaluation of Particle Filters for Figure Tracking, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2006*, pp. 790–797 (2006).
- [34] Saini, M., Venkatagiri, S. P., Ooi, W. T. and Chan, M. C.: The Jiku Mobile Video Dataset, *Proceedings of the 4th ACM Multimedia Systems Conference, MMSys 2013*, pp. 108–113 (2013).
- [35] Chu, W. and Tsai, S.: Rhythm of Motion Extraction and Rhythm-Based Cross-Media Alignment for Dance Videos, *Journal of IEEE Transactions on Multimedia*, Vol. 14, No. 1, pp. 129–141 (2012).
- [36] Castro, D., Hickson, S., Sangkloy, P., Mittal, B., Dai, S., Hays, J. and Essa, I. A.: Let's Dance: Learning From Online Dance Videos, *arXiv preprint arXiv:1801.07388* (2018).
- [37] Stavrakis, E., Aristidou, A., Savva, M., Himona, S. L. and Chrysanthou, Y.: Digitization of Cypriot Folk Dances, *Proceedings of the 4th International Conference on Progress in Cultural Heritage Preservation, EuroMed 2012*, pp. 404–413 (2012).
- [38] Tang, T., Jia, J. and Hanyang, M.: Dance with Melody: An LSTM-autoencoder Approach to Music-oriented Dance Synthesis, *Proceedings of the 26th ACM International Conference on Multimedia, MM 2018*, pp. 1598–1606 (2018).
- [39] Essid, S., Lin, X., Gowing, M., Kordelas, G., Aksay, A., Kelly, P., Fillon, T., Zhang, Q., Dielmann, A., Kitanovski, V., Tournemene, R., Masurelle, A., O'Connor, E. I. N. E., Daras, P. and Richard, G.: A multi-modal dance corpus for research into interaction between humans in virtual environments, *Journal on Multimodal User Interfaces*, Vol. 7, No. 1–2, pp. 157–170 (2013).
- [40] Abu-El-Haija, S., Kothari, N., Lee, J., Natsev, P., Toderici, G., Varadarajan, B. and Vajayanasimhan, S.: YouTube-8M: A large-scale video classification benchmark, *arXiv preprint arXiv:1609.08675* (2016).
- [41] Schindler, A. and Rauber, A.: An audio-visual approach to music genre classification through affective color features, *Proceedings of the European Conference on Information Retrieval, ECIR 2015*, pp. 61–67 (2015).
- [42] Cao, Z., Simon, T., Wei, S. E. and Sheikh, Y.: Realtime multi-person 2D pose estimation using Part Affinity Fields, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, pp. 7291–7299 (2017).
- [43] Hochreiter, S. and Schmidhuber, J.: Long Short-Term Memory, *Journal of Neural Computation*, Vol. 9, No. 8, pp. 1735–1780 (1997).
- [44] Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L. and Lerer, A.: Automatic differentiation in pytorch, *Proceedings of the 31st Conference on Neural Information Processing Systems, NIPS 2017* (2017).
- [45] Boer, V., Jansen, J., Pauw, A. T. and Nack, F.: Interactive Dance Choreography Assistance, *Proceedings of the 14th International Conference on Advances in Computer Entertainment Technology, ACE 2017*, pp. 637–652 (2017).
- [46] Camurri, A., El, R. K., Even-Zohar, O., Ioannidis, Y., Markatzi, A., Matos, J., Morley-Fletcher, E., Palacio, P., Romero, M., Sarti, A., Pietro, S. D., Viro, V. and Whatley, S.: WhoLoDancE: Towards a Methodology for Selecting Motion Capture Data Across Different Dance Learning Practice, *Proceedings of the 3rd International Symposium on Movement and Computing, MOCO 2016*, pp. 43:1–43:2 (2016).
- [47] Großhauser, T., Bläsing, B., Spieth, C. and Hermann, T.: Wearable sensor-based real-time sonification of motion and foot pressure in dance teaching and training, *Journal of the Audio Engineering Society*, Vol. 60, No. 7/8, pp. 580–589 (2012).
- [48] Faridee, A. Z. M., Ramamurthy, S. R., Hossain, H. M. S. and Roy, N.: HappyFeet: Recognizing and Assessing Dance on the Floor, *Proceedings of the 19th International Workshop on Mobile Computing Systems & Applications, HotMobile 2018*, pp. 49–54 (2018).
- [49] Soga, A., Yazaki, Y., Umino, B. and Hirayama, M.: Body-part Motion Synthesis System for Contemporary Dance Creation, *Proceedings of the ACM SIGGRAPH 2016 Posters, SIGGRAPH 2016*, pp. 29:1–29:2 (2016).
- [50] Mousas, C.: Performance-Driven Dance Motion Control of a Virtual Partner Character, *Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2018*, pp. 57–64 (2018).
- [51] Tang, J. K. T., Chan, J. C. P. and Leung, H.: Interactive Dancing Game with Real-time Recognition of Continuous Dance Moves from 3D Human Motion Capture, *Proceedings of the 5th International Conference on Ubiquitous Information Management and Communication, ICUIMC 2011*, pp. 50:1–50:9 (2011).
- [52] Chan, J. C. P., Leung, H., Tang, J. K. T. and Komura, T.: A Virtual Reality Dance Training System Using Motion Capture Technology, *Journal of IEEE Transactions on Learning Technologies*, Vol. 4, No. 2, pp. 187–195 (2011).
- [53] Charbonneau, E., Miller, A. and LaViola, J. J.: Teach Me to Dance: Exploring Player Experience and Performance in Full Body Dance Games, *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology, ACE 2011*, pp. 43:1–43:8 (2011).
- [54] Alexiadis, D. S., Kelly, P., Daras, P., O'Connor, N. E., Boubekeur, T. and Moussa, M. B.: Evaluating a Dancer's Performance Using Kinect-based Skeleton Tracking, *Proceedings of the 19th ACM International Conference on Multimedia, MM 2011*, pp. 659–662 (2011).
- [55] Felice, M. C., Alaoui, S. F. and Mackay, W. E.: Knotation: Exploring and Documenting Choreographic Processes, *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI 2018*, pp. 448:1–448:12 (2018).
- [56] Yamaguchi, T. and Kadone, H.: Bodily Expression Support for Creative Dance Education by Grasping-Type Musical Interface with Embedded Motion and Grasp Sensors, *Sensors*, Vol. 17, No. 5, p. 1171 (2017).
- [57] Raheb, K. E., Tsampounaris, G., Katifori, A. and Ioannidis, Y.: Choreomorphy: A Whole-body Interaction Experience for Dance Improvisation and Visual Experimentation, *Proceedings of the 2018 International Conference on Advanced Visual Interfaces, AVI 2018*, pp. 27:1–27:9 (2018).

- [58] Brown, C.: Machine Tango: An Interactive Tango Dance Performance, *Proceedings of the Thirteenth International Conference on Tangible, Embedded, and Embodied Interaction, TEI 2019*, pp. 565–569 (2019).
- [59] Camurri, A.: Interactive Dance/Music Systems, *Proceedings of the 21th International Computer Music Conference, ICMC 1995*, pp. 245–252 (1995).
- [60] Caputo, F., Gowen, V. M., Geigel, J., Cerqueira, S., Williams, Q., Schweppe, M., Fa, Z., Pembroke, A. and Roffe, H.: Farewell to Dawn: A Mixed Reality Dance Performance in a Virtual Space, *Proceedings of the ACM SIGGRAPH 2016 Posters, SIGGRAPH 2016*, pp. 49:1–49:2 (2016).
- [61] Barkhuus, L., Engström, A. and Zoric, G.: Watching the Footwork: Second Screen Interaction at a Dance and Music Performance, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2014*, pp. 1305–1314 (2014).
- [62] Tsuchida, S., Terada, T. and Tsukamoto, M.: A Dance Performance Environment in Which Performers Dance with Multiple Robotic Balls, *Proceedings of the 7th Augmented Human International Conference, AH 2016*, pp. 12:1–12:8 (2016).
- [63] Hieda, N.: Mobile Brain-Computer Interface for Dance and Somatic Practice, *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology, UIST 2017*, pp. 25–26 (2017).
- [64] Tsuchida, S., Takemori, T., Terada, T. and Tsukamoto, M.: Mimebot: spherical robot visually imitating a rolling sphere, *International Journal of Pervasive Computing and Communications*, Vol. 13, No. 1, pp. 92–111 (2017).
- [65] Fujimoto, M., Fujita, N., Takegawa, Y., Terada, T. and Tsukamoto, M.: A Motion Recognition Method for a Wearable Dancing Musical Instrument, *Proceedings of the International Symposium on Wearable Computers, ISWC 2009*, pp. 11–18 (2009).
- [66] Paradiso, J. A., Hsiao, K. Y. and Hu, E.: Interactive Music for Instrumented Dancing Shoes, *Proceedings of the 25th International Computer Music Conference, ICMC 1999*, pp. 453–456 (1999).
- [67] Fujimoto, M., Fujita, N., Terada, T. and Tsukamoto, M.: Lighting Choreographer: An LED Control System for Dance Performances, *Proceedings of the 13th ACM International Conference on Ubiquitous Computing, UbiComp 2011*, pp. 613–614 (2011).
- [68] Kim, H. and Landay, J. A.: Aeroquake: Drone Augmented Dance, *Proceedings of the ACM Designing Interactive Systems Conference, DIS 2018*, pp. 691–701 (2018).
- [69] Golz, D. P. and Shaw, A.: Augmenting Live Performance Dance Through Mobile Technology, *Proceedings of the 28th International BCS Human Computer Interaction Conference on HCI 2014 - Sand, Sea and Sky - Holiday HCI, BCS-HCI 2014*, pp. 311–316 (2014).
- [70] Nakamura, A., Tabata, S., Ueda, T., Kiyofuji, S. and Kuno, Y.: Multimodal Presentation Method for a Dance Training System, *Proceedings of the SIGCHI Conference Extended Abstracts on Human Factors in Computing Systems, CHI EA 2005*, pp. 1685–1688 (2005).
- [71] Tsuchida, S., Terada, T. and Tsukamoto, M.: A System for Practicing Formations in Dance Performance Supported by Self-propelled Screen, *Proceedings of the 4th Augmented Human International Conference, AH 2013*, pp. 178–185 (2013).
- [72] Soga, A. and Yoshida, I.: Interactive Control of Dance Groups Using Kinect, *Proceedings of the 10th International Conference on Computer Graphics Theory and Applications, GRAPP 2015*, pp. 362–365 (2015).
- [73] Özcimder, K., Dey, B., Lazier, R. J., Trueman, D. and Leonard, N. E.: Investigating group behavior in dance: an evolutionary dynamics approach, *Proceedings of the American Control Conference, ACC 2016*, pp. 6465–6470 (2016).
- [74] Griggio, C. F. and Romero, M.: Canvas Dance: An Interactive Dance Visualization for Large-Group Interaction, *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems, CHI EA 2015*, pp. 379–382 (2015).
- [75] Raheb, K. E. and Ioannidis, Y.: Modeling abstractions for dance digital libraries, *Proceedings of the 14th ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL 2014*, pp. 431–432 (2014).
- [76] Sakashita, M., Suzuki, K., Kawahara, K., Takazawa, K. and Ochiai, Y.: Materialization of Motions: Tangible Representation of Dance Movements for Learning and Archiving, *Proceedings of the ACM SIGGRAPH 2017 Studio, SIGGRAPH 2017*, pp. 7:1–7:2 (2017).
- [77] Kim, E. S.: ChoreoSave: A digital dance preservation system prototype, *Proceedings of the American Society for Information Science and Technology*, pp. 48:1–48:10 (2011).
- [78] Zhang, X., Dekel, T., Xue, T., Owens, A., He, Q., Wu, J., Mueller, S. and Freeman, W. T.: MoSculp: Interactive Visualization of Shape and Time, *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology, UIST 2018*, pp. 275–285 (2018).
- [79] Dreameau, A. and Essid, S.: Probabilistic dance performance alignment by fusion of multimodal features, *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2013*, pp. 3642–3646 (2013).
- [80] Gowing, M., Kell, P., O’Connor, N. E., Concolato, C., Essid, S., Lefevre, J., Tournemenne, R., Izquierdo, E., Kitanovski, V., Lin, X. and Zhang, Q.: Enhanced Visualisation of Dance Performance from Automatically Synchronised Multimodal Recordings, *Proceedings of the 19th ACM International Conference on Multimedia, MM 2011*, pp. 667–670 (2011).
- [81] Tsuchida, S., Fukayama, S. and Goto, M.: Automatic System for Editing Dance Videos Recorded Using Multiple Cameras, *Proceedings of the 14th International Conference on Advances in Computer Entertainment Technology, ACE 2017*, pp. 671–688 (2017).
- [82] Jehan, T., Lew, M. and Vaucelle, C.: Cati Dance: Self-edited, Self-synchronized Music Video, *Proceedings of the ACM SIGGRAPH 2003 Sketches & Applications, SIGGRAPH 2003*, pp. 1–1 (2003).
- [83] Bakala, E., Zhang, Y. and Pasquier, P.: MAVi: A Movement Aesthetic Visualization Tool for Dance Video Making and Prototyping, *Proceedings of the 5th International Conference on Movement and Computing, MOCO 2018*, pp. 11:1–11:5 (2018).
- [84] Aristidou, A., Zeng, Q., Stavrakis, E., Yin, K., Cohen-Or, D., Chrysanthou, Y. and Chen, B.: Emotion Control of Unstructured Dance Movements, *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation, SCA 2017*, pp. 9:1–9:10 (2017).
- [85] Goto, M., Hashiguchi, H., Nishimura, T. and Oka, R.: RWC Music Database: Popular, Classical, and Jazz Music Databases, *Proceedings of the 3rd International Conference on Music Information Retrieval, ISMIR 2002*, pp. 287–288 (2002).
- [86] Goto, M., Hashiguchi, H., Nishimura, T. and Oka, R.: RWC Music Database: Music Genre Database and Musical Instrument Sound Database, *Proceedings of the 4th International Conference on Music Information Retrieval, ISMIR 2003*, pp. 229–230 (2003).