

Query-by-Dancing: 身体動作の類似性に基づくダンス楽曲検索システム

土田 修平* 深山 覚* 後藤 真孝*

概要. ダンス動作をクエリとしてダンス楽曲を検索できる楽曲検索システム Query-by-Dancing を提案する。ダンサーはダンスに用いる楽曲を検索する際に、自分が踊ろうとしているダンス動作に似た動作が含まれるダンス動画を参考にして楽曲を探すことがある。しかし、従来の楽曲検索システムにダンス動作をクエリとするものはなかった。そこで、ダンス動画データベース中を検索するためのクエリとして、ユーザのダンス動作を用いる Query-by-Dancing を開発した。このシステムは、クエリであるダンスに含まれるポーズ（姿勢）とモーション（動作）を抽出し、それらに類似したポーズとモーションが含まれるダンス動画を検索する。得られたダンス動画に付随する楽曲を用いることで、自分のダンスに適した楽曲を検索できる。システムの実装と実験の結果、ダンスと楽曲の関係性を考慮しながらダンスと楽曲をともに探索していく、新しいインタラクションの可能性が示唆された。

1 はじめに

動画投稿サービスの普及に伴い、ダンス動画はネットワーク上に大量にアップロードされるようになった。これら大量の動画から自分が見たい動画にアクセスするには、キーワードやタグを活用した検索に加えて、音楽や映っている人のダンス動作といった動画の内容に基づいた検索が有用であると考えられる。特にダンス動画はダンス動作を含む動画であるため、動画中に含まれるダンス動作を画像処理によって分析し、その動作の特徴に基づいて動画が検索できるようになれば、自分の見たいダンス動作が含まれるダンス動画が検索できるようになる。しかし、これには動画からのダンス動作の高精度な抽出技術が必要であり、これまで実現ができていなかった。

画像から身体の2次元骨格情報を推定できる技術 OpenPose [1] の登場により、動画中に含まれるダンス動作の抽出が容易になった。これにともない、ネットワーク上の大量のダンス動画に含まれるダンス動作を抽出できるようになり、ダンス動画を楽曲とダンス動作の双方の観点から分析できるようになった。楽曲とダンス動作が同期されているこの大規模な動画群を活用することで、ダンスに関する多様な技術を生み出せる可能性が生まれた。

ダンス動作と音楽の関係性に着目すると、様々なインタラクションを生み出すことができる。例えば、ダンスを入力として、そのダンスに類似したダンス動画を検索し、検索によって得られた動画に付随する楽曲を分析することで、入力したダンスのテンポやダンスの雰囲気・印象が推定できる可能性がある。クラブの空間など多くの人が踊っている場所において、踊っている人のダンス動作を抽出したものを入

力して検索することで、その空間にいる人のダンス動作にあった最適な楽曲を提供する DJ システムが実現できる。また、自身のダンスパフォーマンスに使用したいお気に入りの楽曲がある場合に、その楽曲をクエリとして検索することで、その楽曲を用いて誰かが踊っている振り付けを大量に集めることもできる。それら振り付けの中に気に入ったものがあれば、お気に入りの楽曲に合った振り付けをすぐさま練習できる。さらには、そのお気に入りの振り付けをクエリとして検索すると、その振り付けに合った大量の楽曲に出会うことができる。このように、お気に入りの振り付け、お気に入りの楽曲を交互に検索し続けることで、ダンサーの好みに合った楽曲と振り付けを効率的に探索し続けることができる。

可能となるインタラクションはこれに留まらない。例えばダンスで DJ が行える。これは、ダンス動作に対して単に楽曲を割り当てるようなものではない。大規模な動画群を活用すれば、ダンサーにとって妥当に感じられる楽曲を、ダンスによって DJ のように切り替えていくことができる。このような楽曲の切り替え手法をマスターすれば、ダンスで楽曲を切り替えていきつつ、その楽曲に合致したダンスを踊り続けることができる。その他にも、ダンス動作と楽曲中の歌詞の関係に着目して、ダンスから歌が生成できるかもしれない。カラオケなどに応用すると、激しく動いて盛り上がりたい人に適切な楽曲を推薦できるかもしれない。

楽曲とダンス動作の双方に加えて、その他の情報を関連付けていくことで、インタラクションの可能性はさらに拡大する。インターネット上の動画共有サービス、ニコニコ動画などでダンス動画に対して寄せられる多くのコメントと楽曲・ダンス動作の関係性を分析することで、ダンスの盛り上がり具合に合致するコメントを自動生成する、場の雰囲気をより盛り

Copyright is held by the author(s).

* 産業技術総合研究所

上げる演出を行う，といったことが可能である。

上記のことは，すべて楽曲とダンスが対応付いた大規模なコンテンツデータが活用できるからこそ実現可能性を持つ。そのため，そのような大規模コンテンツデータを扱うためのインタラクション手法を次々と生み出し，体系化していく必要がある。

本研究はそのための第一歩として，ダンス動作を入力とするダンスの楽曲検索に取り組む。ダンサーがダンスパフォーマンスや練習，個人的な遊びのためにダンスのための楽曲を決める際，どの楽曲を選ぶかは重要であり，ダンスパフォーマンスの成功やダンスの楽しみを左右する。ダンサーは，楽曲を選ぶときに自分のダンスと似たダンスが踊られているオンライン上のダンス動画を探して，その動画に含まれる音楽を参照することがある。この他には，お気に入りのダンサーやダンスグループがダンスショーで使用している楽曲を参考にする場合もある。これらの作業は時間がかかるにも関わらず，これまでダンスのための楽曲の探索をサポートするシステムは提案されていなかった。これまで多くの楽曲検索システム [2–4] やサービス [5, 6] が提案されているが，ダンスに特化したダンスのための楽曲の検索に焦点を当てたものはなかった。

そこでダンス動作を入力としたダンス楽曲検索システム Query-by-Dancing を提案する。Query-by-Dancing では，ダンス動作を撮影した動画をクエリとしてシステムに与えることで，クエリに類似したダンス動作を含むダンス動画の集合をデータベースから検索することができる。こうして得られたダンス動画に含まれるダンス動作は，ユーザによるダンス動作に類似しているため，ダンス動画に付随する楽曲はユーザのダンスに適していると予想される。また，ユーザがそれら検索結果のダンス動画を容易に閲覧し，最良な楽曲を見つけることができるインタフェースを提供する。本システムでは，身体の2次元骨格情報を推定するための Cao ら [1] によって作成されたライブラリ OpenPose を用いて，入力クエリの動画とデータベース中のダンス動画の分析を行った。ダンス経験者である筆者1人が実際に本提案システムを試用し，得られたダンス動画に含まれるダンスのジャンルと楽曲の音楽ジャンルについて確認をする。さらに，システムについての考察として，有用性や技術的限界についても述べ，最後にシステムの拡張性について議論する。

2 関連研究

多数の楽曲検索および推薦システムが提案されているが，ダンサーがダンス動作を用いて楽曲を検索するシステムはこれまで提案されていない。ここで関連研究として，様々なクエリを用いる楽曲検索や推薦システムに関する研究をいくつか挙げる。

Ghias ら [7] は，ハミングをクエリとして使用す

る Query by humming を提案している。楽曲のオーディオデータベースを検索する効果的かつ自然な手法は，その楽曲をハミングすることであると主張している。Chen ら [8] は，楽曲データベースからリズムで曲を検索するシステム，Query by rhythm を提案した。楽曲情報を音符列として扱い，クエリに類似したパターンを含むすべての楽曲を返す。Jang ら [9] は Query-by-tapping を提案した。このシステムは，ユーザがマイクをタップすることでクエリを作成し，楽曲の最初の数音符の継続時間を入力することで，楽曲データベースから楽曲を検索できる。Maetzawa ら [10] は，テンポを指揮することで，再生される演奏を動的に切り替えることができる Query-by-conducting を提案している。

3 ダンス楽曲検索システム

システムのインタフェース画面を図1に示す。画面左上で，入力したダンス動画を確認できる。画面右上のグラフは，入力動画のダンス動作とデータベース上の動画内のダンス動作との類似度を示しており，取り出される類似度上位10件の動画は，グラフ上に青い円のマークで示されている。このグラフを確認することで，推薦された動画がデータベース内の他の動画と比べてどの程度類似度が高いのかがわかる。画面下側の動画10件は，入力動画のダンス動作との類似度上位10件である。楽曲再生中の動画は赤い枠で囲まれており，ユーザはキー操作によって再生する楽曲を切り替えつつ，楽曲を選定していく。これにより，ユーザは入力動画のダンス動作に合った楽曲を見つけることができる。

次に，システムの概要を図2に示す。Query-by-dancing は前処理とダンス動作の類似度の計算の大きく2段階の処理に分けられる。この2つの段階についてそれぞれ説明する。

3.1 前処理

3.1.1 ダンサー検出

1つ目の段階では，まず OpenPose ライブラリ [1] を使用して，すべてのフレームに映っている人の骨格情報を推定する。踊っている人以外に複数の人が映っていたり，OpenPose が人のいない場所を誤検出したりと，複数の骨格情報が検出されてしまう場合がある。そのため，フレームごとに検出された骨格情報のうち，どの骨格情報がダンサーであるか選択しなければならぬ。得られた骨格情報の x 軸方向の最大値 x_{\max} と最小値 x_{\min} の差と， y 軸方向の最大値 y_{\max} と最小値 y_{\min} の差を掛けて得られる面積をダンサーの占める面積 A_o と定義した。画像の中心位置 P_d から映っている人のすべての骨格情報の平均位置 $P_c(X_{\text{mean}}, Y_{\text{mean}})$ までの距離を D_c として， $R = \frac{A_o}{D_c}$ を最大にする骨格情報をもつ人を，



図 1. UI 画面

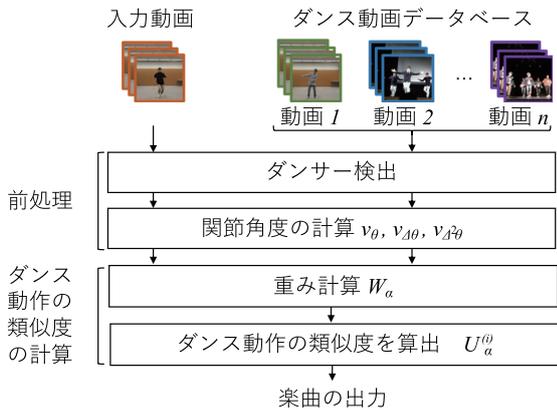


図 2. システム概要

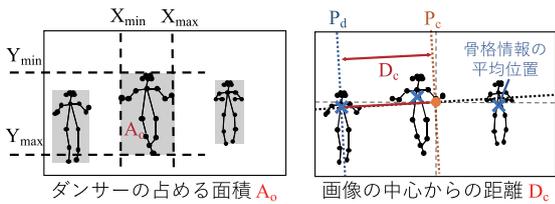


図 3. ダンサーの検出

踊っているダンサーとして選択した (図 3).

3.1.2 特徴量

動画間のダンス動作類似度を計算するために、フレームごとに3つ特徴量を抽出する。ここで、ダンスを特徴付ける要素としてポーズ（姿勢）とモーション（動作）が重要であると考えられる。まずポーズを考慮するために、OpenPoseによって推定された骨格情報から得られる17個の関節角度すべてを1フレームごとに計算する。角度は、画面垂直方向上側を基

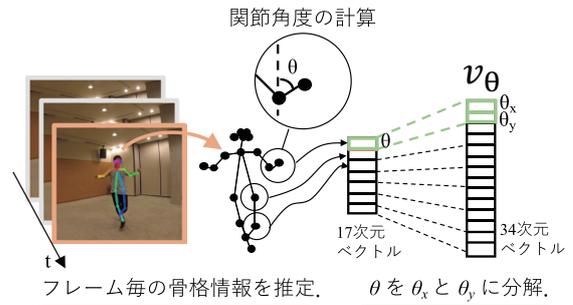


図 4. 関節角度の算出

準の0度として、時計回りに角度を算出する。さらに、これら関節角度を図4に示すように θ_x と θ_y の2つの次元に分解し、 i 番目の動画の n 番目のフレームにおける関節角度を34次元の特徴ベクトル $v_{\theta}^{(i)}(n) (1 \leq n \leq N^{(i)} (1 \leq i \leq I))$ で表す。データベースのビデオ総数を I 、 i 番目のビデオのフレーム数を N とした。骨格情報が検出されなかった関節角度は0を代入した。次に、モーションを考慮するために、フレーム間の関節角度の変化に焦点を当てる。 $v_{\Delta\theta}^{(i)}(n)$ と $v_{\Delta^2\theta}^{(i)}(n)$ を次式に基づいて算出する:

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

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

x の各要素の絶対値を含むベクトルを $\text{abs}(x)$ とした。以上3つの特徴量を102次元の1つのベクトル $v_{\alpha}^{(i)}(n)$ にまとめた。検索対象のすべての動画で計算したベクトルの各要素の平均と分散を求め、それらを用いて $v_{\alpha}^{(i)}(n)$ の各要素が平均0、分散1になるように正規化を行なった。

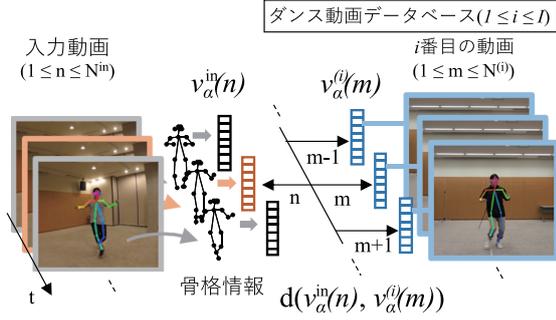


図 5. すべてのフレームの組み合わせにおいてユークリッド距離を計算し、その値を組み合わせの総数で割る。

3.2 類似度計算

2つ目の段階として、フレーム間の類似度を計算する。前節で示した特徴ベクトル $v_{\alpha}^{(i)}(n)$ を用いて、入力動画 in のすべてのフレーム ($1 \leq n \leq N^{in}$) とデータベース上 ($1 \leq i \leq I$) の動画 i のすべてのフレーム ($1 \leq m \leq N^{(i)}$) の間のユークリッド距離 $d(v_{\alpha}^{in}(n), v_{\alpha}^{(i)}(m))$ を計算する (図5)。 x と y のユークリッド距離を算出する式を $d(x, y)$ で表した。本システムでは、すべてのフレームの組み合わせにおいてユークリッド距離を計算し、その値を組み合わせの総数 $N^{in}N^{(i)}$ で割る。算出された値を以下の式で示す:

$$R_{\alpha}^{(i)} = \frac{1}{N^{in}N^{(i)}} \sum_{n=1}^{N^{in}} \sum_{m=1}^{N^{(i)}} d(v_{\alpha}^{in}(n), v_{\alpha}^{(i)}(m)). \quad (3)$$

ここで、 $R_{\alpha}^{(i)}$ を単純にダンス動作の類似度として動画を検索すると、入力クエリのダンスジャンルと関連性の低いダンス動画を得てしまう。そこで、tf-idf に似た手法を用いて、動画を検索する際のダンス動作の重要度を計算し、重み付けすることで、この問題を解決した。重みの計算式を以下に示す:

$$W_{\alpha}(n) = \frac{\frac{1}{N^{(i)}} \sum_{m=1}^{N^{(i)}} d(v_{\alpha}^{in}(n), v_{\alpha}^{(i)}(m))}{\max_{i \in I} \left\{ \frac{1}{N^{(i)}} \sum_{m=1}^{N^{(i)}} d(v_{\alpha}^{in}(n), v_{\alpha}^{(i)}(m)) \right\}}. \quad (4)$$

重みの傾斜を大きくするために、 $W_{\alpha}(n)$ を 30 乗し、 $W'_{\alpha}(n)$ とした。冪乗する値 30 は実験的に最も効果のあった値として選んだ。次に、全てのユークリッド距離に対応する重みを掛け合わせる。計算式を以下に示す:

$$U_{\alpha}^{(i)} = \frac{\sum_{n=1}^{N^{in}} [W'_{\alpha}(n) \sum_{m=1}^{N^{(i)}} d(v_{\alpha}^{in}(n), v_{\alpha}^{(i)}(m))]}{N^{in}N^{(i)}}. \quad (5)$$

次に、入力ビデオのダンス動作に類似した動作を含む動画として、 $U_{\alpha}^{(i)}$ の値が下位 10 件の動画を取り出す。最後に、検索されたダンス動画からダンス楽曲候補として提示する。

4 評価実験

重みあり・なしによる検索結果への影響、ダンスジャンルによる検索結果への影響をそれぞれ調査するために 2 つの評価実験を行なった。データベースとして 100 個の平均 82 秒程度のダンス動画を用意した。ダンスジャンルは Hip-hop, Break, Pop, Waack の 4 つで、1 ジャンルあたり 25 個の動画をもつ。全ての動画に楽曲が含まれており、映像上のダンサーはその楽曲に合わせて踊っている。

4.1 実験: 重みあり・なし

参加者はストリートダンスクラブに所属する学生 12 名 (男性 4 名, 女性 8 名) で、1~15 年 (平均 8.5 年) のダンス経験をもつ。ADD (重みなし), ADD (重みあり), DTW (重みなし), DTW (重みあり) の 4 つ手法を比較する。 $v_{\alpha}^i(n)$ を特徴ベクトルとし、フレーム間のユークリッド距離を求める手法を ADD とした。また、より長い区間での類似度を考慮するために、式 (3, 4, 5) のユークリッド距離計算 $d(v_{\alpha}^{in}(n), v_{\alpha}^{(i)}(m))$ を Dynamic Time Warping による計算に入れ換え、入力ベクトルを 6 フレームごとの系列に変更した手法を DTW とした。

ダンス歴 15 年の Waack ダンサーに踊ってもらい、11 秒の動画を作成した。この動画を入力クエリとして、それぞれの手法ごとに検索を行い、検索上位 5 件の楽曲を取得する。

実験は以下の流れで行う。まず、参加者は予備アンケートにダンス経験を記入する。実験の簡潔な説明を行い、次に、音声を含まない Waack ダンスの動画を見ながら、それぞれの手法の楽曲グループを聴いた。評価項目「この映像上のダンスを「誰か」が音楽に合わせて踊るとした時に、ダンスの雰囲気とこの 5 曲の音楽の雰囲気を考慮すると、踊りやすい楽曲である。」について 5 段階評価でスコアリングを行なった。

アンケート結果を図 6 の左に示す。分散分析を行なったところ、有意差が確認できた ($F_{(3,236)} = 4.21, p < .05$)。LSD 法を用いて多重比較を行なったところ、ADD (重みあり) とその他の手法との間で有意差が確認できた ($p < .05$)。ADD (重みあり) の手法が最も評価が高く、重みを付与することで、より踊りやすい楽曲が取得できることがわかった。

4.2 実験: 検索性能

参加者はストリートダンスクラブに所属する学生 12 名 (男性 6 名, 女性 6 名) で、1~15 年 (平均 5.9 年) のダンス経験をもつ。Waack, Hip-hop, Pop, Break の 4 つのダンスのジャンルで比較する。実験 I で使用した動画を Waack の動画として、Break 経験者である筆者が約 13 秒踊った動画を Break の動画として用いる。また、Pop, Hip-hop のダンサーの約 16 秒のダンスを撮影し、それぞれのジャンル

Query-by-Dancing: 身体動作の類似性に基づくダンス楽曲検索システム

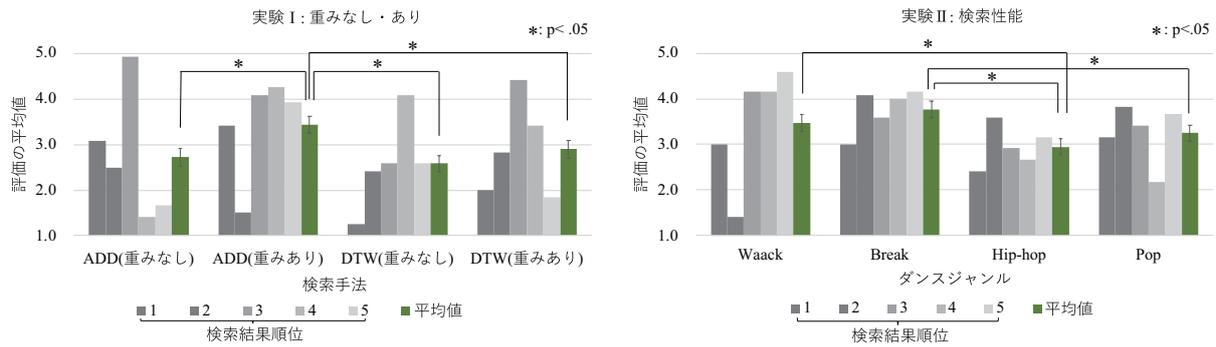


図 6. 左: 実験 I (重みあり・なし) のアンケート回答結果. 右: 実験 II(検索性能) のアンケート回答結果.

の動画とした。これら 4 つの動画を入力クエリとして、ADD (重みあり) の手法を用いて検索を行い、上位 5 件の楽曲を取得する。実験の流れは、実験 I と同様にして、それぞれの動画に対応する楽曲を参加者に 5 段階で評価してもらった。

アンケート結果を図 6 の右に示す。分散分析を行なったところ、有意差が確認できた ($F_{(3,236)} = 3.92, p < .05$)。LSD 法を用いて多重比較を行なったところ、Waack と Hip-hop, Break と Hip-hop, Break と Pop との間でそれぞれ有意差が確認できた ($p < .05$)。

Hip-hop に着目すると、異なるダンスジャンルである Break の楽曲が検索結果となっていた。これについては、2 つの原因が挙げられる。1 つ目は、Hip-hop の中に、さらに多くのジャンルが含まれていることである。今回入力に使用した Hip-hop は Middle Hip-hop であったが、データベース上のダンスは Style Hip-hop, Jazz Hip-hop, Girls Hip-hop などが多くを占めており、入力したダンススタイルと異なっていた。2 つ目は、Middle Hip-hop の動作が Break におけるいくつかの動作と一致することである。Break は床に手をつけて踊ることが多いダンスであるが、床に移るまでのダンスが Middle Hip-hop と似ていた。以上 2 つの理由から、入力クエリと異なるダンスジャンルである Break が取り出された。より細かくジャンル分けを行ない、データベースに追加していくことで、評価は改善されると考えられる。

Pop の評価に着目すると、Break と比較して悪く、Waack と比較しても悪い傾向である。これについて、評価を低くした参加者にインタビューしたところ、クエリとして用いた動画中のダンスには、体を小刻みに震わせる「バイブレーション」や、体に波のように動かす「ウェーブ」というテクニックが含まれていることを原因にあげた。これらの動作は、特定の音に細かく合わせた動きであり、そのような特定の音を含まない曲では、いくらそのダンスジャンルで使用されている曲であっても入力動画上のダンスに適さない楽曲と判断された。このことから、

ダンスモーションとのマッチ度合いをユーザが調整することで、状況に合わせた、より目的に沿った楽曲検索手法が考えられる。

5 拡張性

Query-by-Dancing システムは、ユーザのダンス動作に類似したダンス動画を検索できる。今回、入力動画に含まれる楽曲を検索の際に利用していなかったが、今後、楽曲を扱う視点からの研究を含めることで、より自分の目的に沿った楽曲検索が行えると考えられる。また、本システムを拡張させることで、ダンスに合った楽曲を、楽曲に合ったダンスを探索していく、新たなインタクシヨンの実現が考えられる。さらに、楽曲検索以外のインタラクションへの応用も考えられる。

5.1 ダンス動画検索

例えば、ユーザは、ダンス動画検索として本システムを使用できる。個人のダンスの独自性を保つために、ダンサーは他のダンサーのダンス動作を真似た動作は避ける。そのため、本システムを利用することで、自身のダンス動作と同様の動作がどの程度存在するか見ることができ、自身のダンス動作の独自性を客観的に確認できる。もしダンス動作ごとに細かく独自性を示す数値を表示できれば、自身の独自性ある動きに集中して練習できる。

振り付け作成支援にも応用できる可能性がある。例えば、途中で作成が止まってしまった振り付けを入力することで、その続きに繋がる振り付け候補を、大規模なデータベースから提案できる。これは先ほどの、動作を真似ることを避けることについて相反するが、実際にダンスパフォーマンスを作成する際にパフォーマンス全体の独自性を保つ一方で、次のダンス動作に繋げやすくするために、次のダンス動作は今の動作から繋げやすい動作をあてはめることは多いと考える。そのような振り付け作成をサポートできると考えられる。

ダンス動作をより細かく解析できるようになれば、

自身のダンス動作のレベルの確認にも使えるかもしれない。例えば、有名なダンサーのダンス動作を真似たダンス動画は多くアップロードされている。そのダンス動作を真似た動作を入力すれば、目指すべきダンサーのダンス動画が最も類似性の高い動画と表示されるか、もしくはその人とは関係のない人のダンス動画が表示されるかで、自身のダンス動作のレベルを確認できる。ダンス動作のズレを可視化すれば、実際にどの部分の動画が悪いかを指摘することも可能である。

5.2 自分以外のダンスによる検索

入力クエリに自分以外の人のダンス動画を用いることもできる。そうすれば、例えばお気に入りのダンサーやダンスグループがいた際に、そのダンサー達のダンス動画をクエリとして用いることで、キーワードやタグ検索では見つけづらい動画の発見につながる。また、これまで見たことのなかった、お気に入りのダンサーに似たダンスを行うダンサー達を発見できるかもしれない。

5.3 ダンスのデータ化によるインタラクション

ダンス動作を一度ポーズとモーションという形でデータ化してネットワークで利活用することで、その他インタラクションへの拡張も考えられる。例えば、遠く離れた人と振り付けあわせをすることはこれまで容易ではなかったが、一度それぞれが同じ楽曲に対してダンスを行い、そのダンス動作からポーズとモーションを抽出し、仮想空間で共有・表示すれば、仮想空間上で簡単にグループダンスの確認ができる。また、ネットワーク上にアップされているダンス動画からポーズやモーションを抽出することで、一緒に踊る相手を簡単に生成できるようになり、例えば会うことが難しい有名なダンサーと仮想空間上で一緒にダンスを演じることができるようかもしれない。これは同じ楽曲に同期したダンス動作だからこそ、現実世界の時間差がいくらあろうとも、簡単に同期できる。

以上のように、ダンス動作のポーズやモーションをネットワーク上のデータとして活用したインタラクションは、これから拡大していくと考える。

6 まとめ

ダンス動作で楽曲を検索できるダンス楽曲検索システム Query-by-Dancing を提案した。ダンス動画を検索クエリとし、ダンス動作のポーズとモーションの分析に基づいて、ダンス動画データベースからクエリのダンス動作に似た動作を含むダンス動画を検索することができた。

本研究は、楽曲とダンス動作が対応付いた大規模なコンテンツデータを活用する研究のうち、ダンス動作を扱う視点からの研究である。今後、本研究を

きっかけとして、楽曲を扱う視点からの研究を含め、大規模なコンテンツデータを活かした様々な研究を行い、インタラクション研究の発展に繋げたい。

謝辞

本研究の一部は JST ACCEL (JPMJAC1602) の支援を受けた。

参考文献

- [1] Z. Cao, T. Simon, S.E. Wei, and Y. Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*, 2017.
- [2] W. Chai and B. Vercoe. Using user models in music information retrieval systems. In *Proceedings of the 1st International Society of Music Information Retrieval*, pp. 23–25, 2000.
- [3] K. Hoashi, K. Matsumoto, and N. Inoue. Personalization of user profiles for content-based music retrieval based on relevance feedback. In *Proceedings of the 11th ACM international conference on Multimedia*, pp. 110–119, 2003.
- [4] K. Hoashi, H. Ishizaki, K. Matsumoto, and F. Sugaya. Content-based music retrieval using query integration for users with diverse preferences. In *Proceedings of the 8th International Society of Music Information Retrieval*, pp. 463–466, 2007.
- [5] SoundHound Inc. Soundhound. <https://www.soundhound.com/soundhound> (accessed June 1, 2018).
- [6] Shazam Entertainment Ltd. Shazam. <https://www.shazam.com/> (accessed June 1, 2018).
- [7] A. Ghias, J. Logan, D. Chamberlin, and B. C. Smith. Query by humming - musical information retrieval in an audio database. In *Proceedings of the 3rd ACM international conference on Multimedia*, pp. 231–236, 1995.
- [8] J.C.C. Chen and A.L.P. Chen. Query by rhythm: an approach for song retrieval in music databases. In *Proceedings of the 8th International Workshop on Research Issues in Data Engineering: Continuous-Media Databases and Applications*, pp. 139–146, 1998.
- [9] J.S.R. Jang, H. R. Lee, and C. H. Yeh. Query by tapping: A new paradigm for content-based music retrieval from acoustic input. In *Proceedings of the 2nd Pacific-Rim Conference on Multimedia*, pp. 590–597, 2001.
- [10] A. Maezawa, M. Goto, and H. G. Okuno. Query-by-conducting: An interface to retrieve classical-music interpretations by real-time tempo input. In *Proceedings of the 11th International Society of Music Information Retrieval*, pp 477–482, 2010.