

# 多視点ダンス映像のインタラクティブ編集システム

土田 修平\* 深山 覚\* 後藤 真孝\*

**概要.** 本稿では、自動編集機能を活用することで、動画編集の専門知識がなくとも簡単に魅力的なダンス動画を作成できる多視点ダンス動画のインタラクティブ編集システムを提案する。ダンサーは、パラメータを調整することで、自身の目的に応じた動画に自動で編集させることができる。また、動画編集の原則に基づいて決定されたカメラの優先選択度をヒートマップとして提示することで、カメラ選択を音楽の1拍ごとに修正して好みのダンス動画を制作できる。さらに、被写体とカメラの距離に近いほど画面上側に、遠いほど画面下側に並び替えられており、全体を俯瞰することで近い、遠い映像の割合を容易に確認することができる。自動編集機能の有効性を評価実験を通して確認し、本提案システムの必要性を確認した。

## 1 はじめに

ダンスは踊れても、ダンス動画を魅力的に作れるとは限らない。例えば、ダンサーが、自身で踊っている映像をカメラで撮影しても、魅力的な動画を作成することは難しい。これは、ダンサーが必ずしも魅力的な映像を撮影するための技術を持ち合わせていないからである。ここでいう魅力的な動画とは、専門家が動画を見てかっこいいと思う動画とする。ダンサーの動画編集に関する技能不足を補うために、少しでも魅力的な動画を作成するための手法の1つとして、多視点から撮影を行う場合がある。例えば、1つのカメラを用いて、何度も踊り直し、多視点の動画素材を作成する。複数のカメラが備えられたスタジオなどで、複数のカメラを用いて撮影する場合が考えられる。また、路上でのダンスパフォーマンスでは、観衆が自身のカメラでパフォーマンスを撮影し、その動画を共有することで多視点の動画素材を収集できる場合も考えられる。このように、多視点のカメラ映像を収集するシチュエーションは複数存在し、ダンサーは多様な動画素材を入手できる。しかしダンサーは、ダンスに関する知識を持っていても、これら多様な動画素材を魅力的な1つのダンス動画にまとめる知識を必ずしも持っていない。自身のダンスをダンサー自身の手で魅力的なダンス動画に仕上げるためには、これら多視点動画の編集支援技術が不可欠である。なお、本稿では、映像とは視覚情報のみを指し、動画は映像と音響の組み合わせを指すとする。

動画の編集支援に関する研究は多数存在する。Leake et al. [1] は、対話をメインとした場面において、動画制作で用いられるイデオムを自由にカスタマイズすることで動画を編集できる VidCrit を提案した。これは、対話といった特定の場面における限定

的な情報を活用しており、ダンス動画における情報は想定されていない。Pavel et al. [2] は、動画編集レビューのためのフィードバック記録・表示インタフェースを提案した。専門家がコンテンツを確認しながらレビューを行う必要があり、ユーザとシステムのみインタラクションでは完結しない。また、自動動画編集に関する研究についても多数存在する。Arev et al. [3] は、シネマトグラフィのガイドラインを適応し、多視点の動画を1つの動画にまとめることができる。

本研究では、ダンサー個人でも簡単に魅力的な動画を作成できる、多視点ダンス映像のためのインタラクティブ編集システムを提案する。提案システムでは、ダンサー1人が音楽に合わせて踊っている様子を様々な角度から撮影した多視点カメラの動画を与えると、音楽の1拍ごとに適切な映像が動画編集の原則に基づいて選択された結果（最適な組み合わせパターン）を見ることができる。そのシステムが推奨する最適な組み合わせパターンに対して、ユーザは自身の好みに応じて編集できる。その際、カメラの選択優先度をヒートマップとして参考になるので、知識が乏しくても編集しやすくなっている。さらに、最初の最適な組み合わせパターン自体も、ダンサーのジャンルや利用目的・場面に合わせてカスタマイズ可能である。例えば、膨大なコンテンツが投稿されているSNS用には、カメラが短く切り替わって短時間でも異なる映像が登場するダンス動画が適切である。そこで、そうした組み合わせパターンを自動生成させ、その結果を自分好みに変えられる。ユーザの操作は主にマウス操作で行うことができる。映像は拍毎に区切られており、ダンサーは再生したい映像をクリックするだけで切り替えることができる。また、被写体に対して近い映像は画面上側に、遠い映像は画面下側に並び替えられており、映像の遠近のバランスを容易に確認できる。

以下、2章では、提案システムのインタフェース

Copyright is held by the author(s).

\* 産業技術総合研究所



図 1. インタフェース画面

の特徴、機能、及び操作方法について述べ、3章で提案システム実現方法と具体的な実装について述べる。4章で、自動編集機能についての評価実験の結果と考察をまとめる。5章で本稿のまとめを行う。

## 2 インタラクティブ編集システム

提案システムのインタフェースを図1に示す。図1左上に編集結果、左下にスライダ、右側に編集用のタイムラインが各映像と共に表示されている。編集用のタイムライン上では、各映像が1拍毎に区切られて表示されており、マウスで表示したい動画を選択することによって、1拍区間内の映像を変更することができる。映像は、Gunnar Farneback法 [7] により算出した各カメラのオプティカルフローの平均値を用いて、ダンサーとカメラの距離が近いほど画面上側に、遠いほど画面下側に並び替えられている。赤い三角形のマークは映像の未選択を示す。表示されているヒートマップは1拍における映像毎の選択優先度を示しており、選択優先度が低い順に青、緑、赤で表示される。スライダを調整することで、ヒートマップの濃淡パターンが変更され、スライダのパラメータに基づいた最適な映像の組み合わせが自動的に再選択される。スライダによって変化するヒートマップ生成の詳細については、3章で説明する。

ユーザの利用手順について説明する。ユーザがシステムを起動すると、タイムライン上の映像は最適な組み合わせで選択されており、再生ボタンを押すことですぐに編集結果を確認することができる。ユーザの好みに沿った映像の組み合わせであった場合は、そのまま編集作業に移行するが、好みに合わない場合は、図1左下のスライダを用いてユーザの

好みに沿った組み合わせを探す。組み合わせ確認後、ユーザの好みに合わない箇所へ移動し、ヒートマップを参考にしながら組み合わせを変更する。

## 3 提案システムの実現方法

### 3.1 魅力的な動画の自動編集

#### 3.1.1 動画編集の原則

映像制作の現場では、いくつかの動画編集の原則が存在しており、編集者はこれら原則を考慮した上で、自身の経験に基づいてコンテンツに合った編集を行う。同様に、動画編集によってより魅力的なダンス動画を生成するためには、動画編集の原則を踏まえた上で、ダンスパフォーマンスに合った編集を行う必要があると考えられる。そこで、文献 [4,5] から得られたダンス動画に適応可能な一般的な原則と、映像制作に携わる4名の専門家へのインタビューから得られたダンス動画に適応可能な経験的な原則を列挙する。尚、本稿では、被写体に近いショットをアップショット、被写体から遠い映像をロングショットと定義する。また、1拍あたりの映像を1ショット、同じショットを繋げた映像を1カットと定義する。

**原則1** ビートに合わせてカメラを切り替える。

**原則2** カットの最大長・最小長はユーザの好みに応じて変化させる。

**原則3** 構図の異なるカメラに切り替える。

**原則4** アップなショットほどカット長を短く、ロングなショットほどカット長を長くする。

**原則5** メインカメラが認識できる構成にする。

**原則 6 強調する場所を決めて緩急をつける。**

以上に挙げた6つの原則を踏まえて、魅力的な動画を生成するシステムを構築する。ここで、動画編集の過程に着目すると、多視点のカメラから撮影された複数の動画を1つの動画に編集する過程は、すべての時刻について、どのカメラを用いるか選ぶことと同義といえる。そこで本稿では、動画編集の最小時間単位を音楽の1拍とし、1拍毎に適切なカメラを1つ選択する最適化問題として定式化する。適切なカメラは、ダンサーの動作に応じて変わるだけでなく、前後にどのカメラが選択されているかによっても変わる。そこで、カメラ遷移の確率モデルを動画編集の原則に基づいて定義して、最適なパターンを求める。カメラの選択優先度は、1拍毎のカメラが選ばれる確率をサンプリングによって計算することで求めることができる。

魅力的な動画を生成するシステムを構築する上で、いくつかの原則をそれぞれ以下のように解釈した。原則3では、構図を映像上の動きの大きさと仮定した。遷移するカメラ間の映像上の動きの大きさの差を大きくさせることで、異なる構図に切り替わると考えられる。原則5では、1つのカメラのみ、生成される映像中で使われる割合を増やすことで、そのカメラをメインカメラとして認識するのではないかと考えられる。原則6では、何が魅力的な映像であるかは場合によって異なるため、今回は魅力的な部分を激しい動作と仮定した(逆も技術的には可能)。全体を通して相対的に激しい動作とそうではない場所との画面切り替えの緩急をとる。

**3.1.2 カメラ遷移の確率モデル**

**原則 2. カットの最大長・最小長はユーザの好みに応じて変化させる:** 使われる時間長が  $x$ 、遷移前のカメラが連続して最小長以下である場合は、遷移前と同じカメラに遷移する確率を  $1 - 1^{-20}$ 、それ以外を  $1^{-20}$  とした。また、遷移前のカメラが連続して最大長以上である場合は、同じカメラに遷移する確率を  $1^{-20}$ 、それ以外を  $\frac{1-1^{-20}}{N-1}$  とした。ここで  $N$  は全カメラの個数である。

**原則 3. 構図の異なるカメラに切り替える:** 時刻  $t$  で使用される映像を撮影するカメラを  $x_t$ 、カメラのインデックスを  $n$  とする。全てのカメラについて Gunnar Farneback 法 [7] により算出した各フレームのオプティカルフローの平均値  $W_t(t = 1 \sim t_{max})$  を全てのフレームにおいて算出し、全フレームで算出した値の平均値を算出した。算出した値をカメラ間で正規化し、各カメラが選ばれる確率として用いた。

$$p(x_t = n) = \frac{W_t^n}{\sum_n W_t} \quad (1)$$

**原則 4. アップなショットほどカット長を短く、口**

**ングなショットほどカット長を長くする:** 原則3にて算出したカメラ毎のオプティカルフローの平均値を2乗し(各カメラ毎の差が小さかったため、差を大きくするために実験的に決めた)、正規化したものを確率分布で表す。

前のカメラの場所から上記の確率分布に従って、現在のカメラの場所の確率値を、他のカメラへ移動する確率とみなした。移動先のカメラ番号は一律な確率で決める。こうすることで、アップショットの度合いが高いほど、他のカメラへ移動しやすくなる。

**原則 5. メインカメラが認識できる構成にする:** 次の基準にしたがってメインカメラを決定し、そのメインカメラが用いられる頻度を増やす。メインカメラは身体が全体的に大きく映っているカメラとし、以下のように映像からの骨格推定技術を活用しながら決定した。

1. OpenPose Library [8] を用いてダンサーの骨格情報を検出する。
2. フレーム毎の取得できなかった骨格情報数  $b_f$ 、取得可能な骨格情報数を  $B$ 、全フレーム数を  $F$  として、骨格情報の未検出率  $\gamma$  を以下の式(1)で表す。

$$\gamma = \frac{1}{F} \sum_{f=1}^F \frac{b_f}{B} \quad (2)$$

また、全フレームにおける骨格の検出位置の  $x$  方向、 $y$  方向における分散をそれぞれ  $V(x_{all})$ 、 $V(y_{all})$  とする。

3. 式(2)の値  $m$  が最大となるカメラをメインカメラとする。

$$m = \frac{V(x_{all}) + V(y_{all})}{\gamma} \quad (3)$$

メインカメラを0.6の確率で選択し、残りは一律な確率で遷移する確率分布を用いる。

**原則 6. 強調する場所を決めて緩急をつける:** 全体を通して相対的に激しい動作を行なっている際は、激しい動作が捉えられているショットを用いる。全てのカメラの1拍毎のオプティカルフローの平均値を算出し、中央値を閾値として用いて、閾値より平均値が高ければ激しい動作を行なっているショット、低ければ激しい動作を行なっていないショットとした。1拍毎のオプティカルフローの平均値を正規化して確率分布として表す。激しい動作でない部分であれば、オプティカルフローの最大値からそれぞれの平均値を引いたものを正規化し、各カメラが選ばれる確率として用いた。激しい動作の部分であればオプティカルフローの平均値を正規化し、各カメラが選ばれる確率として用いた。

以上の原則 2~6 に相当する確率分布をそれぞれ、 $P_2, P_3, P_4, P_5, P_6$  とする。あり得るカメラ遷移のサンプリングを行い、評価関数  $O(x_{1:T})$  が最大となる時のカメラの組み合わせで動画を生成する。 $x_{1:T} = x_1, x_2, \dots, x_T, T = t_{max}$  とし、 $a_2, a_3, a_4, a_5, a_6$  はそれぞれの確率の重みづけを示す。ただし、 $a_2 + a_3 + a_4 + a_5 + a_6 = 1$  とする。ここで、

$$O(x_{1:T}) = \sum_{t=1}^T \log P(x_t) \quad (4)$$

$$P(x_t) = a_2 P_2(x_t) + a_3 P_3(x_t) + a_4 P_4(x_t) + a_5 P_5(x_t) + a_6 P_6(x_t) \quad (5)$$

### 3.2 編集用素材の生成

以下の処理を行い、編集用素材として用いた。まず、撮影した各動画の音声信号と、ダンスパフォーマンスに使用した楽曲の音響信号との相互相関関数を求めた。相関が最大となる位置で動画を切り出すことにより、全ての動画を同期させて時間軸を揃えた。次に、楽曲中のダンスパフォーマンス使用区間を手動で切り出した音響信号に対して、既存の拍検出手法 [6] を適用して、編集単位となる拍を求めた。各カメラ動画中のダンサーの骨格情報は、OpenPose Library [8] を使用して取得した。検出された骨格の位置情報の x 軸方向における最大値と最小値の差と、y 軸方向における最大値と最小値の差を掛け合わせた値をダンサーが占める面積とし、その画面上に締める割合を算出する。その割合が 0.1 から 1.0、平均 0.5 から 0.6 の値に収まるように、パフォーマンスの鼻と、パフォーマンスの両足のかかとの位置の平均位置を中心にして映像をクロップする前処理をした。

### 3.3 ヒートマップ提示手法

拍毎のカメラの選択優先度は、以下の通りに決定した。3.1.2 節で作成した確率モデルを用いて、あり得るカメラ遷移のサンプリングを行うことで求める。まず評価関数を考慮せず、ランダムにサンプルする場合を考える。ここである時刻  $t = t'$  においてどのカメラが何回選ばれるかを考える。サンプル  $i = 1, \dots, 1000$  の時に  $t = t'$  でカメラ  $c(i)$  が選ばれたとすると、カメラ  $n$  が  $t = t'$  で選ばれる回数は、

$$\sum_{i=1}^{1000} g^n(c(i))$$

ただし、

$$g^n(c(i)) = \begin{cases} 1 & (c(i) = n) \\ 0 & (otherwise) \end{cases} \quad (6)$$

しかし、実際には評価関数の値が大きいものほど選ばれやすいので、この値を評価関数の値の大きさに

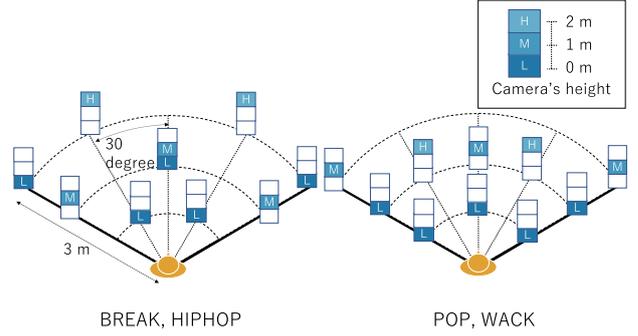


図 2. カメラのレイアウト

応じて重み付けしたものが、各カメラの選ばれやすさに比例すると考えられる。重み付けしたものを  $S_{t'}^n, i = 1, \dots, 1000$  のそれぞれの評価関数の値を  $O(i)$  とすると、

$$S_{t'}^n = \sum_{i=1}^{1000} g^n(c(i)) \exp(O(i)) \quad (7)$$

これを正規化することで、各カメラが選ばれる確率

$$p(x_{t'}^n) = \frac{S_{t'}^n}{\sum_n S_{t'}^n} \quad (8)$$

が得られる。この値を選択優先度として用いた。

## 4 評価実験

### 4.1 実験手続き

3章で提案した、多視点から撮影された複数の動画から魅力的な動画を生成する機能が有効であるかを調査するために、専門家による評価実験を行なった。編集用素材として4名のダンサー(男性2名、女性2名)のダンスパフォーマンスを撮影した。ダンサーのジャンルは異なり、動作の激しい順に BREAK, HIPHOP, WACK, POP である。ジャンル毎に基本的な動作を組み合わせた振り付けを作成してもらった。使用した曲のBPMは109, 129, 93, 92である。スピーカーはBose社のComputer MusicMonitorを用いた。撮影には、GoPro社のHERO5 Sessionを10台使用した。カメラの配置図を図2に示す。それぞれのカメラは、被写体が画面の中心の位置に来るよう調整した。被写体とカメラの間に他のカメラが入ってしまうことを防ぐために、配置を変えて2回に分けて撮影を行なった。

それぞれのダンス動画から7本ずつ動画を用意する。今回、システムに用いた係数を考慮すると、組み合わせが爆発的に増えてしまうために、あらかじめ実験的に数値を決定した。それぞれの係数が、 $a_2 : a_3 : a_4 : a_5 : a_6 = 3 : 2 : 5 : 3 : 3$  の比率になるよう設定した。1本目は原則に合わせた制約にか

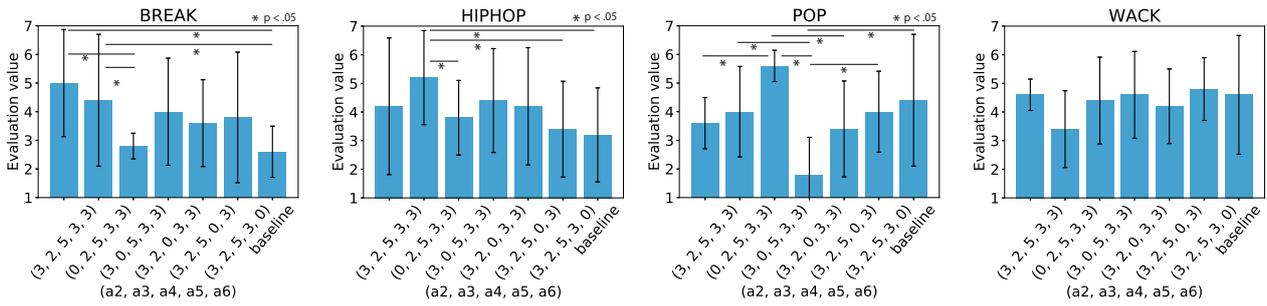


図 3. ダンサー毎のアンケート評価結果

かる係数  $a_2 \sim a_6$  を上記の式の数値を用いた。2~6本目は、それぞれの原則に合わせた制約にかかる係数を0にしたもの、7本目はベースラインとして

1. 各カメラの映像の全てのフレームのオプティカルフローの平均値を算出し、1拍毎の平均をとる。
2. 各カメラ毎にその平均値について正規化を行う。
3. カメラ内で相対的に動作の大きい場面が選ばれやすいように、1拍毎に正規化されたオプティカルフローの平均値が最大となるカメラを選択する。

のアルゴリズムで動作する簡易的な自動編集システムにより生成した。以上の方法で4名のダンサーのダンス動画から7本ずつ、計28本の動画を用意した。

評価者は、動画制作を業務とする、もしくは動画制作に関する教育を受けた専門家10名(男性8名、女性2名)である。動画制作に携わってから1~20年、平均8.6年の経験を有する。評価者1名に付き、ダンサー2名分の動画計14本をランダムに鑑賞し、リカット尺度に従って7段階で評価した。質問に対して非常に同意できる場合は7を、全く同意できない場合は1を回答する。評価項目は編集全体を通しての良さを測る項目として以下の質問を行なった。

Q: かつこいい編集だと思う。

本稿における魅力的な動画とは、専門家が映像を見てかつこいいと思う動画である。魅力的な動画かどうかを評価するために、実際に専門家がかつこいいと思う動画であるかを検証した。

アンケートはGoogleFormを用いて行なった。アンケートの回答中であっても自由に休憩がとれるよう配慮した。また、動画は何度でも見直しても良く、前の動画に戻って再評価を行うことも可能とした。回答終了後にシステムについてのフィードバックをもらった。

## 4.2 結果と考察

今回、すべての係数が代入されている場合、つまりすべての動画の編集原則を満たした場合、最も高い評価を得ると予想した。ダンサー毎のアンケート評価結果について図3に示す。縦軸は評価の平均値、縦線は標準偏差、横軸は係数の組み合わせを示しており、一番左側がすべての係数が0以外の数値の際、中間は1つの係数が0の際、一番右側 baseline が、前節で述べたベースラインとして用いたアルゴリズムを利用した際の結果である。BREAKについて、1要因分散分析を行なったところ、有意差が確認できた ( $F_{(6,24)} = 2.81, p < .05$ )。次にLSD法を用いて多重比較を行なったところ、4つの組み合わせにおいて有意差を確認した ( $p < .05$ )。BREAKにおいては、すべての係数を代入した際が最も良い編集であることがわかる。HIPHOPについて、1要因分散分析を行なったところ、有意な傾向が見られた ( $F_{(6,24)} = 2.43, p < .10$ )。次にLSD法を用いて多重比較を行なったところ、3つの組み合わせにおいて有意差を確認した ( $p < .05$ )。HIPHOPにおいては、カットの最大長、最小長を考慮する制約を無効にした際が最も良い編集であることがわかる。POPについて、1要因分散分析を行なったところ、有意差が確認できた ( $F_{(6,24)} = 2.81, p < .05$ )。次にLSD法を用いて多重比較を行なったところ、6つの組み合わせにおいて有意差を確認した。POPにおいては、構図の異なるカメラへの切り替えを考慮する制約を無効にした際が最も良い編集であることがわかる。WACKについて、1要因分散分析を行なったところ、有意差は確認できなかった。

以上より、4つのうち3つのジャンルにおいて、原則をすべて満たした場合もしくは制約を1つ抜いた場合のシステムを用いた編集結果の方が有意に評価が高かった。このことから、3章で提案した自動編集機能は、パフォーマンスのジャンルによっては有効である。一方で、POPについてはベースラインの場合の方が他の場合と比べて高い傾向が見られたり、WACKについては有意な差が見られないなど、

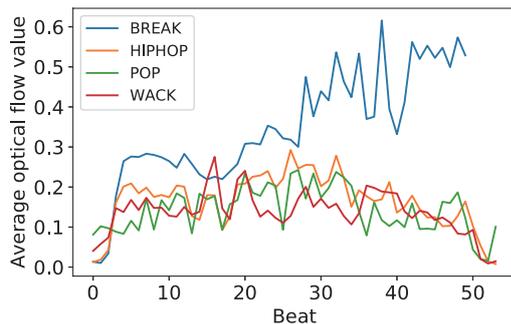


図 4. 拍毎のオプティカルフロー平均値

ダンスのジャンルによっては良い編集ができていない場合がある。これについて、各ダンサーの動作ごとのオプティカルフローの平均値(図4)を分析した。縦軸はオプティカルフローの1拍毎の平均値、横軸は拍を示している。POPはロボットダンスのような、他のジャンルと比べて、比較的止まる動きが多いダンスである。そのため、オプティカルフローは、断続的に値が急激に低くなっている。本システムでは、動作の激しさを緩急として捉えて、オプティカルフローの中央値で場合分けを行なっていたが、これでは断続的に緩急があるとシステムが認識してしまい、他の制約に影響を及ぼす可能性が高い。また、比較的細かい動作が多いために、アップショットが少ないと全体像がわかりにくいといった問題がある。これら影響を及ぼす可能性のある制約にかかる係数を調整するなど、対応が必要である。

評価実験より3章で提案した自動編集機能が有効であることは確認できたが、ダンスのジャンルによっては必ずしもうまく対応できていない場合があり、自動編集後にユーザによる手直しが必要であると考えられる。本研究の提案システムは、ユーザが自分好みに編集できることに加え、こうした自動編集結果で不十分な点をインタラクティブに修正できる利点を持つ。提案システムでは、多数の自動編集結果を用いることで、自動編集機能が選択しやすい映像をヒートマップとして提示した。これにより、編集技術をもたないダンサーでも、動画編集の原則に則した編集ができると考えられる。実際に提案システムが使用した結果、提案した機能が適切に動作することを確認できた。今後、ユーザスタディを通して本研究で提案するシステムのインタラクションデザインについての評価を行う。

## 5 まとめ

本稿では、動画編集の専門的な知識を持たなくとも簡単に魅力的なダンス動画を作成できる、多視点ダンス映像のためのインタラクティブ編集システム

を提案した。本システムでは、ダンス動画の自動編集機能を活用し、ヒートマップとして拍毎のカメラの選択優先度を提示することにより、誰でも動画編集の原則に沿った編集が行える。評価実験より、自動編集機能の有効性を示し、インタラクティブなダンス動画編集システムの必要性を確認した。

今後、小型カメラやスマートフォンに付属しているカメラなどが、アプリを通して簡単に同時制御できるようになり、多視点映像を編集する機会が増えてくると予想している。例えば、ダンスイベントなど、多くの人が自身のカメラでダンスを撮影している状況において、個々のカメラを1つのアプリを通して繋ぎ、多視点映像と捉えれば、多くの動画素材を入手できる。これら大量の動画から、ユーザの好みに合わせてカスタマイズし、瞬時に編集された動画を保存可能にするシステムへの応用が考えられる。

本研究の一部はJST ACCEL (JPMJAC1602)の支援を受けた。また、本稿の内容はACE2017に採録決定済みである[9]。

## 参考文献

- [1] M. Leake, A. Davis, A. Truong, and M. Agrawala. Computational Video Editing for Dialogue-Driven Scenes. In *ACM Transactions on Graphics (TOG)*, Vol. 36, No. 130, 2017.
- [2] A. Pavel, D. B. Goldman, B. Hartmann, and M. Agrawala. VidCrit: Video-based Asynchronous Video Review. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, pp. 517–528, 2016.
- [3] I. Arev, H. S. Park, Y. Sheikh, J. Hodgins, and A. Shamir. Automatic Editing of Footage from Multiple Social Cameras. In *ACM Transactions on Graphics (TOG)*, Vol. 33, No. 81, 2014.
- [4] J. C. Foust, E. J. Fink, and L. S. Gross. Video Production: Disciplines and Techniques, p. 230. Taylor and Francis, 2012.
- [5] C. J. Bowen. Grammar of the Edit, p. 63, p. 211, p. 212. Taylor and Francis, 2013.
- [6] S. Böck, F. Krebs, and G. Widmer. Joint Beat and Downbeat Tracking with Recurrent Neural Networks. In *Proceedings of the 17th International Society for Music Information Retrieval Conference*, pp. 255–261, 2016.
- [7] G. Farnebeck. Two-frame motion estimation based on polynomial expansion. In *Image analysis*, pp. 363–370, 2003.
- [8] Z. Cao, T. Simon, S. E. Wei, and Y. Sheikh. Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2017.
- [9] S. Tsuchida, S. Fukayama, and M. Goto. Automatic System for Editing Dance Videos Recorded using Multiple Cameras. ACE 2017(採録決定済)