

キーポーズを用いた舞踊動作の低次元化表現

Manoj Perera[†] ○工藤俊亮[†] 白鳥貴亮[‡] 池内克史[†]
[†] 東京大学 [‡] Carnegie Mellon University

あらまし 本論文では、舞踊動作を効果的に低次元空間中に表現する手法について提案する。本手法ではキーポーズに注目する。キーポーズとは「とめ動作」とも呼ばれ、これをタイミング良く正確に再現することが舞踊を自然で美しく見せるために重要であるとされる。このキーポーズを考慮した次元圧縮手法により動作を低次元化することにより、より人間の目に自然に見える次元圧縮が可能になった。

Low-dimensional Representaion for Dance Motion Based on Keyposes

Manoj Perera[†] ○ Shunsuke Kudoh[†] Takaaki Shiratori[‡] Katsushi Ikeuchi[†]
[†] The University of Tokyo [‡] Carnegie Mellon University

Abstract This paper proposes a method for representing dance motion effectively into a low-dimensional space. We focus on *keyposes*. Keyposes are often called *motion pauses*, and it is considered important that dancers represent keyposes precisely in appropriate timing. We use the keyposes to reduce dimension of motion. Comparing to ordinary methods, this method can generate natural dance motion from the viewpoint of human.

1 はじめに

近年キャラクターアニメーションへの需要は大きく高まっており、人間の動作を生成・編集するための手法も数多く開発されている。そのような手法の中に、動作を低次元の空間で表現することによって、編集などの取り扱いを容易にするいは効果的にしようとするものがある [1, 2]。一般に人間の姿勢を表現するには 30 から 50 程度の自由度が必要であり、極めて高次元の量を扱う必要がある。しかしそれらの自由度はかなりの部分強い相関関係をもっており、効果的な次元圧縮によって動作の本質的な部分を取り出すことができれば、容易かつ効果的な生成・編集等が可能になる。

我々のグループでは、これまで舞踊動作を対象として「キーポーズ」という概念を提案し、CGアニメーションの生成やロボットによる舞踊動作模倣などの研究を行ってきた [3, 4, 5]。キーポーズとは舞踊動作中でも特に重要とされる「とめ動作」に相当し、これをきちんと再現することが、「上手に踊る」ための必要条件となるもので

ある。本論文では、このキーポーズを用いて舞踊動作を効果的に低次元空間で表現する手法を提案し、キーポーズが動作の低次元化表現にも有効であることを示す。

キーポーズという概念において重要な観点として、それが「人間にとって重要な」舞踊中の動作であるという点がある。実際に、数値的には全体として元動作に近い再現された動作と、キーポーズのみ集中的に元動作に近付けその他の部分は元動作から比較的遠い再現動作を比較した場合、人間の目には後者の方がはるかに自然に見える。今回提案する手法においてもそういった観点を重視し、手法の評価において、数値的な比較のみならず結果を人間が見比べる感性実験も行った。

本論文の構成は、まずキーポーズの抽出について述べ、次にそれを用いた舞踊動作の低次元化表現について述べる。最後に提案手法に関する実験とその結果について考察する。

2 キーポーズの抽出

舞踊では要所要所において、全身の動作を一時的に停止させる「とめ動作」と呼ばれる姿勢が現れる。舞踊者の間では、これをタイミング良く正確に再現することがとりわけ重要と考えられている。我々はこれまでの研究において、このとめ動作をキーポーズとして抽出する手法を考案した [3].

これまでのキーポーズ抽出手法においては、動作の停止を単純に手先軌道の速度のみによって判断していた。しかしたとえば上体に対して手の位置を固定したまま平行移動する場合などのように、踊り手の意識としては「とめ」であるにもかかわらず、単純な速度による比較では「とめ」と認識できないケースが存在することが分かってきた。そこで動作の停止をより正確に判断するため、エネルギー関数による判定手法を新たに開発した。

本節ではこの新しい判定手法を解説したあと、音楽情報の解析について述べ、最後にこれらの情報からキーポーズを抽出する手法について述べる。

2.1 動作情報の解析

本研究では、モーションキャプチャシステムを用いて踊り手の動作を取得する。使用するマーカー数が N 個の時、各マーカーの位置・速度を時間 t の関数として

$$\mathbf{x}^i(t), \mathbf{v}^i(t) \quad (i = 1, 2, \dots, N) \quad (1)$$

と表すことにする。

動作情報の解析には、2つのエネルギー関数 $E_G(t)$, $E_L(t)$ を用いる。どちらも踊り手の動きの大きさを示す指標であるが、前者は環境に対してのグローバルな動きの大きさを示し、後者は体中心に対してのローカルな動きの大きさを示す。これらは以下のように書き表される。

$$E_G(t) = \left| \sum_{i=1}^N w_i \mathbf{v}^i(t) \right| \quad (2)$$

$$E_L(t) = \left| \sum_{i=1}^N w_i (\mathbf{v}^i(t) - \mathbf{v}^c(t)) \right| \quad (3)$$

ただし \mathbf{v}^c は体中心（腰部の中心位置）の速度を表し、 w_i は各マーカーごとに定められた重み係数を表す。

動作が完全に停止した場合どちらの指標も 0 となるが、実際の動作ではそうなることはありえない。そこで2つのエネルギー関数それぞれについて極小値を求め、それが閾値以下である場合に動きが停止しているとみなすことにする。

このようにして得られた姿勢は「動作が止まる姿勢」という意味においての「とめ動作」であるが、本手法ではこれをそのままキーポーズとしては用いずに、次節で述べるように音楽情報の解析による結果に基づいて絞り込みを行ったものをキーポーズとして用いる。

2.2 音楽情報の解析

理論的にはキーポーズの抽出には動作データから全身が静止した瞬間を取り出せば良いはずであるが、実際には前小節の手法で抽出されたとめ動作は、人間の目から見ると過剰であることが分かった。これは踊り手の意識としては動きをとめたつもりはなくても、数値的にはエネルギー関数が 0 に近くなる場合があることによるものである。そこで動作データの解析に加えて音楽情報の解析も行い、とめ動作が音楽のリズムと一致したところをキーポーズとすることで、キーポーズを絞り込んだ。

音楽情報の解析は、後藤らの手法 [6] を用いてスペクトルパワーの変化に基づき行った。ある時刻 t , 周波数 f において、スペクトルパワーの増加分を表す関数 $d(t, f)$ を次のように定義する。

$$d(t, f) = \begin{cases} \max(p(t, f), p(t+1, f)) - p' \\ (\min(p(t, f), p(t+1, f)) \geq p') \\ 0 \quad (\text{otherwise}) \end{cases} \quad (4)$$

$$p' \equiv \max(p(t-1, f), p(t-1, f \pm 1))$$

ここで $p(t, f)$ は時刻 t , 周波数 f におけるスペクトルパワーを表す。次にこの $d(t, f)$ に対し、周波数領域全体に関しての和 $D(t) = \sum_f d(t, f)$ を定義する。これは時刻 t に発せられた音の強さを表す。続いて $D(t)$ の自己相関関数を考えるこ

とにより、リズムの平均間隔を推定する。さらに推定されたリズム間隔からパルス列を生成し、 $D(t)$ との相互相関関数を考えることで、リズムの開始時刻を推定する。こうして求められたリズム情報を、キーポーズの絞り込みに用いる。

2.3 動作情報と音楽情報の統合

音楽のリズムを用いたキーポーズ候補点の絞り込みとして、音楽のリズム時刻の前後一定時間内にキーポーズ候補点が存在するかを確認し、存在する場合はそのリズム時刻を候補点とし、存在しない場合は候補から外す処理を行う。

この処理により会津磐梯山踊りから抽出されたキーポーズの例を図1に示す。上段はエネルギー関数と音楽のリズム、抽出されたキーポーズ時刻であり、中段はキーポーズと認識された姿勢をCGモデルで表したものであり、下段は舞踊の師範が弟子を指導する目的で描いたテキストのイラストである。抽出されたキーポーズと師範のイラストの間には極めて強い一致が観察される。図に示した例以外でも、ほぼすべてのキーポーズが師範によるスケッチの図と過不足なく一致した。師範によるイラストは舞踊者がどのように舞踊を捉えているかを表しており、これと抽出されたキーポーズが一致しているということは、本手法により抽出されたキーポーズが舞踊の本質を十分に表している証拠と考えられる。

3 動作の低次元化

動作の低次元化には主成分分析 (PCA) を用いる。その際、低次元空間での基底を求めるのに、前節の手法により求められたキーポーズを用いるのが本手法の特徴である。

時刻 t における踊り手の姿勢 $\mathbf{P}(t)$ は、以下のようなベクトルとして定義される。

$$\mathbf{P}(t) = (\mathbf{x}^1(t)^T \mathbf{x}^2(t)^T \dots \mathbf{x}^N(t)^T)^T \quad (5)$$

このうちキーポーズとして抽出された姿勢が M 個あったとすると、それらを

$$\mathbf{K}_1, \mathbf{K}_2, \dots, \mathbf{K}_M \quad (6)$$

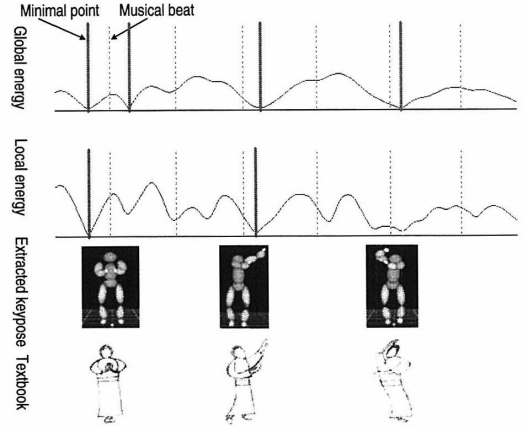


図1: 師範のスケッチと抽出されたキーポーズ (会津磐梯山踊り)

とする。これらに関してPCAを用いた次元圧縮により、 d 次元部分空間 Σ_d へ射影することを考える。

$$\bar{\mathbf{K}} = \sum_{i=1}^M \mathbf{K}_i \quad (7)$$

とした時、行列 $(\mathbf{K}_1 \mathbf{K}_2 \dots \mathbf{K}_M)$ の共分散行列に対し、固有値の大きい順に d 個の固有ベクトルからなる行列 W_d を用いることにより、

$$\mathbf{K}_i^d = W_d(\mathbf{K}_i - \bar{\mathbf{K}}) \quad (\mathbf{K}_i^d \in \Sigma_d) \quad (8)$$

のように、 \mathbf{K}_i を d 次元部分空間 Σ_d へと射影することができる。

このようにして得られた射影行列 W_d を、すべての姿勢 $\mathbf{P}(t)$ に対して用いて低次元化を行う。すなわち、 $\mathbf{P}(t)$ を d 次元に圧縮したベクトル $\mathbf{P}^d(t)$ は以下のようにして計算される。

$$\mathbf{P}^d(t) = W_d(\mathbf{P}(t) - \bar{\mathbf{K}}) \quad (9)$$

逆に、圧縮された姿勢を通常の姿勢に戻すには、 W_d の転置行列を掛ければ良い。

$$\mathbf{P}^l(t) = \bar{\mathbf{K}} + W_d^T \mathbf{P}^d(t) \quad (10)$$

PCAの性質により射影行列 W_d は、元の姿勢と再現された姿勢の差 $\sum^M (\mathbf{K}_i - \mathbf{K}_i^l)^2$ が最小になるような行列となっている。

4 実験と考察

本研究では踊り手の動作を取得するために、VICON 社製の光学式モーションキャプチャシステムを利用した。踊り手の体に装着したマーカー数 N は 33 であり、それぞれのマーカーについて 3 次元位置が得られる。対象動作として会津磐梯山踊りを選び、師範を含む複数の踊り手によるそれぞれ複数回の舞踊を計測した。

この動作についてキーポーズの抽出を行った結果は、2.3 節に示したとおりである。会津磐梯山踊りは、師範の手によるテキストによれば、8 個の決め動作からなる一連の動作を繰り返すという構造になっており、本手法を用いることでそのうち 6 個の決め動作をキーポーズとして抽出することができた。余分な動作はいっさい抽出されなかった。抽出に失敗した 2 個の姿勢については、テキストでは決め動作となっているものの、実際の舞踊においては前後を含めた一連の動作のように滑らかな動きで表現されており、体の動きの停止に基づいた本手法では抽出することができなかったものと思われる。

このようにして抽出されたキーポーズを用いて計算された射影行列により、動作を低次元化したのが図 2、図 3 である。ここでは 3 次元まで次元を落して表示している。図 2 はキーポーズのみを射影したものであり、図 3 は踊り手の一人について全ての動作を射影した結果である。

次元として 3 次元を選んだのは、人間の目から見て不自然ではない動作が再現可能な最小の次元が 3 次元だったからである。これより低い次元では情報の欠落が大き過ぎ、再現された動作が元の動作と大きくかけ離れてしまった。寄与率等の数値的指標によらずに見た目によって次元を選んだのは、はじめに述べた通り、人間の感性において自然な動作の記録手法を考えるのが本研究の目的になっているからである。

比較のために、キーポーズと同じ個数の姿勢を等間隔にサンプリングしたものをを用いて射影行列を作成し、同様に 3 次元空間に次元圧縮した結果を図 4 に示す。前のキーポーズに基づく次元圧縮の結果と比べると、姿勢の分布にばらつきが目立つことが分かる。このことは、キーポーズ

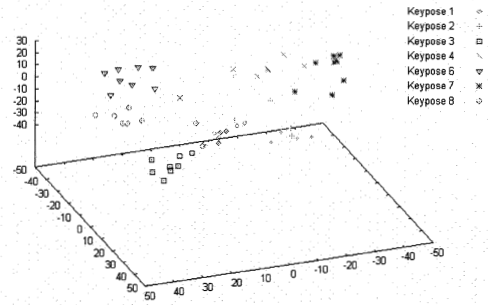


図 2: 3 次元空間に射影されたキーポーズ

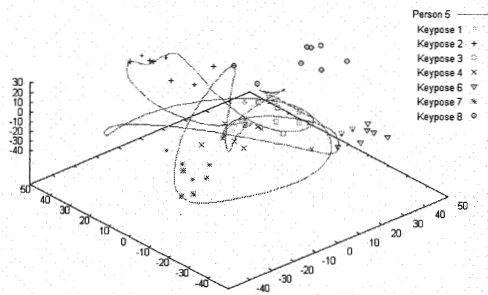


図 3: 3 次元空間に射影された舞踊動作

ズに基づく次元圧縮が、舞踊全体の中から特徴的な部分をきちんと抽出しつつ次元圧縮を行っていることの、ひとつの証拠であると言える。

両手法における寄与率を比較すると、表 1 のようになる。これは共分散行列の固有値 λ_i (大きい順に整列) に関して、

$$\rho = \frac{\sum_{i=0}^d \lambda_i}{\sum_{i=0}^M \lambda_i} \quad (11)$$

の値をそれぞれ計算したものである。これを見ると、キーポーズに基づく場合も等間隔サンプリングの場合も、寄与率に関してはほとんど差がないことが分かる。

また動作の再現に関する性能について、キーポーズに基づく手法と等間隔サンプリングによる手法の比較を行った。同じ動作 P_0 について、キーポーズに基づく手法により次元圧縮を行ったのちに復元した動作 P'_0 と、等間隔サンプリングによる手法により次元圧縮を行ったのちに復

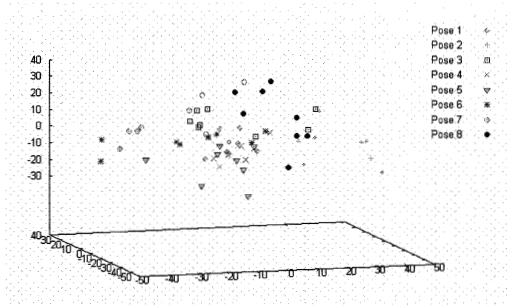


図 4: 等間隔にサンプリングされた姿勢から生成された行列による 3 次元空間への次元圧縮

dim.	1	2	3	4	5
keypose	34.0	58.2	70.3	80.1	88.8
uniform	27.1	50.9	62.2	72.4	79.3

表 1: キーポーズに基づく手法と等間隔サンプリングによる手法における次元ごとの寄与率 (%)

元した動作 P_0'' について、元の動作との差

$$d' = \sum_t |P_0' - P_0|, \quad d'' = \sum_t |P_0'' - P_0| \quad (12)$$

を計算したところ、大きな差はみられなかった ($(d'' - d')/d' \sim 4\%$)。

これらの寄与率に関する検証と動作の再現性に関する検証の結果から、数値的に見れば、両手法における再現性に大きな差は見られないということがいえる。

次にこれらの結果が人間の感性という面から見てどのような違いがあるかということを探るため、感性実験を行った。図 5 のように、元の動作と両手法によって生成された動作の 3 つを並べた映像を用意し、被験者に A と B とのどちらの動きがより「自然」かを選ばせた。何をもって「自然」とするかは定義は特に与えず、被験者が自由に判断して良いことにした。動作はすべて会津磐梯山踊りを使用した。異なる踊り手による 3 種類の舞踊動作について実験した。被験者については、東京大学生産技術研究所のオープンハウスへの来訪者からランダムに 34 人を選んで協力を依頼した。その結果、表 2 のように、9 割以上の確率でキーポーズに基づく手法により生成された動作の方が自然であるという

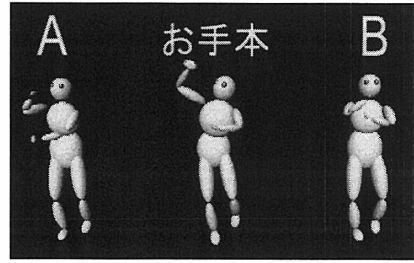


図 5: 感性実験に用いた映像

	dance 1	dance 2	dance 3	total
ratio	32/34	29/34	31/34	92/102

表 2: 感性実験によりキーポーズに基づく手法による動作の方が「自然」だと感じた人の割合

回答が得られた。

以上の結果から、キーポーズに基づく低次元化表現においては、数値的には単純な等間隔サンプリングによる手法とほとんど差がないにもかかわらず、人間の目から見た時に有意に自然に見える動作が生成可能であることが分かった。これは、キーポーズが踊りにとってとりわけ重要な瞬間を示しているため、これを集中的に考慮した手法により、踊りの本質的な情報をより多く保存することが可能になったためと思われる。

5 まとめ

本論文では、キーポーズに注目することによって、舞踊動作を効果的に低次元空間中で表現する手法について提案した。キーポーズに基づいて張られた空間中で表現された動作は、数値的には単純なサンプリングによる方法と大差ないものであったが、感性実験の結果、人間の目には有意に自然なものに見えることが判明した。

謝辞

本研究は科学技術振興機構戦略的基礎研究事業 (CREST) デジタルメディア作品の製作を支援する基盤技術の支援を受けて行われた。

参考文献

- [1] Keith Grochow, Steven L. Martin, Aarno Hertzmann, and Zoran Popvić. Style-based inverse kinematics. *ACM Transactions on Graphics (SIGGRAPH 2004)*, Vol. 23, No. 3, pp. 522–531, 2004.
- [2] Alla Safonova, Jessica K. Hodgins, and Nancy S. Plooard. Synthesizing physically realistic human motion in low-dimensional, behavior-specific spaces. *ACM Transactions on Graphics (SIGGRAPH 2004)*, Vol. 23, No. 3, pp. 514–521, 2004.
- [3] 白鳥貴亮, 中澤篤志, 池内克史. モーションキャプチャと音楽情報を用いた舞踊動作解析手法. 電子情報通信学会論文誌 D-II, Vol. J88-D-II, No. 8, pp. 1583–1590, 2005.
- [4] Takaaki Shiratori, Atsushi Nakazawa, and Katsushi Ikeuchi. Dancing-to-music character animation. *Computer Graphics Forum*, Vol. 25, No. 3, pp. 449–458, Sep 2006.
- [5] Shin'ichiro Nakaoka, Atsushi Nakazawa, Fumio Kanehiro, Kenji Kaneko, Mitsuharu Morisawa, Hirohisa Hirukawa, and Katsushi Ikeuchi. Learning from observation paradigm: Leg task models for enabling a biped humanoid robot to imitate human dances. *The International Journal of Robotics Research*, Vol. 26, No. 8, pp. 829–844, Aug 2007.
- [6] Masataka Goto. An audio-based real-time beat tracking system for music with or without drum-sounds. *Journal of New Music Research*, Vol. 30, No. 2, pp. 159–171, 2001.