

## Time-Varying Mesh の類似動作検索

山崎俊彦, 相澤清晴

東京大学大学院情報理工学系研究科電子情報学専攻

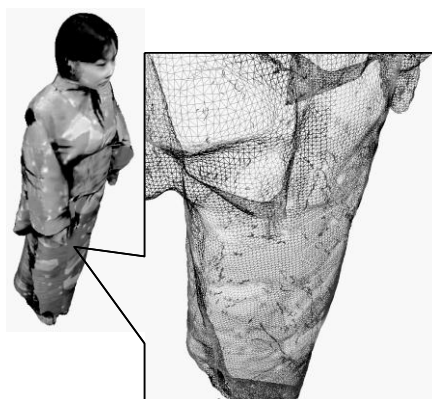


図 1. TVM の例。

**目的:** 本研究の目的は、実写 3 次元映像表現の一つである Time-Varying Mesh (TVM) を対象として、効率よく類似動作の検索を行うことである。TVM は、複数台のカメラから撮影された多視点映像を元に、オブジェクトの大まかな 3 次元(凸)形状を求める視体積交差法と、オブジェクトのへこみ部分や衣服の皺などを忠実に再現するためのステレオマッチング法などの組み合わせによって作られる、3 次元メッシュシーケンスである[1]-[5] (図 1 参照)。TVM をはじめとするこのような実写 3 次元映像は、従来の 3 次元 CG やモーションキャプチャなどとは異なり、実物体の 3 次元形状や動きを忠実に記録・再生できる点が特徴である。TVM 生成技術を利用して、能などの伝統舞踊を 3 次元アーカイブ化する試みがすでになされており[4]、今後様々な TVM のデータベースが構築されたとき、それを効率的に運用するためには検索・編集などの基礎技術が必要不可欠である。

しかしその一方で、TVM は受動素子であるカメラ映像のみから作成され、また被写体が非剛体であることから、フレーム間で頂点や結線情報の対応関係はとらず、1 フレームずつ独立にメッシュを生成することが多い。生成される 3 次元メッシュデータには関節の位置がどこか、などといった構造的な記述はなく、多数の頂点の位置と結線情報が無作為に記述されているだけである。また、フレーム間で頂点同士の明示的な対応関係もとられないため、仮にある 1 フレームにおいて人手で関節の位置などを指定したとしても、それに続くフレームにおいてその位置を追跡できない。そのため、TVM では構造的な特徴点の特定、ならびに追跡が特に困難で、様々な TVM の処理を困難にしている。ごく最近になって、すべてのフレームにおいて頂点の対応関係を取り、頂点数と結線情報に一貫性を持たせながら 3 次元メッシュシーケンスを生成する手法も提案されているが[6][7]、これらはリファレンスフレームとして、予めレーザスキャンなどで人体の形状を詳細に取得しておかなければならないなどの課題が残っている。

そこで本研究では、構造的な特徴点が記述されない TVM に対し、類似動作の検索を行う手法を開発する。TVM からどのような特徴量を抽出し、またどのようにしてシステムにクエリを与えるかが要となる。本論文では Query-by-Example (キーワードで検索するのではなく、例となる動作シーケンスを何らかの形で与えて検索する)を基本とし、クエリとして別の TVM を用いるもの[8]-[15]、モーションキャプチャ・データを用いるもの[16]-[23]、2 次元動画像を用いるもの[24]を提案する。また、モーションキャプチャを用いる手法に関しては、モーションキャプチャ・データを TVM のデータに近づけて動作の類似度を評価する手法と[16]、TVM からモーションキャプチャ互換の動きデータを抽出して検索する手法[17]-[23]について述べる。TVM、モーションキャプチャ、2 次元動画像と、順

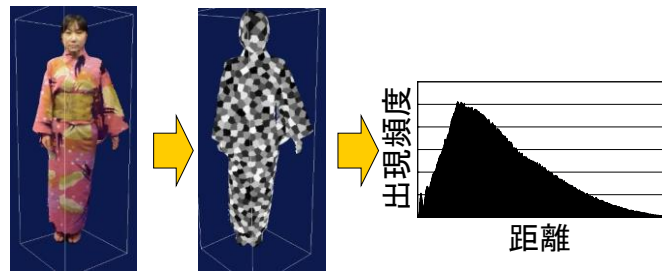


図 2. Modified Shape Distribution 法による形状特徴ベクトル抽出

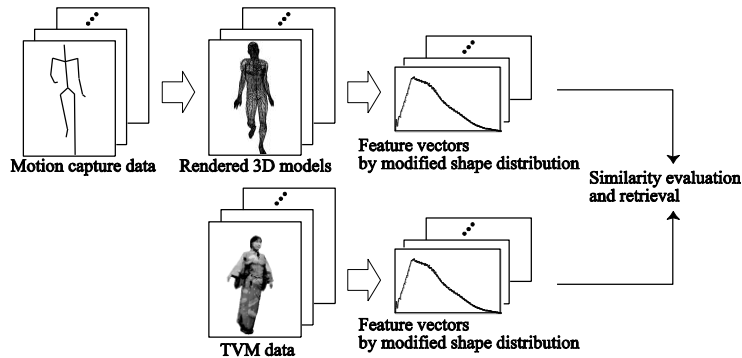


図 3. モーションキャプチャ・データから TVM 互換の動き特徴量を抽出する場合の検索処理の流れ。

にシーケンスの生成コストは低くなるが、元の TVM との共通的な性質が少なくなるために、処理は逆に難しくなる。しかし、いずれの場合においても 60-90%程度の高い精度で検索を行える手法を確立できた。

**方法:** TVM の類似動作検索を実現するために、下記に示す 4 種類の方式を開発した。

#### ①別の TVM をクエリとする方法

TVM から、各フレームの姿勢の特徴を記述できる特徴ベクトルを抽出し、これにより動作の類似度を評価する手法である。これについては、静止 3 次元メッシュモデルの形状特徴記述アルゴリズムとして有名な Shape Distributions 法[25]を頂点のクラスタリングによって安定化させた Modified Shape Distribution 法[8]-[12]、3 次元点群のレジストレーションアルゴリズムとして有名な Iterative Closest Point (ICP)法を用いて対応頂点同士の距離を類似度とする方法[13]-[15]を開発した。概念図を図 2 に示す。シーケンスは、予め動きの大きさによってセグメンテーション（文節化）されて短いクリップに分割されているものとし、クリップ同士の類似度は Dynamic Programming (DP) マッチング法によって計算する。

#### ②モーションキャプチャをクエリとする方法 ～モーションキャプチャから TVM 互換の特徴量を抽出～

TVM を検索するために別の TVM を用いる手法は、新たな TVM を作成しなければならず、人的・時間的コストが膨大である。そこで、比較的設備の利用が容易で、またフリーのデータベースがインターネット上で公開されている [26][27]、モーションキャプチャ・データの利用を考える。

モーションキャプチャのデータをまず 3 次元メッシュに変換した後、①と同様の形状特徴を抽出することを考える (図 3)。まずモーションキャプチャのデータを一般的な 3 次元 CG レンダリングソフトを用いて 3 次元メッシュモデルのシーケンスに変換する。本稿では Poser を使用した。その後、TVM とレンダリングされた 3 次元 CG モデルの双方から特徴ベクトルを抽出する。これによって、①で開発した Modified Shape Distribution 法を用いて効率的に検索を実現することができる。

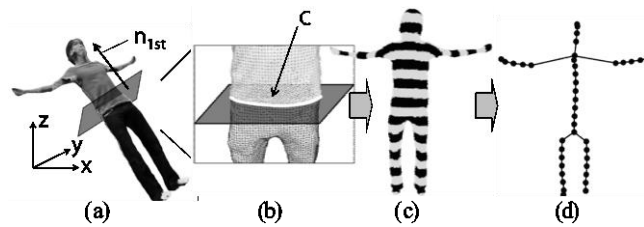


図4. TVMからのスケルトン構造抽出。

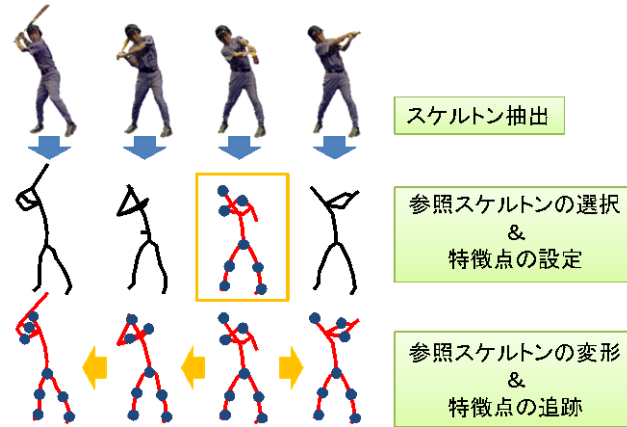


図5. TVMからモーションキャプチャ・データ互換の動き特徴量を抽出する場合の検索処理の流れ。

### ③モーションキャプチャをクエリとする方法 ～TVMからモーションキャプチャ互換の特徴量を抽出～

モーションキャプチャを TVM 互換の形式にする方法は、比較的容易に実現できるが、その反面モーションキャプチャのもつ構造的な特徴量が失われる。仮に TVM からモーションキャプチャと同等な、動き情報を抽出することができれば、検索の効率が格段に向上し、また今までに提案されてきた多数のモーションキャプチャの為の検索技術を利用できる。

TVM からモーションキャプチャ互換のデータを生成する手法[17]-[23]では、まず Reeb-graph と呼ばれるスケルトン構造抽出を行った。スケルトン構造は、メッシュ表面に何らかの連続する関数を定義し、一定値間隔でモデルを輪切りにし、それぞれの輪切りの重心を順に接続することにより求める。このようなスケルトン構造抽出は[28]など多くの手法が提案されているが、演算コストが高い上、人体モデルには適さないなどの問題があった。そこで、3次元モデルの頂点座標の主成分分析により、モデルの重心付近を通り第一主成分を法線ベクトルとするような平面によってモデルを切断し、切断面からの測地線距離を用いることで安定かつ高速な処理を可能にした。TVM からスケルトンを抽出する処理の流れを図4に示す。

得られたスケルトンを用いて、TVM の動きを追跡する。最初に、各フレームのスケルトンの中からユーザによって被写体の構造を最も良く表しているものが選択される。ここで選択されたスケルトン(参照スケルトンと呼ぶ)を、その他のフレームのスケルトン(目標スケルトンと呼ぶ)に対してフィッティングすることで動きの追跡が行われる。この処理においては、手先や足先など端点の対応付けを先に行い、その結果を元にスケルトン全体に対して線形補間によるフィッティングを行う。この処理はスケルトンの中心点(重心点から最も近いスケルトン上の点)と端点を結ぶパス毎に行う。端点の数は必ずしも一定ではないので、対応付けができない場合がある。このような場合は、目標スケルトンの周囲にエネルギー関数を定義して、ポテンシャル場を通過するコストが最小になるように参照スケルトンを変形することで対応する。一連の処理の流れを図5に示す。

検索は、得られた動きの情報から、いくつかの関節点の情報を取り出して行われる。腕・足・胴体に相当する部分の向きを3次元空間上のベクトルとして取り出し、それぞれのベクトルの距離を計算して足し合わせることで類似度

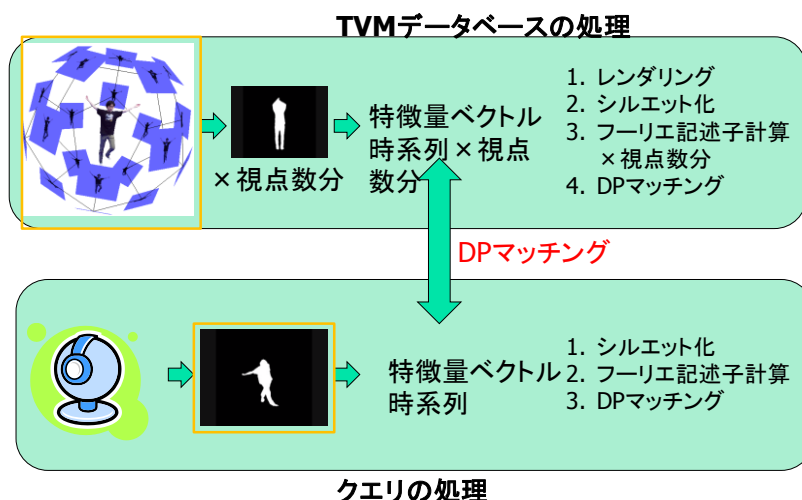


図6. 2次元映像をクエリとしたTVMの検索。

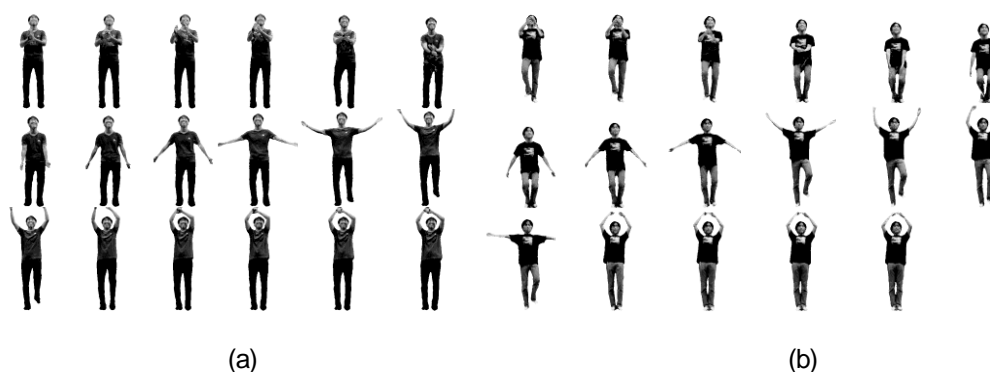


図7. TVMをクエリとした類似動作検索の例: (a) クエリ、(b) 検索結果。

とする。モーションキャプチャ・データからも同様のベクトルが取り出せるので、モーションキャプチャをクエリとした類似検索が可能となる。検索は①、②と同様DPマッチングに基づく手法を採用する。

#### ④2次元動画をクエリとする方法

②、③で述べたモーションキャプチャ・データを用いる手法は、①のTVMをクエリとする方法にくらべるとコストは低い。しかし、特殊なスタジオが必要であったりするなど、一般のユーザへの負担は依然として高い。そこで、Webカメラなどで撮影した映像をクエリとした、直感的かつ低コストな検索手法を提案する[24](図6)。具体的には、3次元静止モデルの検索手法[29]を応用し、TVMを異なる20視点からレンダリングした映像群とクエリ動画とのマッチングをとることにより実現する。マッチングは、両動画から人物のシルエット・シーケンスを抽出し、その形状特徴ベクトル(P型フーリエ記述子)のシーケンス間でDPマッチングを行った。

**結果:** 盆踊り、ラジオ体操などのシーケンスを用いて検索の実験を行った。TVMをクエリとした検索結果例を図7に、モーションキャプチャ・データをクエリとした検索例を図8に示す。図7では動作を行う人物が異なっても正しく類似動作検索が行えていることを示している。また、図8ではモーションキャプチャ・データをレンダリングした歩行シーンを用いて、TVMの歩行シーンを正しく検出出来ていることがわかる。TVM、モーションキャプチャ・データ、2次元映像を用いたクエリによる検索では、概ね60-90%の検索精度を実現している。両手を大きく開いて円を描くなど、大きな動きの場合は検索精度が高く、逆に体の前で手拍子をうつなど小さい動きは、形状特徴ベクトルを抽出する場合でも2次元映像を用いる場合でも検索が難しいことが明らかとなった。



(a)



(b)



(c)



(d)

図7. モーションキャプチャをクエリとした TVM の検索例: (a) クエリ、(b)-(d) 検索結果。

**結論:** 本稿では TVM、モーションキャプチャ・データ、2次元映像などをクエリとして、効率的に形状得手用ベクトルを抽出し、TVM を検索する手法について述べた。今後はこれまで提案してきた様々な手法の統合、さらにはインタラクティブな検索などについて検討を行っていく予定である。

#### 参考文献

- [1] T. Kanade, P. Rander, and P. Narayanan, "Virtualized reality: constructing virtual worlds from real scenes," IEEE Multimedia, vol. 4, no. 1, pp. 34-47, Jan/March 1997.
- [2] W. Matusik, C. Buehler, R. Raskar, S. Gortler, and L. McMillan, "Image based visual hulls," ACM SIGGRAPH2000, pp.369-374, 2000.
- [3] T. Matsuyama, X. Wu, T. Takai, and T. Wada, "Real-time dynamic 3-D object shape reconstruction and high-fidelity texture mapping for 3-D video," IEEE TCSVT, vol. 14, no. 3, pp. 357-369, March 2004.
- [4] K. Tomiyama, Y. Orihara, M. Katayama, and Y. Iwadate, "Algorithm for dynamic 3D object generation from multi-viewpoint images," Proc. SPIE, vol. 5599, pp. 153-161, 2004.
- [5] J. Starck and A. Hilton, "Surface Capture for Performance-Based Animation," IEEE Computer Graphics and Applications, Vol. 27, No.3, pp. 21-31, May-June 2007.
- [6] D. Vlasic, I. Baran, W. Matusik, and J. Popovic "Articulated Mesh Animation from Multi-view Silhouettes," SIGGRAPH08, #97, 2008.
- [7] E. de Aguiar, C. Stoll, C. Theobalt, N. Ahmed, H.P. Seidel, and S. Thrun "Performance Capture from Sparse Multi-view Video," SIGGRAPH08, #98, 2008.
- [8] T. Yamasaki and K. Aizawa, "Temporal 3D Video Segmentation Using Modified Shape Distribution," IEEE ICME2006, pp. 1909-1912, 2006.
- [9] T. Yamasaki and K. Aizawa, "Similar Motion Retrieval of Dynamic 3D Mesh Based on Modified Shape Distribution," Proceeding of Eurographics2006 short papers, pp. 9-12, 2006.
- [10] 山崎俊彦, 相澤清晴, "Modified Shape Distribution 法による 3D ビデオの類似動作," 情報科学技術レターズ, LI-011, pp. 201-204, 2006. (letter)

- [11] T. Yamasaki and K. Aizawa, "Motion Segmentation and Retrieval for 3D Video Based on Modified Shape Distribution," EURASIP Journal on Applied Signal Processing, Special Issue on Search and Retrieval of 3D Content and Associated Knowledge Extraction and Propagation, vol. 2007, Article ID 59535, 11 pages, 2007.
- [12] 山崎俊彦, 相澤清晴, "Modified Shape Distribution 法によるダイナミックメッシュの動きセグメンテーション," 映像情報メディア学会, vol. 61, no. 4, pp. 495-501, April 2007.
- [13] T. Yamasaki and K. Aizawa, "Motion Segmentation for 3D Video Based on Spherical Registration," 3DTV-Conference, #25, 2007.
- [14] 山崎俊彦, 相澤清晴, "3次元表面位置合わせによる Time-Varying Mesh の動き" 情報科学技術レターズ, LH-010, pp. 241-244, Sep 2007 (letter).
- [15] T. Yamasaki and K. Aizawa, "Motion Segmentation for Time-Varying Mesh Sequences Based on Spherical Registration," EURASIP Journal on Advances in Signal Processing Volume 2009 (2009), Article ID 346425, 9 pages doi:10.1155/2009/346425.
- [16] T. Yamasaki and K. Aizawa, "Content-based cross search for human motion data using time-varying mesh and motion capture data," IEEE ICME2007, pp. 2006-2009, 2007.
- [17] 唯野隆一, 山崎俊彦, 相澤清晴, "測地線距離による高速かつロバストな 3次元ビデオからのスケルトン," 電子情報通信学会論文誌 D, Vol. J90-D, No.7, pp.1729-1732, 2007. (letter)
- [18] 唯野隆一, 山崎俊彦, 相澤清晴, "スケルトンを用いた Time-Varying Mesh からの動き抽出と類似動作検索" 映像情報メディア学会誌, vol. 62, no. 5, pp. 729-742, May, 2008.
- [19] R. Tadano, T. Yamasaki, and K. Aizawa, "Fast and robust motion tracking for time-varying mesh featuring Reeb-graph-based skeleton fitting and its application to motion retrieval," IEEE ICME2007, pp. 2010-2013, 2007.
- [20] T. Yamasaki, Y. Hamazaki, and K. Aizawa "Interactive refinement and editing for time-varying mesh," Proc. 3DTV-Conf 2008, pp. 285-288, 2008.
- [21] N.S. Lee, T. Yamasaki, and K. Aizawa, "Hierarchical mesh decomposition and motion tracking for time-varying-meshes," IEEE ICME2008, pp. 1565-1568, 2008.
- [22] T. Maeda, T. Yamasaki, and K. Aizawa, "Model-Based Analysis and Synthesis of Time-Varying Mesh," AMDO2008 Springer LNCS 5098, pp. 112-121, 2008.
- [23] N.S. Lee, T. Yamasaki, and K. Aizawa, "Motion Tracking of Time-Varying Mesh Through Surface Gradient Matching with Multi-Temporal Registration," ACM SIGGRAPH2008, poster #0196, 2008.
- [24] D. Kasai, T. Yamasaki, and K. Aizawa, "Time-Varying Mesh Retrieval using 2D Video Queries Based on Silhouette Shape Descriptors," PSIVT2009. (submitted)
- [25] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin, "Shape distributions," ACM Transactions on Graphics (TOG), vol. 21, issue 4, pp. 807-832, 2002.
- [26] <http://mocap.cs.cmu.edu/>
- [27] <http://www.mocapdata.com/>
- [28] M. Hilaga, Y. Shinagawa, T. Kohmura, and T. L. Kunii, "Topology matching for fully automatic similarity estimation of 3D shapes," Proc. ACM SIGGRAPH 2001, pp. 203-212, 2001.
- [29] D.-Y. Chen, X.-P. Tian, Y.-T. Shen, and M. Ouhyoung, "On Visual Similarity Based 3D Model Retrieval," Computer Graphics Forum, vol. 22, no. 3, pp. 223-232, Sep. 2003.