

動的点群の統合処理

新谷 幹夫[†] 塩谷 裕作[†] 白石 路雄[†]

3次元デジタル化技術の発展により、運動物体の実時間3次元形状獲得が可能になりつつある。しかし、得られる時系列形状データはフレーム毎に独立であり、圧縮や編集加工を効果的に行うためには、各々のフレームを関連づける必要がある。これまで、剛体運動や人体など特定の運動学的モデルを前提とした研究例はあるが、より一般的な枠組が必要である。そこで本研究では、柔軟物体の点群時系列を効果的に統合する手法の開発を行った。各時間 t において、順序が任意の点群 $\{x_i\}[t]$ が与えられたとき、形状の変化が小さくなるような点群の軌跡 $\{X_i[t]\}$ を、変形エネルギーと距離エネルギーの和の最小化により求める。変形エネルギーは仮想的なバネを考え、Baraff の変形エネルギーで表現する。また、距離エネルギーはターゲット点群の最近点までの距離の和で表す。最小化手法には最急勾配法を用いる。人物像の実測3D時系列データを用いて統合処理を行い、その有効性を実証的に示した

Unification of Dynamic Point Data

MIKIO SHINYA,[†] YUSAKU SHIOYA[†] and MICHIO SHIRAISHI[†]

Recent progress in digitizing technologies is making it possible to capture the 3D shapes of moving objects. To efficiently utilize time series records of spatial data, the information must be unified to yield coherent deforming models.

This paper presents a general method that unifies unregistered 3D point sequences to generate deforming mesh models. The method does not assume any specific kinematic structure, and is applicable to any digitizer. The method first polygonizes the initial points and then deforms meshes to best fit the subsequent data points while minimizing the deformation energy. Experiments are conducted on real measured data and CG data, and successful results are obtained. As an application of the method, we examine data compression and achieve a 380 fold reduction rate for a measured data sequence.

1. はじめに

3次元デジタル化技術の発展により、運動物体の実時間3次元形状獲得が可能になりつつあり、3次元ビデオシステムの研究が盛んになっている^{4),9)}。これらのシステムにより得られる原データは、時系列の点群データである。この時系列点群データはフレーム毎に独立であり、圧縮や編集加工を効果的に行うためには、各々のフレームを関連づけ、単一の動的CGモデルに統合することが好ましい。

物体の運動が剛的な場合は、静的なレンジデータの統合と等価な問題となり、ICP(Iterative Closest Point)法⁶⁾により統合が可能であるが、一般運動に対しては物体の変形を配慮する必要がある。Carranzaら⁵⁾, Sandら¹¹⁾は人体モデル、Zhangら¹⁵⁾は顔の標準モデルを

事前情報として与え、モデル駆動的な手法で単一CGモデルの抽出に成功している。しかし、3次元ビデオの一般的な用途を考えると、特定の運動モデルを仮定しない方式が望まれる。リンク構造などの運動モデルを仮定しない統合化手法として、筆者らはエネルギー最小化に基づく手法を提案している¹³⁾。この手法では、まず初期フレームの点群からポリゴンモデルを作成する。次いで、このモデルを観測点との距離および変形量が最小になるように変形させ、単一の変形モデルを導く。本手法では、特定のリンク構造などの運動モデルを仮定せずに、安定に柔軟物体の動的点群データを単一モデルに統合化することができる。本稿では、手法の詳細と、実験的な検証を示す。

2. 関連研究

静的なレンジデータ統合に関しては多数の研究があるが、実時間測定は比較的新しい技術であるため、動的統合の研究はあまり多くない。モデルベースの3次

[†] 東邦大学理学部情報科学科
Dept. of Information Science, Toho University

元ビデオシステムは必然的に単一のモデルを出力するが^{5),11),15)}、リンク構造や形状に関する事前知識が必要なため、適用可能なシーンは強く制限される。プレートモデルを計測点にフィッティングする問題も動的統合に関連が深い。Marschner ら⁸⁾、Allen ら²⁾が最小化による手法を提案しているが、多数の対応点を手動で指定する必要があり、3次元ビデオには直接適用することが困難である。レベルセットを用いた手法¹⁶⁾は対応点を指定する必要がないが、初期形状の変形を制御することが難しい。画像処理における類似した問題として、画像輪郭のマッチングがある。Snakes はよく知られた方法であり⁷⁾、提案手法との関連も深い。また、Nobuhara らは変形エネルギーと色情報を用いて復元形状の高精度化を行っている¹⁰⁾。

3. 手法の概要

提案する手法は、各フレーム毎に対応付けられていない点群データを入力とし、単一の変形するポリゴンモデルを出力する。

3.1 問題設定

時系列形状データの各時間 t における点群 $\{x_i\}[t]$ が与えられたとする。これを基に頂点の軌跡 $\{X_i[t]\}$ を

- $\{x_i\}[t]$ になるべく近く、
- 形状の変化がなるべく小さく、

なるように定めたい。このとき、実用性の観点から

- 点群データのみを入力とし、リンク構造や対応点などの情報を必要としない
- 柔軟物体の変形を許容する
- 点群データの部分的欠落を許容する

などの条件を満たすようにする。

3.2 定式化と処理の流れ

ICP 法や Snakes と同様に、この問題をエネルギーの最小化問題として定式化する。初期の頂点 $\{X_i[0]\}$ は、以下のエネルギーを最小化するように移動してゆく。

$$E = E_{int} + E_{dis},$$

ここで、 E_{int} は変形エネルギーで形状の変形を測り、 E_{dis} は距離エネルギーで計測点群と軌跡のマッチングの良さを評価する。

この処理は図1のように実現できる。まず、最初のフレームの点群をポリゴン化し、必要に応じて簡略化する。このポリゴンの頂点を変形頂点と呼ぶことにする。ついで、次フレームの点群(ターゲット点群と呼ぶ)を読み込み、点群からの距離場を前処理として計算しておく。変形頂点のスタート位置を補外により求

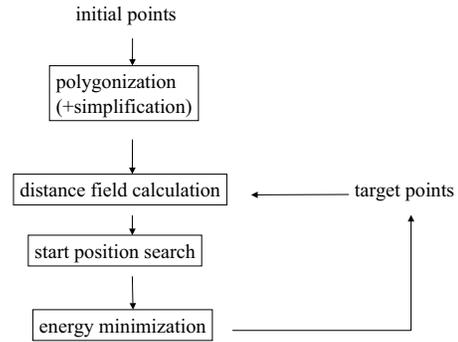


図1 Outline.

め、最小化を行う。

4. エネルギー

4.1 変形エネルギー

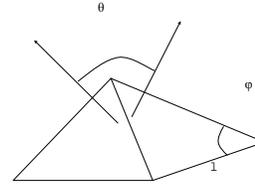


図2 Deformation energy.

変形頂点を仮想的なばねで結合し、そのばねのエネルギーで変形エネルギーを定義する。ばねエネルギーとしては、布のシミュレーションなどでよく用いられる Baraff の変形エネルギー³⁾を用いた。変形エネルギーは、伸張成分 E_{st} 、捻れ成分 E_{sh} 、ずれ成分 E_{ben} の和として、

$$\begin{aligned} E_{int} &= \sum (E_{st} + E_{sh} + E_{ben}) \\ &= 1/2 \sum (k_{st}(l - l_0)^2 + k_{sh}(\theta - \theta_0)^2 \\ &\quad + k_{ben}(\cos \phi - \cos \phi_0)^2), \end{aligned} \quad (1)$$

と表される。ここで、 l, l_0 はばねの長さとおよび自然長、 θ, θ_0 はメッシュ間の角度および初期値、 ϕ, ϕ_0 は頂点およびメッシュ間の角度およびその初期値を示している(図2)。 k_{st}, k_{sh}, k_{ben} はバネ定数で、決定方法は後述する。また、系のバネ行列 $K = (k_{ij})$ を

$$k_{ij} = -\partial^2 E_{int} / \partial x_i \partial x_j,$$

と定義する。ここで x_i は i 番目の頂点の座標を表している。

4.2 距離エネルギー

距離エネルギーは、変形頂点とターゲット点群の最近点までの距離の和で表す。

$$E_{dis} = \sum_i m_i D_i,$$

ここで、 m_i は変形頂点 i の仮想的な質量、 D_i は最近点までの距離を表す。

距離 D_i を最小化の反復時に行くと処理量が膨大となる。そこで、前処理により距離場を計算しておき、これを参照することで距離エネルギーを評価する。前処理では、まずターゲット点群の周辺の空間を、計測精度に対応する程度の解像度でボクセル化する。ついで、各ボクセルの中心点に対して最近距離を計算し、格納する。最近距離の計算は距離変換（例えば¹²⁾）などにより、ボクセル数に対して線形時間で効率的に算出できる。

4.3 パラメータの選択

物理ベースの手法において、パラメータの決定はやっかいな問題であり、機械的な決定法が望まれる。そこで本手法では、同一形状を細分割してもエネルギーの総和が変わらないようにバネ定数および質量を決めるようにする。このためには、 m_i, k_{ben} は頂点に割り当てられる面積に比例させ、 k_{st}, k_{sh} は一定値とすればよい。

5. 最小化手法

5.1 最急勾配法

本手法では最小化手法として最急勾配法を用いた。これは、以下の反復で実現できる。

$$x[t+1] = x[t] + h * \nabla E. \quad (2)$$

この方法は記憶容量をあまり必要とせず、局所解に対する対処が比較的容易である。またタイムステップ h を

$$h < 2/k_{max}, \quad (3)$$

とすることで、安定に反復処理を行うことができる¹⁴⁾。ただし、 k_{max} はバネ行列 $\{k_{ij}\}$ の最大固有値である。

5.2 処理の高速化

処理の効率化を図るために以下のような工夫を施した。

- 初期値の改良 (MP) : 線形・2次外挿曲線の最小化探索により動き予想を行い、最小点を初期値とする。
- 仮決定 (PD) : 動きの小さな頂点を仮決定とし、処理対象からはずす。ただし、隣接頂点が動いた場

合は復活する。

- 法線の比較 : 変形頂点の法線方向と最近ターゲット点の法線方向が著しく異なる場合は、エネルギーから除外する。これにより誤対応による速度低下を防ぐ。

6. CG データを用いた実験

まず、精度を定量的に評価するため、CG データによるシミュレーション実験を行った。実験データはポリゴンモデルを基に点群を生成し、モーションキャプチャーデータおよび物理シミュレーションにより動きをつけることにより作成した。

6.1 人体動作

モーションキャプチャーデータにより動きをつけたデータに対し本手法を適用した。ポリゴンモデルは453頂点からなる。距離場は $200 \times 200 \times 200$ のボクセルで計算した。ターゲット点群は表面上をランダムにサンプルして作成した。数はおよそ2万点である。ジャンプや宙返りなど8つの動作、合計945フレームに関して実験を行った。

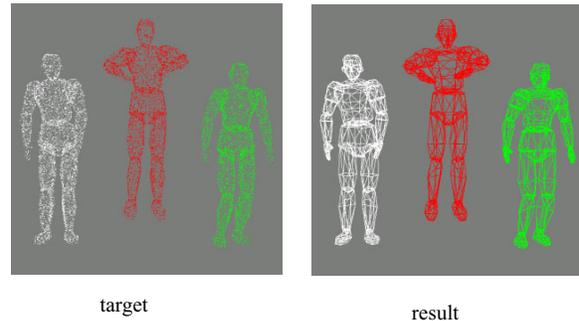


図3 JUMP sequence.

結果の一例を図3に示す。また表1はRMS誤差と平均処理時間を示している。ただし、誤差はターゲットとなるデータと追跡結果との間の差

$$\left(\sum_t \sum_i (|X_i[t] - x_i[t]|)^2 / (N_{ver} N_t) \right)^{1/2},$$

で定義している。JUMPのシーケンス(図2)は170cmの人物を想定しており、RMS誤差は3.886mmという結果になった。ボクセルサイズは12mmとしているので、誤差はボクセルサイズの半分以下に抑えられた

表 1 Average CPU time and RMS errors measured in the motion-captured sequences.

	# of frames	rms	CPU (s/frame)
JUMP	81	0.003886	99.62
RUN	53	0.004316	88.97
KICK	82	0.004745	90.21
WALK2	112	0.003524	65.65
SOMERSAULT	149	0.005267	121.9
SHAKING	198	0.003643	64.70
WALK	105	0.003585	65.69
SITTING	165	0.002991	57.34
Total	945	0.0039	79.23

表 2 Average CPU time and RMS errors with/without the motion prediction (MP) and the provisional determination (PD).

	NONE	MP only	PD only	BOTH
CPU (s/frame)	151.4	82.28	88.78	65.69
RMS error	0.004641	0.003691	0.003605	0.003585

ということになる。平均処理時間は、2 GHz Pentium III で 1 フレーム当たり 79 秒であった。

表 2 に高速化処理の効果を示す。MP は初期値改良、PD は仮決定処理を示し、それぞれ単独では 46%、42%、両方用いると 57% 処理時間が短縮された。誤差も改善されており、有効性が示された。

6.2 布のダイナミックシミュレーション

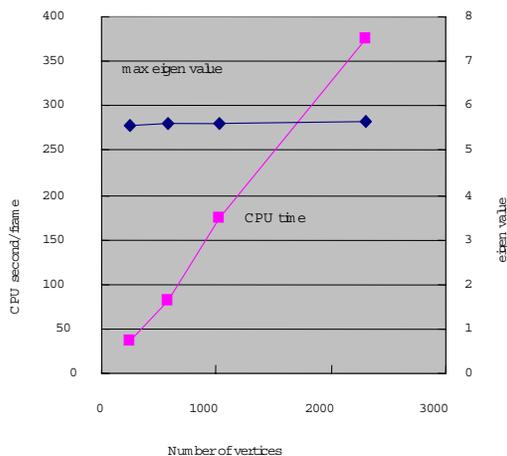


図 4 CPU time and number of vertices.

より柔軟な運動への適応性を調べるため、布のダイナミックシミュレーション結果に対して実験を行った。メッシュサイズに対する処理時間を図 4 に示す。図の

ように頂点数にほぼ比例して計算時間がかかる。このことは以下のように解釈できる。バネ定数 k_{st}, k_{sh} は一定に保たれるので、バネ行列の最大固有値 k_{max} は一定であり、時刻 h も式 3 により一定である。反復当たりのコストは頂点数に比例するので、トータルの処理量は頂点数に比例する。処理結果の一例を図 5 に示す。

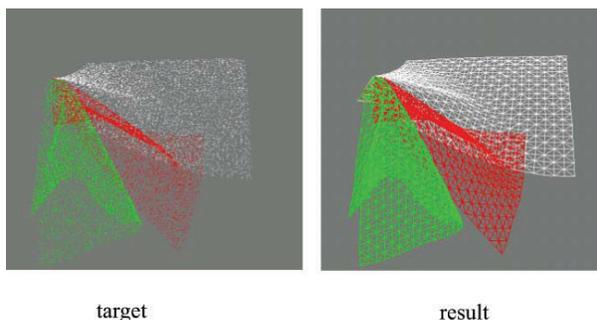


図 5 Cloth sequence.

7. 実データを用いた実験

7.1 3次元ビデオ

3次元ビデオシステム⁹⁾で得られたデータに本手法を適用した。被写体は12台のカメラで撮影され、PCクラスタにより実時間でボクセル表現が得られ、マーキングキューブ法によりポリゴン化される。原データのポリゴンは約13Kであるが、初期ポリゴンは500頂点に簡略化した。結果の一例を図6に示す。原データはノイズが多く、しばしば部分的な欠落を起こしているが、視覚的に良好な結果が得られている。

7.2 顔データ

顔画像を1コマ毎にレーザスキャンし、30フレームを獲得した。原データ40Kポリゴンを1000頂点に簡略化し、提案手法を適用した。結果の一例を図7に示す。口の開け閉めなど表情変化も取り込まれている。

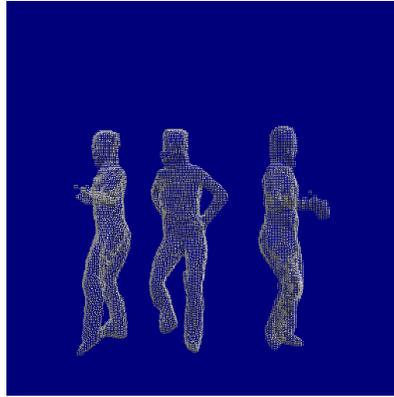
7.3 手のCTデータ

人間の手のCTデータに対して処理を行った。原データ10Kポリゴンを1K頂点に簡略化した。処理結果を図8に示す。

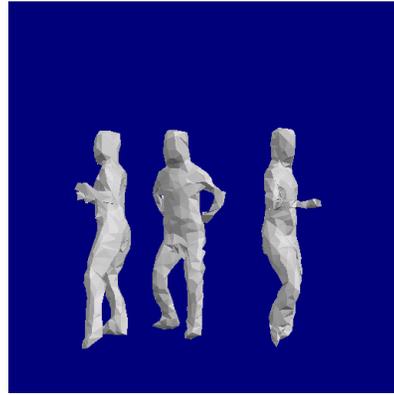
8. データ圧縮

データ圧縮は提案手法の重要な応用分野である。時

データは京都大学松山研究室に提供いただいた。
データはATR人間情報科学研究所で測定した。
産総研デジタルヒューマン研究センターから提供いただいた。



target



result

図 6 Dance sequence.

系列点群が単一の変形ポリゴンに統一化されれば、様々なデータ圧縮が可能となる。まず、エッジや面などの位相構造は不変なので、自動的に半分以上の圧縮がなされる。また、ポリゴンの単純化により、空間的な圧縮が可能となる。さらに、主成分分析 (PCA) を適用することにより、動きの冗長成分を除去し、時間的なデータ圧縮を行うことができる¹⁾。

変形頂点 $X_i[t]$ から PCA により主成分 $Y_j[t]$ を抽出する。これを用いると元の頂点は、少数 (n_Y) の主成分の線形結合

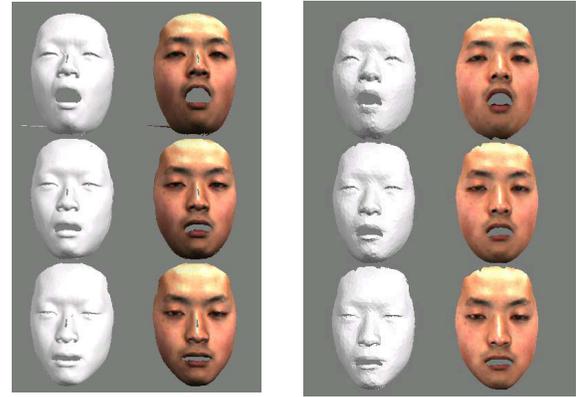
$$X_i[t] \simeq \sum_j^{n_Y} c_{ij} Y_j[t]$$

で表される。元の点群データの大きさ D_0 は、点数 N_v 、フレーム数 N_t を用いて、

$$D_0 = 3 \times N_t N_v$$

と評価される。一方、圧縮されたデータは、主成分 Y_j と各頂点に対する係数 c_{ij} なので、その大きさは、

$$D_1 = 3 \times N_t n_Y + N_v n_Y$$



target

result

図 7 Laser scanned face.

表 3 Data compression by the PCA. N_{v0} : average number of input vertices, N_v : number of deforming vertices, N_t : number of frames, n_Y : number of principal components, D_0 : input data amount, D_1 : compressed data amount.

	JUMPING	DANCING
N_{v0}	453	13135
N_v	453	504
N_t	82	89
n_Y	15	12
D_0	111K	3.5M
D_1	10.5K	9.2K
D_0/D_1	10.6	380

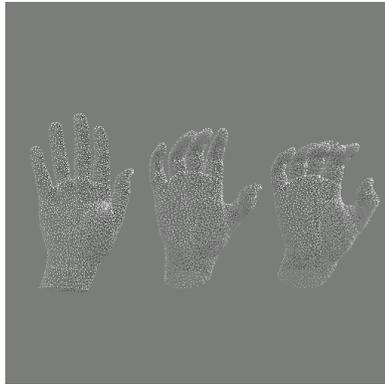
である。主成分数 n_Y が頂点数 N_v 、フレーム数 N_t より十分小さければ、大きな圧縮が可能となる。

CG データ (JUMP) および実データ (DANCE) に主成分分析を施した結果を表 3 に示す。誤差 0.1% の条件で、それぞれ 15 個、12 個の主成分に圧縮された。JUMP では 1/10 の圧縮が得られた。DANCE では、主成分分析で 1/15、ポリゴン簡略化で 1/26、トータルで 1/380 の圧縮が実現されている。

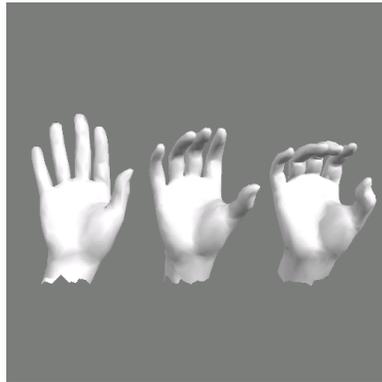
9. ま と め

本稿では、点群時系列の統合手法を提案した。提案手法は、統合化を変形エネルギーと距離エネルギーの和の最小化問題として定式化し、点群の軌跡を最急勾配法により求めた。また、特定の運動モデルを前提としない変形を許容するという特長を持ち、局所解から逃れるための工夫や処理の高速化も行った。

CG データおよび実データに提案手法を適用し、良好な結果を得た。また、主成分分析によりデータ圧縮を行い、大きな圧縮率を得た。



target



result

図 8 CT data.

謝辞 貴重なデータの提供を頂いた松山 京都大学教授、倉立 ATR 研究員、栗原 日立中央研究所主任研究員に深謝する。本研究の一部は、文科省科研費(基盤研究(C) 14550380)の補助を受けている。

参 考 文 献

- 1) M. Alexa and W. Muller. Representing animations by principal components. *Computer Graphics Forum*, 19(3):411–418, 2000.
- 2) B. Allen, B. Curless, and Z. Popovic. The space of human body shapes: reconstruction and parameterization from range scans. In *Proceedings of SIGGRAPH'2003*, pages 587–594. ACM Press / ACM SIGGRAPH, 2003.
- 3) D. Baraff and A. Witkin. Large steps in cloth animation. In *Proceedings of SIGGRAPH'98*, pages 43–54. ACM Press / ACM SIGGRAPH, 1998.
- 4) E. Borovokov and L. Davis. A distributed system for real-time volume reconstruction. In *Proceedings of Computer Architectures for Ma-*

- chine Perception*, pages 183–189, 2000.
- 5) J. Carranza, C. Theobalt, M. A. Magnor, and H. Seidel. Free-viewpoint video of human actors. In *Proceedings of SIGGRAPH'2003*, pages 569–577. ACM Press / ACM SIGGRAPH, 2003.
- 6) Y. Chen and G. Medioni. Object modelling by registration of multiple range data. *Image and Vision Computing*, 10(3):145–155.
- 7) M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1988.
- 8) S. Marschner, B. Guenter, and Raghupathy. Modeling and rendering for realistic facial animation. In *Proceedings of 11th Eurographics Rendering Workshop*, pages 231–242, 2000.
- 9) T. Matsuyama and T. Takai. Generation, visualization, and editing of 3d video. In *Proceedings of 1st International Symposium on 3D data processing, visualization and transmission*, pages 234–245, 2002.
- 10) S. Nobuhara and T. Matsuyama. Heterogeneous deformation model for 3d shape and motion recovery from multi-viewpoint images. In *Proceedings of 2nd International Symposium on 3D data processing, visualization and transmission*, pages 566–573, 2004.
- 11) P. Sand, L. McMillan, and J. Popovic. Continuous capture of skin deformation. In *Proceedings of SIGGRAPH'2003*, pages 578–586. ACM Press / ACM SIGGRAPH, 2003.
- 12) R. Satherley and M. W. Jones. Vector-city vector distance transform. *Computer Vision and Image Understanding*, 82(3):238–254, 2001.
- 13) M. Shinya. Unifying measured point sequences of deforming objects. In *Proceedings of 2nd International Symposium on 3D data processing, visualization and transmission*, pages 904–911, 2004.
- 14) M. Shinya. Theories for spring-mass simulation in computer graphics: Stability, costs and improvements. *IEICE Trans. Info & Sys*, (4):767–774, 2005.
- 15) L. Zhang, Snavely N., B. Curless, and S. Seitz. Spacetime faces: high resolution capture for modeling and animation. *ACM Trans. on Graphics*, 23(3):548–558, 2004.
- 16) Osher S. Zhao, H.-K. and R. Fedkiw. Fast surface reconstruction using the level set method. In *Proceedings of 1st IEEE Workshop on Variational and Level Set Methods*, pages 194–202, 2001.