

## 弾性メッシュモデルを用いた複雑な3次元形状・運動の同時復元

延原 章平<sup>†a)</sup>松山 隆司<sup>†</sup>A 3D Deformation Model for Complex 3D Shape and Motion Estimation  
from Multi-Viewpoint VideoShohei NOBUHARA<sup>†a)</sup> and Takashi MATSUYAMA<sup>†</sup>

あらまし 本論文では弾性メッシュモデルを用いた、多視点ビデオからの対象の3次元形状および運動の同時復元手法を提案する。本手法の特長は(1)観測画像のみから形状と運動の復元を行うことで、和服のように複雑な形状を持つ対象に対しても扱うことができる点と、(2)身体の一部が接触するといった対象形状の見かけ上のトポロジー変化に対処できる点である。3次元形状と運動の復元にあたっては、シルエット、テクスチャ、形状の滑らかさと連続性、物体の剛性、局所的な運動情報、向かい合う面同士の距離など、多様な情報を弾性メッシュモデルという1つの枠組みに統合し、3次元形状と運動を同時に推定する。

キーワード 弾性メッシュモデル, 多視点映像, 3次元ビデオ

## 1. はじめに

本論文では、対象の3次元形状および運動情報を、対象を囲むように配されたカメラ群によって撮影された多視点映像を用いて推定することを目的とする。特に本論文では図1のような(a)和服の袖などの非剛体運動や、(b)足の交差・接触のような対象形状のトポロジー変化、といった複雑な形状と運動の推定を目的とする。このような問題に対し、我々は[1]で提案した弾性メッシュモデルによる3次元形状復元を、形状と運動の同時復元に拡張したアルゴリズムを提案する。提案する3次元形状・運動推定によって、たとえば能や日本舞踊のように、踊り手の姿勢や動作に加えて、袖や裾などの微妙な動きが重要な意味を持つと考えられる伝統芸能の分析・記録への応用や、フレーム間圧縮による3次元ビデオの効率的な保存・伝送が可能となる[2][3]。

## 1.1 関連研究

観測画像に基づいて対象の形状および運動を同時に推定する問題は、コンピュータビジョンの分野ではこれまでも様々なアプローチで取り組まれてきた。たとえばHeapら[4]は、対象(人の手)の形状は既知と

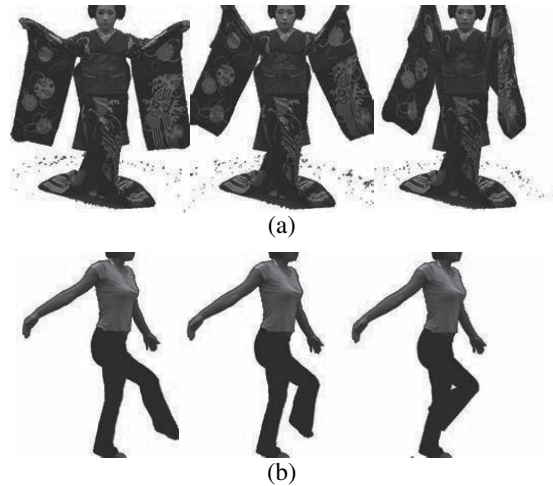


図1: 複雑な(a)形状と(b)運動の例(部分撮影のため、対象の一部が枠外となっている)

Fig.1 Examples of complex (a) shape and (b) motion

してメッシュモデルによって表現し、これを变形させることで運動を推定する手法を提案している。この手法ではメッシュを構成する頂点がどのように動きうるかという学習データを与えた上で、撮影画像に一致するようにメッシュを变形することで対象の形状および運動の同時復元を行った。しかしこの場合の形状はあくまでも元のメッシュが単純に多関節剛体として变形

<sup>†</sup> 京都大学 大学院情報学研究所

Graduate School of Informatics, Kyoto University, Japan

a) E-mail: nob@vision.kuee.kyoto-u.ac.jp

しただけであり、真の形状との差が大きいものであった。このような問題に対して Plänklers ら [5] は、平均的な形状と運動モデルを備えた人体モデルを用いた手法を提案している。この手法では、撮影データに基づいてモデルの形状を対象の形状に一致するように微修正することで解決を図った。しかし依然として、(1) 運動モデルを事前に与え、(2) 衣服が変化した場合は個別にモデルを用意する必要がある、といった制約が残っている。

これに対して Vedula らは、完全に観測データのみから 2 時刻間の形状と運動を推定する手法を提案している [6] [7]。しかしこれらは対象の持つテクスチャに完全に依存し、図 1(b) のようにテクスチャの乏しい対象部位が複雑な運動を行う場合には有効ではない。

これに対して我々は [1] において、弾性メッシュモデルを用いた 3 次元形状復元と、それを発展させた 3 次元形状・運動の同時復元手法を提案した。この手法はテクスチャ・シルエットといった複数の情報を統合し、対象の形状や運動に関する情報を事前に用意することなく、対象の形状と運動を同時復元することを目指したものである。しかし対象の運動に関するモデル化を行っていないために、時刻間での形状変化が少なく、かつ対象の見かけ上のトポロジーの変化のないような運動しか復元することができなかった。

そこで本論文では [1] の弾性メッシュモデルに運動に関する制約を導入し、これによって複雑な運動に対処することを目指す。本論文で扱う“複雑な運動”とは、図 1 に示したように (a) 対象のある一部は剛体、別の一部は非剛体といった、運動モデルとしての複雑さと、(b) 身体の一部接触によって対象形状の見かけ上のトポロジーが時間とともに変化するという、形状に関する複雑さの 2 種類である。

## 1.2 提案手法

本論文では、弾性メッシュモデルと呼ぶ 3 次元メッシュモデルを変形させることによって、対象の 3 次元形状・運動の同時推定を行う。まずある時刻  $t$  の対象形状に一致するようなメッシュモデルが得られているとする。このメッシュモデルを頂点の移動によって変形させ、その結果別の時刻  $t+1$  の対象形状を復元することができれば、変形結果が  $t+1$  における対象の形状を、そして各頂点の移動ベクトルが  $t$  から  $t+1$  への運動情報を表しているといえる。このようにある時刻の対象形状から別の時刻の対象形状へと変形する 3 次元メッシュモデルを、以下弾性メッシュモデルと

呼ぶ。

この変形を実現させるため、まずメッシュが対象の真の形状を表しているときに満たすべき制約条件を定義する。次にメッシュを構成する頂点に対して、各制約条件が満たされる位置へ移動させる力を定義する。そして各制約条件がそれぞれ頂点に作用する力となつて、メッシュが全体として平衡する状態に変形したとき、変形終了とする。このように複数の制約条件をそれぞれ頂点に作用する力として扱うことから、複数の情報を統合することのできる形状・運動復元アルゴリズムとなっている。

提案手法では対象の形状に関する制約として [1] で提案した frame-and-skin モデルを使用する。そして対象の運動に関する制約として、対象運動は剛体部位と非剛体部位に分けることができると仮定する。この仮定に基づいて、観測データから各頂点を剛体・非剛体に分類し、剛体に属する頂点の運動に制約を課した上で変形を行う。またメッシュを構成する各頂点を有意なテクスチャを持つものとそうでないものに分類し、有意なテクスチャを持つものをキーとして変形を進める。つまり [1] では全ての頂点を画一的なモデルで変形させていたが、本論文では各頂点が持つ運動およびテクスチャに関する特徴に基づいて変形プロセスを制御し、より頑健な形状・運動復元を実現する。

さらに本論文では、対象形状の大局的なトポロジー変化を伴う複雑な運動を扱うため、ユークリッド距離に基づく衝突検出処理と、向かいあうメッシュ表面間に作用する斥力を導入する。これはそもそも [1] ではすべての処理が頂点単位かつ測地距離に基づいた近傍のみに着目しているために、原理的に大局的なトポロジー変化に対処することが不可能だからである。

## 2. アルゴリズム

先に述べたように、提案手法のポイントは (1) 弾性メッシュモデルが満たすべき制約条件を定義し、その条件を満たすような状態へとメッシュを変形すること、(2) メッシュを構成する頂点を運動情報およびテクスチャ情報に基づいて分類し、各頂点に分類に応じた変形プロセスを適用すること、(3) 対象形状の大局的なトポロジーの変化に対応するために、メッシュ表面間の衝突を回避する斥力を導入することである。

### 2.1 弾性メッシュモデルが満たすべき制約条件

提案手法では以下の 6 つの制約条件: photometric, silhouette, motion flow, inertia, repulsive をメッシュ

が満たすべき制約条件として使用する .

1. **Photometric 制約** 各頂点におけるテクスチャは, それを観測可能な時刻  $t$  および  $t+1$  の視点において一致する .
2. **Silhouette 制約** メッシュを各カメラ視点に投影して得られる投影像は, 時刻  $t+1$  において観測された対象のシルエットと一致する .
3. **Smoothness 制約** メッシュは局所的に滑らかな形状を保ち, 自己交差を起こさない .
4. **Motion flow 制約** 各頂点は visual hull から推定された motion flow [1] に沿って変形を行う .
5. **Inertia 制約** 各頂点の運動は時系列的に滑らかかつ連続的である .
6. **Repulsive 制約** 向かい合うメッシュ表面同士は交差しない .

これらのうち制約条件 1~3 は [1] 同様に frame-and-skin モデルとして対象の形状をモデル化している . これに対して制約条件 4~6 は頂点の移動に関する制約を表している .

以降ではまずこれらの制約条件それぞれを満たす位置へと頂点  $v$  を移動させる力  $F_e(v)$ ,  $F_s(v)$ ,  $F_i(v)$ ,  $F_d(v)$ ,  $F_n(v)$ ,  $F_r(v)$ , を定義する . そしてこれらの和

$$\begin{aligned} \mathbf{F}(v) \equiv & \alpha \mathbf{F}_e(v) + \beta \mathbf{F}_s(v) + \gamma \mathbf{F}_i(v) + \\ & \delta \mathbf{F}_d(v) + \epsilon \mathbf{F}_n(v) + \zeta \mathbf{F}_r(v) \end{aligned} \quad (1)$$

が各頂点に作用し, 全頂点でこの力が平衡するようにメッシュを変形させるアルゴリズムについて述べる .

## 2.2 頂点に働く力の定義

前述の 6 つの制約条件に対応する力の定義を行うにあたり, 以下の記号を定義する . まず以降ではある時刻  $t$  のメッシュが  $t+1$  へと変形を行うものとする . そして時刻  $t+1$  における頂点を  $v$ , その位置を  $q_v$  とし, 時刻  $t$ , すなわち初期状態における頂点とその位置を  $v_0$  および  $q_{v_0}$  とする .

1. **Photometric Force  $F_e(v)$ :**  $F_e(v)$  は, 2 つの時刻間でのテクスチャが一致する位置へと各頂点を移動させる力である . これは [1] と同様に以下の定義を用いる .

$$\mathbf{F}_e(v) \equiv -\nabla E_e(q_v), \quad (2)$$

ここで  $E_e(q_v)$  は “photo-consistency” 関数 [8] [6] を表し,

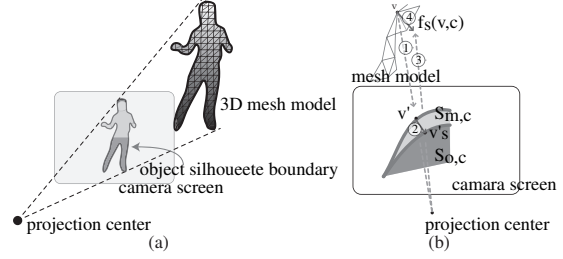


図 2: シルエット制約  
Fig. 2 Silhouette constraint

$$\begin{aligned} E_e(q_v) \equiv & \frac{\sum_{c \in C_{v_0}} \|p_{v_0,c} - \bar{p}_{v_0,v}\|^2}{N(C_{v_0}) + N(C_v)} \\ & + \frac{\sum_{c \in C_v} \|p_{v,c} - \bar{p}_{v_0,v}\|^2}{N(C_{v_0}) + N(C_v)}, \end{aligned} \quad (3)$$

とする . ただし  $C_v, C_{v_0}$  は頂点  $v, v_0$  をそれぞれ観測可能な視点の集合を,  $c$  は視点集合にふくまれる 1 視点を,  $N(C_v), N(C_{v_0})$  は  $C_v, C_{v_0}$  にそれぞれ含まれる視点の数を,  $p_{v,c}$  は時刻  $t+1$  において視点  $c$  で観測されるテクスチャを,  $p_{v_0,c}$  は時刻  $t$  において視点  $c$  で観測されるテクスチャを,  $\bar{p}_{v_0,v}$  は 2 つの時刻のテクスチャ  $p_{v_0,c}$  および  $p_{v,c}$  の平均をあらわす .  $F_e(v)$  は  $v$  を, テクスチャが 2 つの時刻を通して共に一致する位置へと移動させる .

2. **Silhouette Force  $F_s(v)$ :**  $F_s(v)$  は [1] での以下の定義を用いる . 図 2(b) において  $S_{o,c}$  は時刻  $t+1$  においてカメラ  $c$  で観測された対象のシルエットを表し,  $S_{m,c}$  は  $c$  上に投影されたメッシュのシルエットを, そして  $v'$  が  $v$  の  $c$  における投影先を表すとして,  $C_v$  に含まれる  $c$  それぞれについて,  $f_s(v, c)$  を以下のように計算する .

- もし  $v'$  が  $S_{m,c}$  の輪郭上にあるとき, すなわち  $v'$  が contour generator であるときに,  $v'$  が  $S_{o,c}$  の輪郭上に無ければ,  $v'$  から  $S_{o,c}$  の輪郭への最短ベクトル (図 2(b)②), すなわち  $v'_s$  を 3 次元空間中に逆投影したベクトルを  $f_s(v, c)$  とする (図 2(b)④, 後述) .
- それ以外の場合,  $f_s(v, c) = 0$  .

ここで  $f_s(v, c)$  (図 2(b) ④) は以下のように計算する:

$$\mathbf{f}_s(v, c) = (\mathbf{n}_v \cdot \mathbf{d}_{v,v'}) \mathbf{n}_v, \quad (4)$$

ただし  $\mathbf{n}_v$  は  $v$  での法線ベクトルであり,  $\mathbf{d}_{v,v'}$  は  $v'$

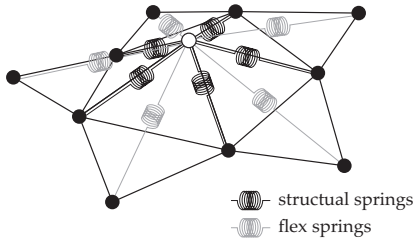


図 3: Smoothness 制約  
Fig. 3 Smoothness constraint

から  $v'_s$  へのベクトルである．そして， $f_s(v, c)$  を観測可能なカメラすべてについて合成し， $f_s(v, c)$  を以下のように定義する．

$$F_s(v) \equiv \sum_{c \in C_v} f_s(v, c). \quad (5)$$

ここで， $F_s(v)$  は contour generator [9] となる頂点にのみ選択的に働く．このことはこの変形アルゴリズムを評価関数の解析的な最適化という形では表現できないことを意味している．

3. Smoothness Force  $F_i(v)$ :  $F_i(v)$  は図 3 に示すような接続関係のパネモデルを用いて以下のように定義する．

$$F_i(v) \equiv \sum_j^n k_j (\|q_{v_j} - q_v\| - l_{v,v_j}) - d_v \dot{q}_v \quad (6)$$

ここで  $v_j$  と  $n$  は  $v$  に図のように接続された頂点とその総数であり， $k_j$  は  $v$  と  $v_j$  の間のパネ定数， $l_{v,v_j}$  はパネの自然長， $d_v$  はパネのダンパ定数である．図中 structural springs は頂点間の距離を保つ働きを持ち，flex springs は頂点の曲率を保つ働きを持つ．この力は (1) メッシュの局所的な自己交差を防ぎ，(2) 他の力にとってローカルサポートとして機能することでアウトライアの除去を行う，という 2 つの役割を持つ．

4. Drift Force  $F_d(v)$ :  $F_d(v)$  は [1] と同様に以下のように定義する．まず各時刻の多視点シルエットが得られると仮定しているため，各時刻の visual hull が得られる．そこで時刻  $t$  のメッシュを構成する頂点集合と，時刻  $t+1$  の visual hull の間で点集合間の変形アルゴリズム [10] によって運動推定を行い，推定結果を線群

$$L = \{l_i \mid i = 1, \dots, N(V_0)\}, \quad (7)$$

とする．ここで  $V_0$  と  $N(V_0)$  は時刻  $t$  におけるメッ

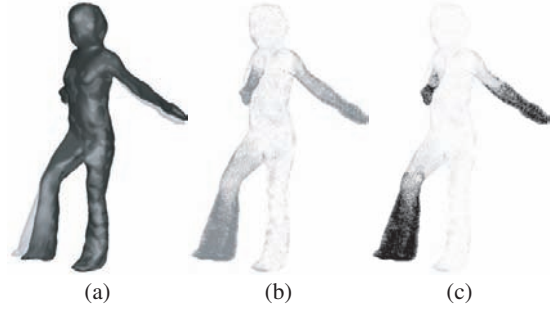


図 4: Motion flow 推定  
Fig. 4 Roughly estimated motion flow lines

シュの頂点群とその頂点数を， $l_i$  は  $V_0$  の  $i$  番目の点を始点とする線を表す．図 4(a) および (b) に得られた推定結果の一例を示す．同図 (a) は入力となる 2 つの時刻の形状を重ねて表示しており，同図 (b) が得られた線群である（同図 (c) については後述）．

このようにして motion flow が線群として得られた後，これを元にして各頂点ごとに局所ポテンシャル場  $E_d(v)$  を以下のように定義する．まず  $l_v$  を頂点  $v$  から  $L$  の中でもっとも近い線とし， $p_{l_v,v}$  が  $v$  から  $l_v$  上で最も近い点， $s_{l_v}$  を  $l_v$  の始点とする．これらを用いて，局所ポテンシャル場  $E_d(v)$  を  $v$  から  $l_v$  までの距離と  $s_{l_v}$  から  $p_{l_v,v}$  までの距離の関数として以下のように定義する．

$$E_d(q_v) \equiv \|s_{l_v} - p_{l_v,v}\|^2 - \|q_v - p_{l_v,v}\|^2. \quad (8)$$

そして，drift force  $F_d(v)$  を局所ポテンシャル場  $E_d(q_v)$  の勾配ベクトルとして

$$F_d(v) \equiv \nabla E_d(q_v) \quad (9)$$

と定義する．式 (8) から，頂点  $v$  が  $L$  中のフローにより近く，またよりフローの終点へと向かうときに， $E_d(q_v)$  は最大となり，式 (9) の定義から，頂点  $v$  は (8) を最大化する方向に，つまりフローに沿ってその終点へと向かうように移動される．

5. Inertia Force  $F_n(v)$ : 提案手法によって逐次的に各時刻への変形を行うと，時刻  $t-1$  から  $t$  へのメッシュ変形が先になされており，前の時刻における各頂点の位置を知ることができる．ここで時刻間隔が十分短いならば，対象の運動は連続かつ滑らかに変化すると仮定することができるため，次の時刻  $t+1$  における運動先を予測することができると思われる．そこで外挿によって時刻  $t+1$  における推定位置

$\hat{q}_v^{t+1}$  を計算する .

$$\hat{q}_v^{t+1} = q_v^t + (q_v^t - q_v^{t-1}) = 2q_v^t - q_v^{t-1}. \quad (10)$$

ただしここでは時刻間隔が一定であるとしている .

そしてこれを元に drift force の定義と同様に , inertia force を

$$F_n(v) \equiv \nabla E_n(q_v), \quad (11)$$

のように局所ポテンシャル場  $E_n(q_v)$  の勾配ベクトルとして定義する .

**6. Repulsive Force  $F_r(v)$ :** 図 1(b) のように , 対象の見かけ上のトポロジーが変化する場合 , 測地距離的には遠いもののユークリッド距離的には近いメッシュ表面同士で接触が起こる . しかしここまでの制約条件とそれに対応する力の定義は , すべて各頂点とその測地距離的な近傍のみに基づいた定義であり , このような状況での自己交差を防ぐことができない .

そこで我々はこのような接触はメッシュ全体では軽微な範囲で起こるとモデル化し , ユークリッド距離に基づいた力として大域的な衝突判定とそれに基づく repulsive force  $F_r(v)$  を導入する . これは , 提案するモデルではあくまでもカメラによって観測可能な情報 (テクスチャ・シルエットなど) によって姿勢・運動を推定することを前提としているため , そもそも接触領域 , つまり観測不能領域が大半を占める場合には対象としないからである .

まず (1)  $F_e(v)$  の計算において頂点ごとにそれを観測可能な視点群  $C_v$  を得ていること , (2) カメラが対象を囲むように位置していることの 2 点を考慮すると , 他の頂点に接する可能性のある頂点では  $C_v = \emptyset$  であることは明らかである . したがって衝突判定はこのような頂点に対してのみ行えばよいといえるので ,

1. すべての頂点において  $F_r(v) = 0$  とする .
2.  $C_v = \emptyset$  である頂点の集合  $V_\emptyset$  に含まれる頂点それぞれについて ,
  - 2.1.  $V_\emptyset$  に含まれる他の頂点へのユークリッド距離を計算し , 距離が  $l_{\min}(v)$  より近い頂点を探す . ここで  $l_{\min}(v)$  は  $v$  に接続された辺のうち , 最も短いものの長さである . こうして見つかった頂点の集合を  $V_d(v)$  とする .
  - 2.2.  $V_d(v)$  に含まれる各頂点  $v'$  について ,  $f_r(v, v')$  を以下のように計算して  $F_r(v)$  に

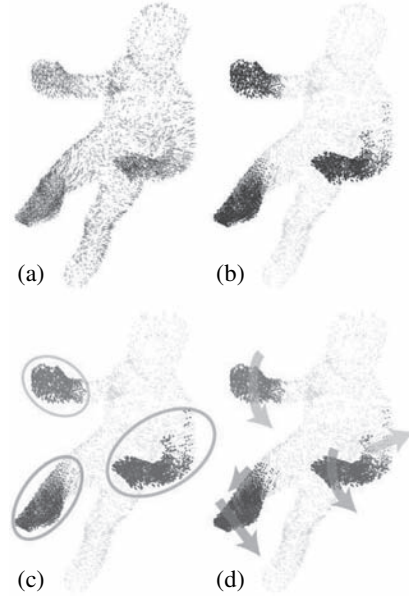


図 5: 剛体部分の推定  
Fig. 5 Rigid parts estimation

加える .

$$f_r(v, v') = -\frac{q_{v'} - q_v}{\|q_{v'} - q_v\|^3}. \quad (12)$$

と repulsive force を定義する . この repulsive force は接触する直前の頂点に , 互いに遠ざかる力を加える . このことにあわせて , 第 3.1 節で述べるように他の力とバランスする係数を与えることによって , 提案手法では自己交差を防ぐ .

### 2.3 頂点の分類

#### 2.3.1 運動に基づく頂点の分類

先に述べたように motion flow 制約と inertia 制約に対応する  $F_d(v)$  ,  $F_n(v)$  の計算に際して , 時刻  $t+1$  における頂点位置を各頂点ごとに推定している . この推定ベクトルを以下の手順でクラスタリングすることで , 頂点を剛体部分と非剛体部分に分類し , 特に剛体部分について下に述べる制約を加えることでより頑健な形状・運動復元を実現する .

1. 各頂点と推定された移動先を結んだ線群を得る (図 5(a)) .
2. このうち , 長さが一定の閾値以下のものは非剛体領域として分類する (図 5(b) の薄灰の線群) .
3. 残りの線群 (図 5(b) の濃灰の線群) をその基点の

空間的近接性（ユークリッド距離）に基づいてクラスタリングする（図 5(c)）。

4. 得られた各クラスタ毎に、方向ベクトルの近接性に基づいて更にクラスタリングを行う（図 5(d)）。こうして得られたクラスタそれぞれを剛体部分とみなす。

また先に示した図 4 のクラスタリング結果を図 4(c) に示す。この図は図 5 と同様に薄灰と濃灰の 2 種類の線群から構成されており、薄い灰色が非剛体領域を表し、濃い領域がそれぞれ個別の剛体領域である。この例では振り上げている右足と振り下げている両手がそれぞれ個別の剛体運動とみなせる領域として検出され、それ以外の部分はすべて非剛体部分となっている。これは右腿のように実際には剛体として動いている箇所であっても、観測上は剛体・非剛体いずれとも解釈可能であるために上記 2. において排除しているからである。つまりここでは、対象を多関節剛体とみなすことができる箇所について、あくまでも時刻  $t$  および  $t+1$  における観測から推定し、それ以外の部分はすべて非剛体とみなす。また非剛体とされた部分が本来剛体であったか非剛体であったかについては取り扱わない。こうして剛体領域として分類された頂点に関しては、推定ベクトルを共通の回転・並進で記述されるものに修正する [11] [12] とともに、smoothness 制約に対応する  $F_i(v)$  におけるバネ定数を大きくすることで剛体変形を実現する。具体的アルゴリズムは 2.4 節で述べる。

### 2.3.2 テクスチャに基づく分類

本論文では (1) ある頂点を観測可能なカメラ間でテクスチャが一致しており、かつ (2) テクスチャが強いエッジ成分を持っているとき、その頂点が有意なテクスチャを持つとみなす。我々はこの条件を式 (3) で定義されたテクスチャの一致度  $E_e(q_v)$  を用いて

$$\frac{\min \left\{ \min_{c \in C_{v_0}} \|\nabla p_{v_0, c}\|, \min_{c \in C_v} \|\nabla p_{v, c}\| \right\}}{1 + E_e(q_v)} > T \quad (13)$$

と定義する。ここで  $T$  は閾値であり、 $\|\nabla p_{v, c}\|$  は頂点  $v$  をカメラ  $c$  に投影した位置におけるテクスチャの微分値の強さを表している。そしてこの条件を満たす各頂点  $v$  について、周囲の頂点  $v_j$  に働く力を以下のように修正する。

$$\begin{aligned} F(v_j) &= \omega F(v) + (1 - \omega) F_{\text{prev}}(v_j), \\ \omega &= e^{-D_g(v, v_j)}, \end{aligned} \quad (14)$$

ここで  $F_{\text{prev}}(v_j)$  は  $v_j$  に本来作用する力であり、 $D_g(v, v_j)$  は  $v$  と  $v_j$  の間の測地距離である。このようにして有意なテクスチャを持つ頂点の移動ベクトルをその周囲の頂点へと伝播させ、有意なテクスチャを持つ頂点が周囲を導くようにする。

### 2.4 Heterogeneous Deformation Model

以上で各頂点に作用する力の定義と、頂点の分類を行った。本研究では頂点の分類に基づくメッシュ変形アルゴリズムとして、下記のような heterogeneous deformation model を提案する。

1. まず時刻  $t = 0$  における対象の形状を  $J[1]$  の intra-frame deformation によって得る。これを弾性メッシュモデルの初期形状とし、 $t = 1, 2, \dots$  について、以下の手順で形状と運動を推定する。
  2.  $F_d(v)$  と  $F_n(v)$  の運動ベクトル推定を行う。
  3. 頂点を運動とテクスチャに基づいて分類する。
    - 3.1. 推定された運動ベクトルをクラスタリングすることで、頂点を Ca-1: 剛体部分、あるいは Ca-2: 非剛体部分に分類する。
    - 3.2. Ca-1 とされた頂点のバネ定数を大きくする。
  4. 反復的に変形計算を行う。
    - 4.1. 各頂点に働く力  $F(v)$  を計算する。
    - 4.2. 式 (13) によって有意なテクスチャを持つとされた各頂点  $v$  について、式 (14) によってその周囲の頂点に働く力を修正する。
    - 4.3. 計算された力に沿って、頂点を移動する。
    - 4.4. 移動量が十分小さいときは 5. へ、そうでなければ 4.1 へ。
5. 得られたメッシュ形状を時刻  $t$  における対象の形状とみなし、各頂点の遷移ベクトルを  $t-1$  から  $t$  の間の対象の運動とみなす。
6.  $t = t+1$  として 2. へ。

### 3. 評価実験

提案したアルゴリズムを評価するために、図 6 のように配置された 15 台のカメラを用いた環境で多視点

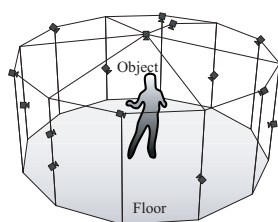


図 6: カメラ配置  
Fig. 6 Camera arrangement

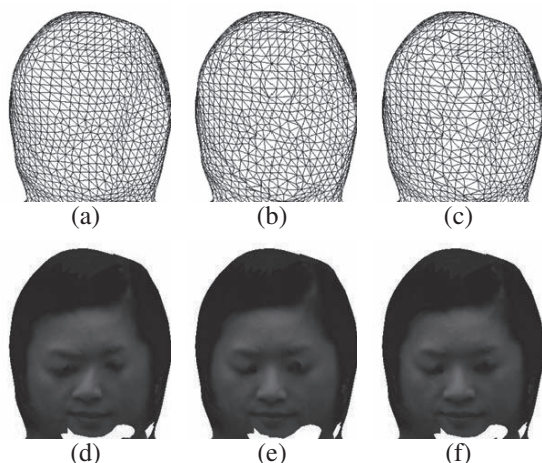


図 7: Photometric 制約の評価  
Fig. 7 Evaluation of the photometric constraint

映像を撮影して実験を行った．使用したカメラの撮影解像度は XGA，フレームレートは 25fps（外部同期），露光時間は 1ms である．また各視点におけるシルエット抽出には [13] の手法を使用した．なお以下では時刻  $x$  を # $x$  と表記する．

### 3.1 係数の設定と評価

式 (1) の係数は，第 2.2 節で定義した各力の重みであるので，これらを互いに意味のあるものとするためには，基本的に均等な重みを与える必要がある．以降の実験ではこの考察と実データに基づいて，経験的に定めた固定値  $\alpha = 0.2$ ,  $\beta = 0.15$ ,  $\gamma = 0.15$ ,  $\delta = 0.2$ ,  $\epsilon = 0.1$ ,  $\zeta = 0.2$  を使用した．ここでスケールは各力の最大の大きさを 1 としている．

次に係数を 0，つまり対応する力を無効にして形状と運動の推定を行い，2.2 節で設計した力が持つ効果を検証する．

まず  $\alpha = 0$  として Photometric 制約を除外して #1 から #2 の形状を推定した結果を図 7 に示す．図は上段が左から順に (a)  $\alpha = 0$  として推定した #2 のメッ

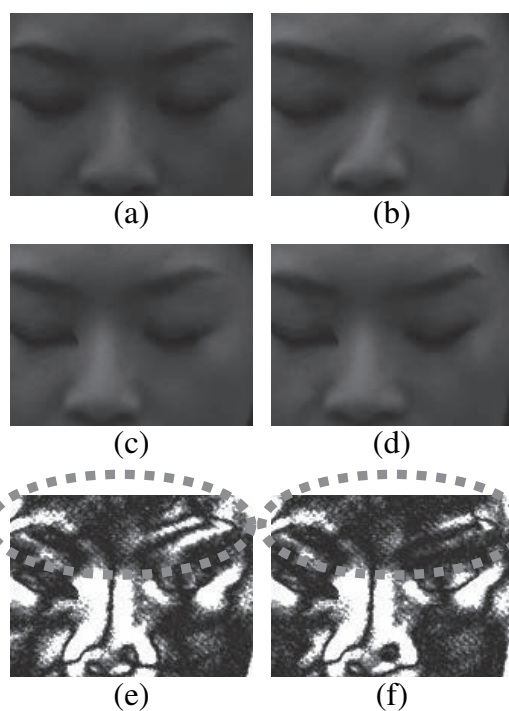


図 8: Photometric 制約の評価（拡大図）  
Fig. 8 Evaluation of the photometric constraint

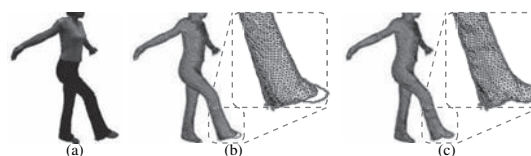


図 9: Silhouette 制約の評価  
Fig. 9 Evaluation of the silhouette constraint

シュ，(b)  $\alpha = 0.2$  として式 (14) を用いずに推定した #2 のメッシュ，(c)  $\alpha = 0.2$  として式 (14) を用いて推定した #2 のメッシュを示し，下段が左から順に (d), (e), (f) がそれぞれ (a), (b), (c) に “#1 の撮影画像を用いて” テクスチャマッピングを行った結果である．つまり (1) まず #1 のメッシュに #1 の多視点画像を使ってテクスチャを生成し，(2) #1 のメッシュを提案した手法で #2 の形状へと変形し，(3) そして (1) で生成したテクスチャをそのまま使ってレンダリングを行っている．

次に図 8(b), (c), (d) は図 7(d), (e), (f) をそれぞれ拡大したものであり，図 8(a) は #2 の撮影画像つまり対象テクスチャの真値である．また図 8(e), (f) は，図 8(c), (d) に示したレンダリング画像と，8(a) に示した

撮影画像との二乗誤差であり、ピクセルあたりの平均 2 乗誤差は各ピクセル 256 階調として順に 10.20, 9.82 であった。

図 7(a),(b) および (c) を比べると、メッシュの概形には特に変化がない。これはメッシュ全体の形状は Silhouette 制約などその他の制約条件によって保たれるためである。しかしテクスチャマッピングした結果を比較すると  $\alpha = 0$  とした場合 (図 8(b)) はテクスチャが本来の位置 (図 8(a)) からずれており、各頂点の位置は本来対応すべき位置へは移動していないことが確認される。一方、Photometric 制約を用いた場合 (図 8(c) および (d)) はテクスチャが (a) に比べて適切に配置されている。さらに図 8(e),(f) を見ると、特に楕円で囲まれた領域のように画像の微分値が大きい箇所において、(f) のほうがより誤差が小さいことがわかる。これは (f) では式 (14) によってより強いテクスチャエッジ (画像微分値) に対応する頂点移動が優先される一方で、(e) ではすべての頂点移動が対等であるために周囲の頂点との smoothness 制約によって頂点位置がずれてしまうためである。

以上から Photometric 制約によって、頂点が変形前後の時刻間で対象上の同じ位置へと移動していること、また式 (14) によって頂点がより正確な位置へと移動し、有効に機能していることが確認できた。なお図 8(a) と (b),(c),(d) の間には特に目の下や鼻周辺に明度差がある。これは対象が移動したことによって照明との相対位置が変化し、#1 と #2 の間で対象表面の明度が変化したためである。先に述べた図 8(e),(f) それぞれのピクセルあたりの 2 乗誤差はこの変化分を含んでいるため、数値上は差が小さくなっている。

次に  $\beta = 0$  として Silhouette 制約を除外して #1 から #2 の形状を推定した結果を図 9 に示す。図は左から (a) 撮影画像、(b)  $\beta = 0$  の場合、(c)  $\beta = 0.15$  の場合であり、(b)・(c) 中の太線は撮影画像から得られた対象シルエットの輪郭線である。図 (b) と (c) を比較すると、 $\beta = 0$  の場合に足の先などの部分で対象本来のシルエットとメッシュの投影像が一致していない。これは該当部分のテクスチャ情報が乏しく、Photometric 制約では位置を正確に決定できないためである。一方 Silhouette 制約を有効にした同図 (c) では撮影シルエットとメッシュの投影像が一致しており、この制約が有効に機能していることが確認できる。

$\gamma = 0$  として Smoothness 制約を無効にした場合、提案手法は図 13#1 から #2 のように類似度の高い時

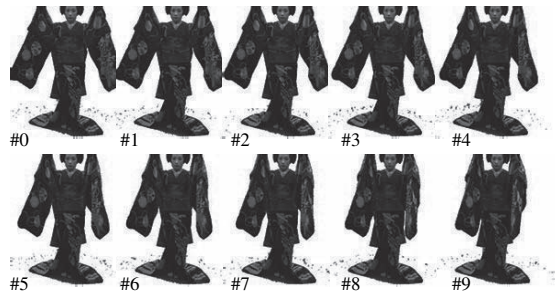


図 10: 入力多視点映像  
Fig. 10 The input multi-viewpoint images

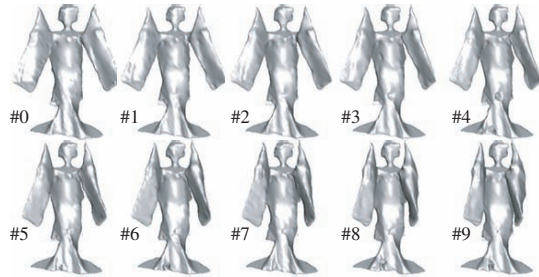


図 11: 提案手法による推定結果  
Fig. 11 Estimated shapes



図 12: 推定された 3 次元運動  
Fig. 12 Estimated motion flow

刻間でも正しく形状を推定することができない。これは (1)Photometric 制約などで発生するアウトライアの除去が全く機能しない、(2)局所的な自己交差が多発して頂点の可視判定 ( $C_v$ ) が正しく行われなくなり、photometric force  $F_e(v)$  や repulsive force  $F_r(v)$  の計算が破綻する、という理由によって、頂点に作用する力が正しく計算できないからである。

同様に  $\delta = 0$  あるいは  $\epsilon = 0, \zeta = 0$  とした場合も提案手法は破綻する。これは剛体・非剛体部分推定を行うことができず、かつ大域的な衝突判定も行わないことによって、対象が複雑に運動し、さらに体節間の接触によって見かけ上のトポロジーが変化する場合に [1] と同様に対処できないからである。

### 3.2 実験 1

複雑な形状を持つ対象の例として、図 10 のような



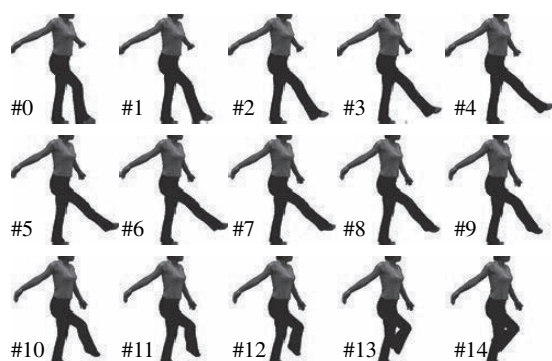


図 13: 入力多視点映像  
Fig. 13 The input multi-viewpoint images

多視点映像から対象の形状と運動を推定した結果を図 11 に示す．また特に開始と終了時刻に注目したものを図 12 に示す．この図は左から順に (a) #1 の形状，(b) #10 の形状であり，(c) は #1 から #10 の間の運動ベクトルを濃灰の始点 (#1 における頂点位置) から薄灰の終点 (#10 における頂点位置) へと向かう線分で表している．

この実験では #1 の多視点画像から [1] の手法によって 3 次元形状を復元し，初期形状として使用した．メッシュを構成する頂点数は約 35000，計算時間は Intel Core2 2.4GHz のコンピュータで 1 時刻あたり約 1.5 時間である．

これらの結果から，提案手法を用いることで複雑な形状を持つ対象についても形状と運動を同時に推定することが可能であることが確認できる．

### 3.3 実験 2

次に見かけ上のトポロジーが変化する複雑な運動に対する例として，図 13 のような多視点映像から対象の形状と運動を推定した結果を図 14 に示す．この例では対象は足を交差・接触させている．この実験においても初期モデルは [1] の手法で復元した #1 のメッシュを用いた．使用したメッシュの頂点数は約 15000，計算時間は Intel Core2 2.4GHz のコンピュータで 1 時刻あたり約 45 分である．

得られたメッシュを骨格線へと変換すると図 15 が得られる [14]．提案手法では頂点の移動によって運動を表現しているため，見かけ上のトポロジーの変化に影響されることなく，骨格線の構造が常に同じであることが確認できる．

特に見かけ上のトポロジーの変化が起きている #10 に注目すると，提案手法では図 16(a) のような形状が

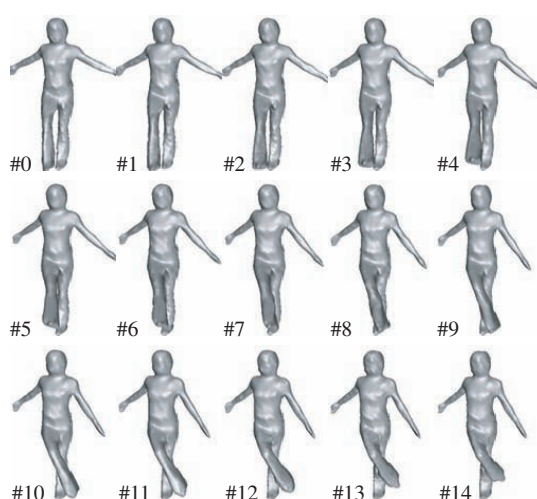


図 14: 提案手法による推定結果  
Fig. 14 Estimated shapes

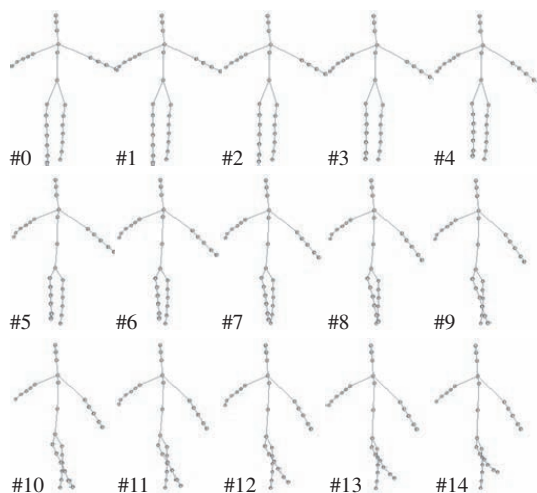


図 15: 提案手法による推定結果 (骨格線)  
Fig. 15 Skeletons of estimated shapes

得られ，その骨格線は同図 (c) のようになる．一方 [1] の手法によってこの時刻のみ独立に形状復元した場合は同図 (b) のように両足が融合したメッシュが得られ，その骨格線は同図 (d) のようになる．図 15・16 から，(1) 独立に形状復元した場合は見かけ上のトポロジー変化に伴ってメッシュのトポロジーも変化していること，(2) 一方提案手法では実際の運動に沿ってメッシュが変形され，トポロジーが保存されていることが確認できる．

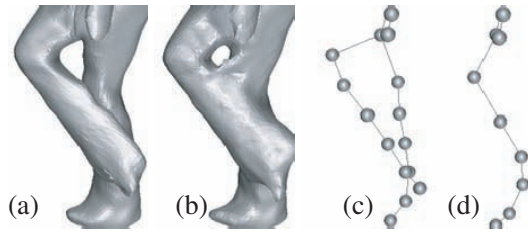


図 16: トポロジー変化を伴う運動の推定

Fig. 16 Shapes and skeletons of frames with topological change

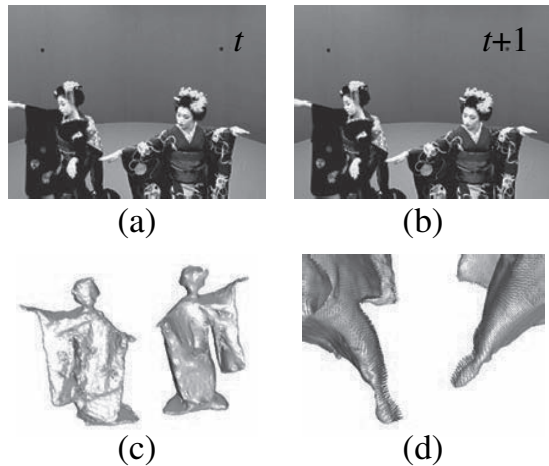


図 17: 複数人物の形状・運動復元 (1)

Fig. 17 Simultaneous estimation of 3D shapes and motions of multiple objects

#### 4. ま と め

本論文では、弾性メッシュモデルの変形という枠組みを用いて、対象の3次元形状および運動を同時に推定する手法を提案すると共に、実画像によってそれが有効に機能することを示した。提案手法の特徴は以下の点である。(1) 対象の形状自身を観測データから得るため、和服のように複雑な衣服を纏った場合のように、モデル化が複雑で困難な対象に関しても、適用可能である。(2) 対象の見かけ上のトポロジー変化を扱うことができる。これにより、図 17 および 18 に示すように、複数の人物に対しても本手法をそのまま適用することが可能である。図 17 は左から (a) 時刻  $t$  での入力画像、(b) 時刻  $t+1$  での入力画像、(c) 推定された時刻  $t+1$  での形状、(d) 時刻  $t+1$  での形状と推定された運動軌跡を天井方向から表示したものであり、図 18 は左から順に入力画像、復元結果の全身像、

復元結果のうち接触部分に注目したレンダリング結果を示している。実験環境は第 3 節と同様であり、アルゴリズム上対象が 1 人か 2 人かの区別はなく、全体で 1 つの弾性メッシュモデルとして扱われている。

また提案した手法は頂点の移動のみで運動を表現するため、メッシュのトポロジーが一定であることを前提としたデータ圧縮 [2] [3] への展開が可能であることが本手法の特長である。さらに提案した手法は、対象の形状と運動情報だけでなく、対象の剛体領域や関節位置など、運動モデルの学習へも応用が可能である。

しかし提案したモデルでは、形状・姿勢推定をあくまでカメラから観測可能な情報のみによって行うため、接触領域、つまり観測不可能な領域が大半を占めるような姿勢については対応できない。今後はこのような重度の接触を含む姿勢について、身体モデルを導入することで解決を目指す。

#### 謝 辞

本研究の一部は、文部科学省プロジェクト「知的資産の電子的な保存・活用を支援するソフトウェア技術基盤の構築」における研究開発課題「大型有形・無形文化財の高精度デジタル化ソフトウェアの開発」、および総務省 SCOPE-C プロジェクト「和装文化振興を目的とした実写 3 次元映像のインターネット配信技術の研究開発」の補助を受けて行った。

#### 文 献

- [1] T. Matsuyama, X. Wu, T. Takai and S. Nobuhara: "Real-time 3d shape reconstruction, dynamic 3d mesh deformation and high fidelity visualization for 3d video", *CVIU*, **96**, pp. 393-434 (2004).
- [2] H. M. Briceño, P. V. Sander, L. McMillan, S. Gortler and H. Hoppe: "Geometry videos: A new representation for 3d animations", *Proc. of SIGGRAPH*, pp. 136-146 (2003).
- [3] H. Habe, Y. Katsura and T. Matsuyama: "Skin-off: Representation and compression scheme for 3d video", *Proc. of Picture Coding Symposium*, San Francisco (2004).
- [4] T. Heap and D. Hogg: "Towards 3d hand tracking using a deformable model", *Proceedings of the 2nd International Conference on Automatic Face and Gesture Recognition*, pp. 140-145 (1996).
- [5] R. Plänkers and P. Fua: "Articulated soft objects for multiview shape and motion capture", *PAMI*, **25**, 9, pp. 1182-1187 (2003).
- [6] S. Vedula, S. Baker, S. Seitz and T. Kanade: "Shape and motion carving in 6d", *Proc. of CVPR* (2000).
- [7] S. Vedula, S. Baker, P. Rander, R. Collins and

- T. Kanade: “Three-dimensional scene flow”, PAMI, **27**, 3 (2005).
- [8] K. N. Kutulakos and S. M. Seitz: “A theory of shape by space carving”, Proc. of ICCV, pp. 307–314 (1999).
- [9] G. Cross and A. Zisserman: “Surface reconstruction from multiple views using apparent contours and surface texture”, Proceedings of NATO Advanced Research Workshop on Confluence of Computer Vision and Computer Graphics, Kluwer Academic Publishers, pp. 25–47 (2000).
- [10] D. Burr: “A dynamic model for image registration”, Computer Graphics and Image Processing, **15**, pp. 102–112 (1981).
- [11] M. D. Wheeler and K. Ikeuchi: “Iterative estimation of rotation and translation using the quaternion”, Technical Report CMU-CS-95-215, Computer Science Department, Carnegie Mellon University, Pittsburgh, PA (1995).
- [12] N. Ohta and K. Kanatani: “Optimal estimation of three-dimensional rotation and reliability evaluation”, Proc. of ECCV, Germany, pp. 175–187 (1998).
- [13] T. Horprasert, D. Harwood and L. S. Davis: “A statistical approach for real-time robust background subtraction and shadow detection”, ICCV Frame-Rate WS (1999).
- [14] T. Tung and F. Schmitt: “The augmented multiresolution reeb graph approach for content-based retrieval of 3d shapes.”, International Journal of Shape Modeling, **11**, 1, pp. 91–120 (2005).

(平成 xx 年 xx 月 xx 日受付)

論文賞, 1990 年人工知能学会論文賞, 1993 年情報処理学会論文賞, 1994 年電子情報通信学会論文賞, 1995 年第 5 回国際コンピュータビジョン会議 Marr Prize, 1996 年国際パターン認識連合 Fellow, 1999 年電子情報通信学会論文賞, 2000 年画像センシングシンポジウム優秀論文賞, 2004 年 FIT 優秀論文賞, 2006 年情報処理学会, 電子情報通信学会フェロー・日本学術会議連携会員.

#### 延原 章平 (正員)



2000 年京大工学部卒 . 2002 年京大大学院情報学研究科修士課程修了 . 2005 年同博士後期課程修了 . 2005 年京大大学院情報学研究科特任助手 . 2007 年より京大大学院情報学研究科 GCOE 助教 . 博士 (情報学) .

#### 松山 隆司 (正員:フェロー)



1976 年京大大学院修士課程修了 . 京大助手, 東北大助教授, 岡山大教授を経て 1995 年より京大大学院電子通信工学専攻教授 . 現在同大学院情報学研究科知能情報学専攻教授 . 2002 年学術情報メディアセンター長, 京都大学評議員 . 2005 年情報環境機構長 . 工博 . 画像理解, 人工知能, 分散協調視覚, 3 次元ビデオの研究に従事 . 最近は「人間と共生する情報システム」の現に興味を持っている . 1980 年情報処理学会創立 20 周年記念

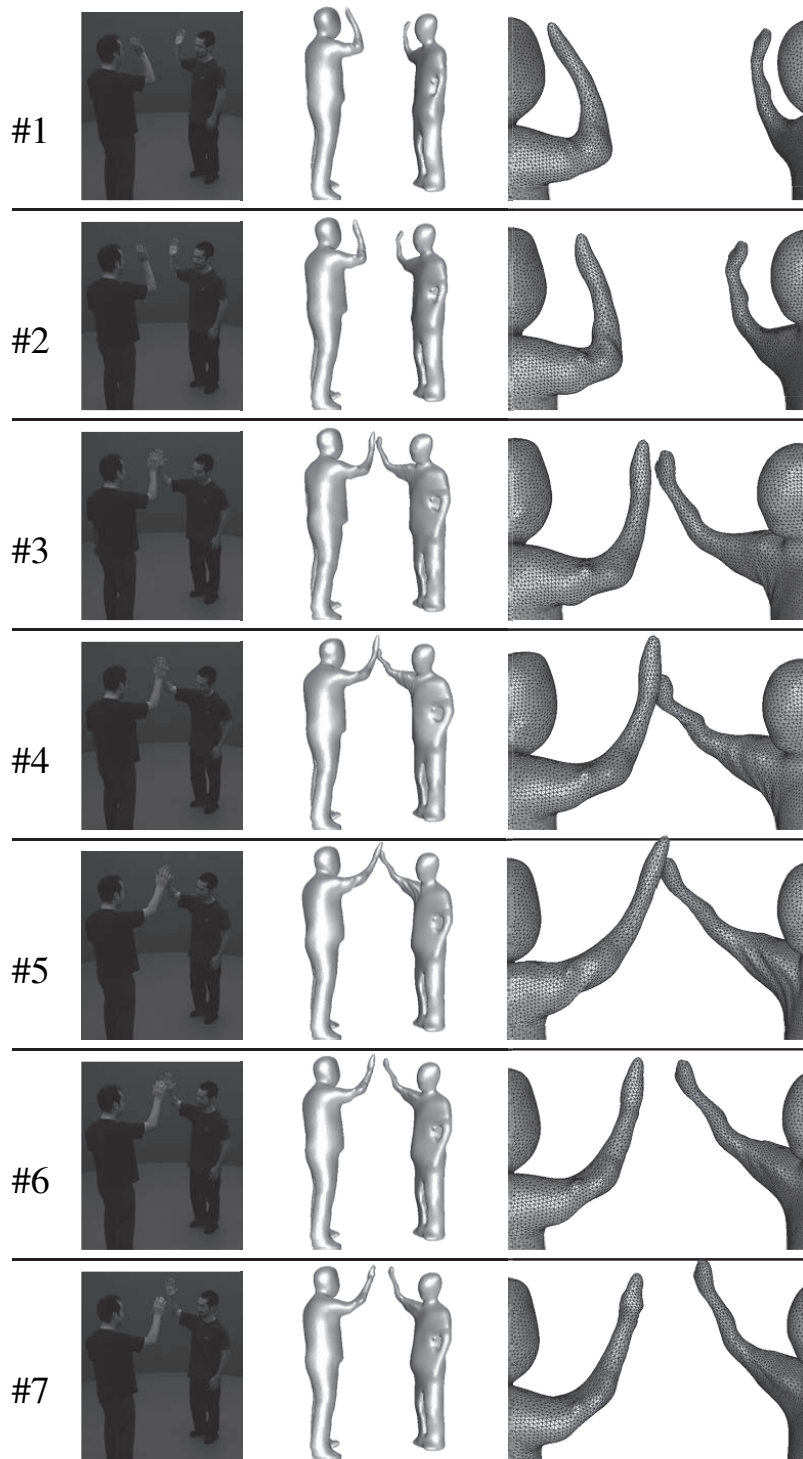


図 18: 複数人物の形状・運動復元 (2)

Fig. 18 Simultaneous estimation of 3D shapes and motions of multiple objects

**Abstract** We propose a new framework for 3D shape and motion estimation using deformable mesh model. Since our method does not require object model given a priori, our method can cope with complex object shape and motion which are difficult to be modeled, e.g., a dancing human with *kimono*. Our method integrates several estimation cues such as: 2D silhouettes, textures, smoothness, rigidity, local motion flow fields, and global collision detection between surfaces into single computation scheme and realizes simultaneous estimation of 3D shape and motion of the object.

**Key words** deformable mesh model, multi-viewpoint video, 3D video