

多視点映像からの3次元形状・運動復元のための 弾性メッシュモデル

延原章平 松山隆司

京都大学大学院情報学研究科知能情報学専攻

概要 本論文は多視点ビデオからの対象の3次元形状および運動の復元という問題に対して、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手法を提案する。

Deformable Mesh Model for 3D Shape and Motion Estimation from Multi-Viewpoint Video

Shohei Nobuhara and Takashi Matsuyama

Graduate School of Informatics, Kyoto University

Abstract We propose a new framework for 3D shape and motion estimation using deformable mesh model. It 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.

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元にして推定することを目的とする。この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、あるいは3次元ビデオのフレーム間圧縮による効率的な保存・伝送が可能となる [1] [2] [3]。

従来よりコンピュータビジョンの分野では、対象を多視点から撮影してその3次元形状や運動を計測する研究 [4] [5] [6] [7] [8] [9] が行われてきたが、その多くは形状あるいは運動のいずれかを復元の対象としていた。これに対して本研究では、対象の運動解析や3次元ビデオのフレーム間圧縮などの応用のために3次元形状と運動の同時復元を目的とする。このような観点から、まずはこれまでの関連研究と我々の提案手法との関係について述べる。

1.1 関連研究

まずはじめに、以下本論文では2.5次元形状と3次元形状を区別して扱う。多視点映像からこれらはしばしばまとめて“3次元形状”と表現されることがあるが、以降では特に記述の無い限り、2.5次元形状は含まないものとする。

対象の3次元形状を映像から推定するためには、対象を異なる複数の視点から撮影する必要があるが、このときこれまでの研究は“空間的”多視点撮影を用いるものと“時間的”多視点撮影を行うものとの分類

することができる。前者は対象を囲むように配されたカメラ群によって撮影を行い [6]、後者は1つのカメラを対象の周りを移動させながら撮影を行う [10]。明らかに、後者は運動する対象の形状を得るという目的には適さないため、本研究では前者の撮影方式を採用する。

次に各視点の映像からは、対象のテクスチャや輪郭などさまざまな情報を得ることができる。2.5次元形状復元手法では、主にテクスチャ [5]、陰影 [11]、オプティカルフロー [12] を用いて2.5次元形状を各視点で復元し、その後この2.5次元形状を1つの3次元形状へと統合する [13]。一方、各視点の情報から直接3次元形状を計算する手法には、対象のシルエット [14] [15] やテクスチャ [16] を用いるものがある。

これらの手法には映像から得られるどの情報を使用するかに応じてそれぞれ一長一短がある。例えばテクスチャマッチングに基づくステレオ法やスペースカービング法では、原理的に任意の観測可能な対象表面の形状を復元可能であるが、対象表面すべてに有意なテクスチャが存在しなくてはならない。一方シルエット輪郭に基づく手法では、テクスチャマッチングに比較して安定に対象の形状が得られるが、得られる形状 (visual hull) は撮影された輪郭形状のみを反映したものであり、映像の上では輪郭として現れない部分、例えば対象の凹領域を復元することができない。そこでこのような問題点を克服し、形状の正確さと復元の安定性を両立させるために、複数の情報を統合するアルゴリズムが提案されてきた。例えば Fua [17] [18] [19] は2.5次元のメッシュで表現

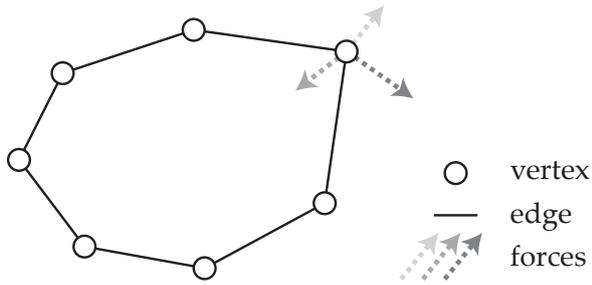


図1 Active contour model

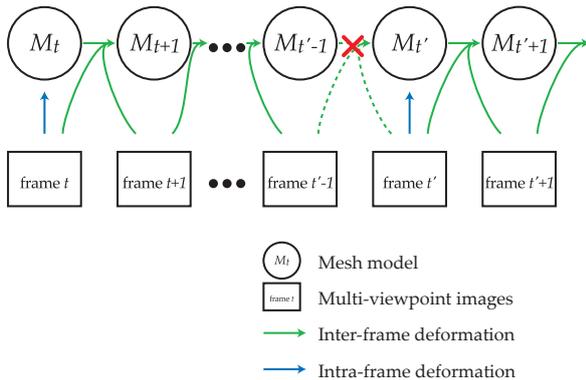


図2 3D video capturing using deformable mesh model

した対象形状をテクスチャとシルエットなどを元にして最適化することで2.5次元の形状復元を行い、あるいはCross [20] や Esteban [21] はシルエットから得られた visual hull を初期形状として space carving を行っている。

これらに対して本研究の提案手法は、形状復元についてのみ考えると Fua の手法に近く、対象形状を3次元の弾性メッシュモデルで表現し、これがさまざまな拘束条件を同時に満たすように変形をすることで対象の形状を復元する。

次に形状と運動の同時復元という点では、従来手法は既知の対象モデルを必要とするものとそうでないものに分類することができる。例えば前者ではHeapによる手のモデルを用いたもの [22] や Plänkner [23] による人体モデルを用いたものを挙げることができ、後者では Vedula [24] による6次元の space carving を挙げることができる。Vedula の手法は入力映像のみを使用するため、より汎用的であるといえるが、従来の space carving と同様に対象のテクスチャのみに依存しているために安定な復元は困難である。

これに対して本手法は形状と運動と同時復元に際してテクスチャだけでなくシルエットなどの複数の情報を統合することで、より安定な復元を可能としている。

1.2 提案手法

本論文で提案する“弾性メッシュモデル”は SNAKES のような動的輪郭モデル [25] [26] の一種であり、我々はこれを複数の情報を統合するための枠組みとして使用する。弾性メッシュモデルは図1のように頂点、辺、面から構成され、各頂点に働く力が平衡するように変形を行う。ここで頂点に作用する力は、対象の形状および運動が満たすべき制約条件を反映して、それらが満たされる位置へと各頂点を移動させるように設計する(後述)。こうして、ある初期形状から変形を開始し、各頂点に作用する力が平衡した状態へと至ったならば、そのときの形状をもって推定結果とする。

本論文では、上記のような弾性メッシュモデルの変形として、大きく以下の2つを提案する。

- フレーム内変形 ある時刻の多視点画像をもとにして、対象のその瞬間の3次元形状を推定する
- フレーム間変形 2つの時刻 t と $t+1$ の多視点画像と、一方の時刻 t における対象形状を表している弾性メッシュモデルをもとにして、 $t+1$ における対象の3次元形状と、 t から $t+1$ への運動を推定する

ここで本研究のポイントは、特に後者のフレーム間変形において、時刻 t の弾性メッシュモデルから、その頂点の移動による変形によって時刻 $t+1$ の対象形状を推定するという点である。このとき、各頂点の移動ベクトルは、まさに対象表面の各点ごと運動ベクトルとみなすことができ、3次元形状と運動の同時推定を行っているといえる。

そして図2に示すように、この2つの変形を用いた全体としての形状および運動推定の枠組み

- Step 1. フレーム内変形によって、時刻 t の形状 M_t を得る。
- Step 2. 次の時刻 $t+1$ の形状 M_{t+1} を、 M_t をフレーム間変形させることによって推定する。
- Step 3. Step 2. を繰り返し、 M_{t+1} から M_{t+2} 、 M_{t+2} から M_{t+3} を順次推定する。
- Step 4. 時刻 t' へのフレーム間変形の結果 $M_{t'}$ がエラーの累積などによって一定の推定の確度を達成できなかった場合、 $M_{t'}$ をフレーム内変形によって推定し、再び Step 2. に戻って $M_{t'}$ から $M_{t'+1}$ を推定する。

を提案する。このようにして、提案する2つの変形による形状および運動推定は、2次元の映像圧縮におけるキーフレームと差分フレームの関係と同様にして3次元形状および運動を推定する。

本論文では、以下まず第2節においてフレーム内変形について説明し、対象の形状を frame-and-skin モデルという形でモデル化することを述べる。次に第

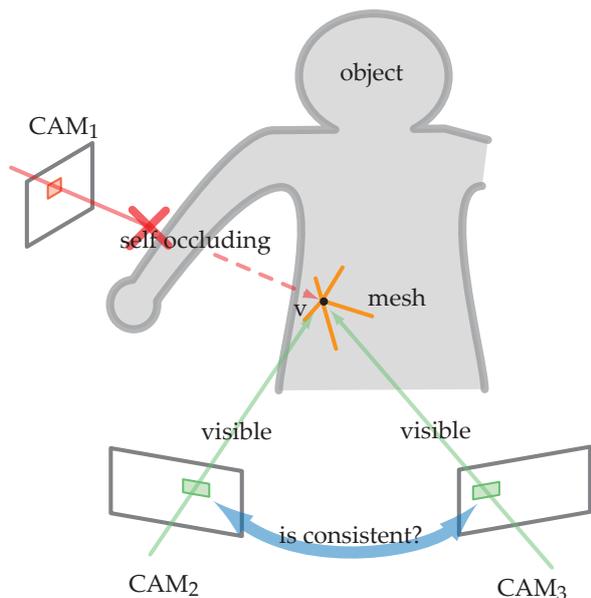


図3 Photometric consistency and visibility

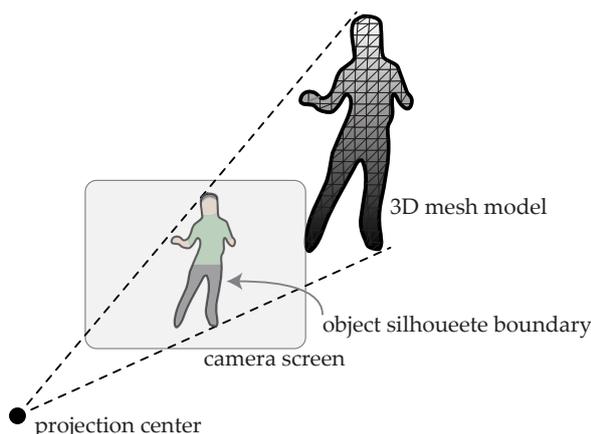


図4 Silhouette constraint

3節において対象の運動に対するモデル化を導入し、heterogeneous deformation モデルを定義することで、対象の3次元形状と運動が同時に推定できることを示す。そして第4節において対象の形状トポロジーが変化した場合にも対処できるように heterogeneous deformation を拡張し、最後に第5節において本論文のまとめとともに今後の課題について考察する。

2 フレーム内変形による3次元形状推定

フレーム内変形の目的は、多視点画像を入力として対象の3次元形状を推定することである。本論文では、各視点の画像から対象のテクスチャとシルエットが得られるものとし、また、対象表面は滑らかかつ連続的で、拡散反射をするものと仮定する。

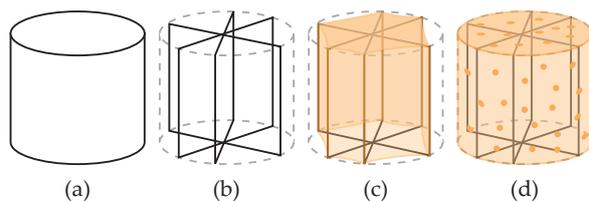


図5 Frame and skin model

フレーム内変形では、以下の3つの制約条件が満たされるように対象の変形を行う。

Photometric 制約 メッシュを構成する各面は、それを観測可能な各視点へと投影したとき、互いにテクスチャが一致しなくてはならない(図3)。

Silhouette 制約 メッシュを各視点に投影したとき、その2次元シルエットは実際に観測された対象のシルエットと一致しなくてはならない(図4)。また各視点での2次元シルエットの輪郭に対応する頂点(contour generator)は、3次元メッシュ上では互いに連続するように位置しなければならない。

Smoothness 制約 対象の形状は局所的には連続かつ滑らかであり、それ自身と交差することは無い。

この3つの制約は以下のような frame-and-skin モデルによって対象の3次元形状をモデル化する。

1. 図5(a)のような対象の3次元形状のモデル化について考えると、
2. まず silhouette 制約が対象の“梁”をつくり(図5(b))、
3. つぎに smoothness 制約がこの“梁”の上に滑らかな“幕”を張る(図5(c))。これにより、対象の大まかな形状が決まる。
4. 最後に photometric 制約によって、有意なテクスチャをもつ部分が“幕”を支持する点をつくり、対象の形状をより正確なものとする(図5(d))。

なおここでは便宜上、梁・幕・点が順に作られるように記述したが、実際にはこれらは同時に推定される。

2.1 頂点に作用する力

Photometric Force テクスチャに基づく力として、頂点 v に働く photometric force $F_p(v)$ を以下のように定義する。

$$F_e(v) \equiv \begin{cases} \nabla E_e(q_v), & \text{if } N(C_v) \geq 2, \\ 0, & \text{otherwise,} \end{cases} \quad (1)$$

ここで C_v と $N(C_v)$ は v を観測可能なカメラの集合とその要素数を表し、 $E_e(q_v)$ は v におけるテクスチャ間(図3)の一致度を表し、以下のように定義す

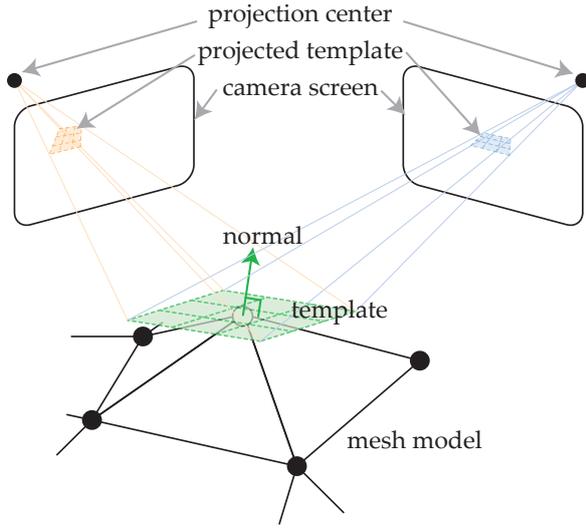


図 6 Template matching

る [16] [24]

$$E_e(\mathbf{q}_v) \equiv \frac{\sum_{c \in C_v} \|p_{w_v, c} - \bar{p}_{w_v}\|^2}{N(C_v)}, \quad (2)$$

ただし w_v は v に接する微小な平面 (図 6) を, $p_{w_v, c}$ は w_v をカメラ c に投影して得られるテクスチャを表し, \bar{p}_{w_v} は $p_{w_v, c}$ の平均を表す. この定義において, 我々は計算量の観点から

- v を観測可能なカメラ集合 C_v は v が $\nabla(E_e(\mathbf{q}_v))$ を計算するために微小移動する間は変化せず,
- v における対象表面は, v に接する微小平面で近似できる

ということを仮定し, また,

- 対象の表面は完全拡散面である

ということを仮定している. これは, 形状と反射パラメータを同時に推定することがきわめて困難であることによる.

この $F_p(v)$ は v を

- 観測可能なカメラが十分に多ければ, テクスチャがより一致する位置へと移動させ,
- 観測可能なカメラが 1 台以下ならば, その位置が他の制約から決められるようにする.

Silhouette Force 各視点で観測される対象シルエットの輪郭を保つための silhouette force $F_s(v)$ を以下のように定義する. 図 7 において $S_{o, c}$ はカメラ 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,

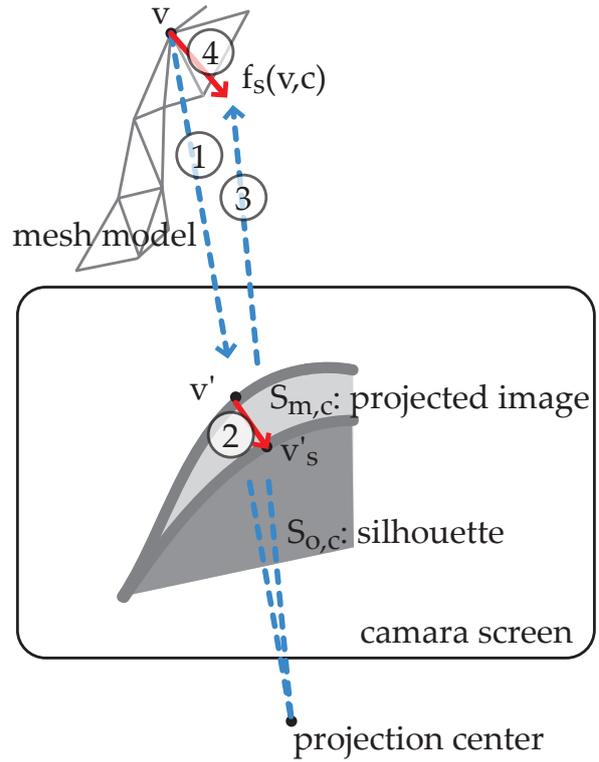


図 7 Silhouette force

1. 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.
2. 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 7 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 7 ④, 後述).
3. それ以外の場合, $f_s(v, c) = 0$.

ここで $f_s(v, c)$ (図 7 ④) は以下のように計算する:

$$f_s(v, c) = (\mathbf{n}_v \cdot \mathbf{d}_{v, v'}) \mathbf{n}_v, \quad (3)$$

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' 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.

$$F_s(v) \equiv \sum_{c \in C_v} f_s(v, c). \quad (4)$$

ここで, $F_s(v)$ は contour generator [20] 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 SNAKE における評価関数の最適化という形で表現できないことを意味している.

Internal Force メッシュを局所的に滑らかに保ち, 自己交差を防ぐために internal force $F_i(v)$ を以下の

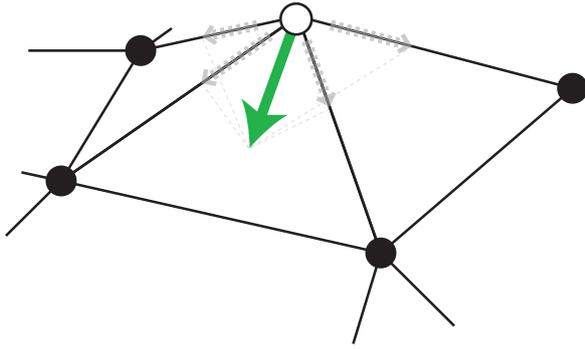


図 8 Internal force

ように定義する (図 8)。

$$F_i(v) \equiv \frac{\sum_j^n q_{v_j} - q_v}{n}, \quad (5)$$

ただし q_{v_j} と n は v に隣接する頂点とその総数を表す。

ここで、このように定義された $F_i(v)$ は頂点間の張力のように働き、メッシュを局部的に滑らかにするだけでなく、メッシュを全体として収縮させる効果も持つ。この収縮という作用は、後述するように、フレーム内変形の初期形状が visual hull であることに関係する。すなわち初期形状が visual hull である場合、真の対象形状は必ずその内部に存在するため、初期形状から内側へと対象表面を探索することはきわめて妥当な戦略であり、internal force の収縮作用はこれを実現する。

以上から、頂点 v に働く力 $F(v)$ は係数 α, β, γ を用いて以下ようになる。

$$F(v) \equiv \alpha F_i(v) + \beta F_e(v) + \gamma F_s(v). \quad (6)$$

2.2 変形アルゴリズム

初期形状 先に述べたように、我々はフレーム内変形の初期形状として、視体積交差法から得られる visual hull を使用する。これは限られたカメラ台数でも visual hull が対象形状に概ね近くなること、そして対象の真の形状を必ず内包するという visual hull の性質が初期形状として適しているためである。

係数の動的制御 SNAKE に類する動的輪郭モデル一般の問題として、internal force とそれ以外の力を固定された係数で合成したとき、動的輪郭モデルが一定以上の曲率をとりえないというものがある。特に今回のように収縮方向に変形が進む場合は、一定以上の凹の部分には変形が進まないということがおきる。これを頂点の移動のみという変形の枠組みの中で解決するためには、力の合成係数を動的に制御する手法が知られているが、本論文ではこの制御の指針として、以下の条件を導入する。

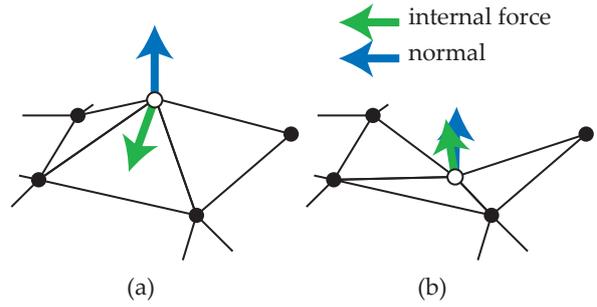


図 9 Vertex normal and internal force. (a) convex, (b) concave surface.

まずはじめに、頂点が凹領域にあるかどうかは法線と $F_i(v)$ が同じ方向を向いているかどうかで簡単に判断できる (図 9)。ここでさらに photometric force $F_e(v)$ がより曲率が大きくなる方向、すなわち法線と逆の向きを示しているときに、internal force の係数 α を小さくする。また対象の輪郭を保つことを優先するため、 $F_s(v) = 0$ であることも条件とする。

変形プロセス 以上からフレーム内変形の計算プロセスは以下ようになる。

Step 1. visual hull を初期形状とする。

Step 2. 各頂点の F_v を計算する。

Step 3. もし

- 法線 n_v と internal force $F_i(v)$ が同じ方向を指し、
 - $F_e(v)$ がこれらと逆を指し、
 - $F_s(v) = 0$ ならば、
- α を小さくする。

Step 4. 頂点を F_v に沿って移動させる。

Step 5. 全頂点の移動量が十分小さいとき、変形を終了する。そうでなければ Step 2. に戻る。

Step 6. 平衡状態となったメッシュを推定結果とする。

2.3 評価実験

図 10 のような撮影空間で 25 台のカメラで撮影された画像 (図 11) を使用した実験結果を示す。初期メッシュは図 12 のような約 12,000 頂点からなる visual hull であり、これを係数の動的制御を行わずに変形した結果が図 13、動的制御を行った結果が図 14、図 16 である。明らかに動的制御を行った場合がより対象本来の形状に近いことが確認できる。

3 フレーム間変形による 3 次元形状・運動推定

前節のフレーム内変形において、我々は対象の形状に対するモデル化として frame-and-skin モデル (図

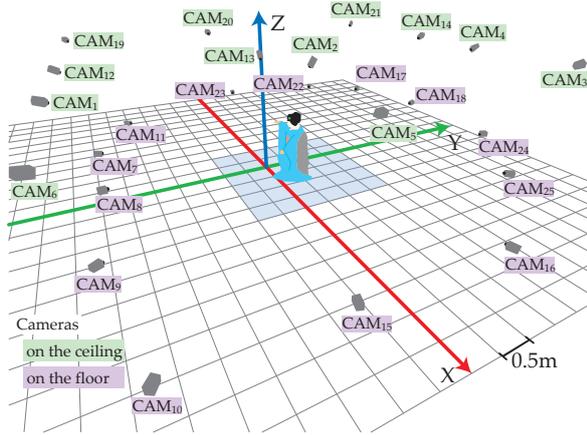


図 10 Camera arrangement



図 11 Input images

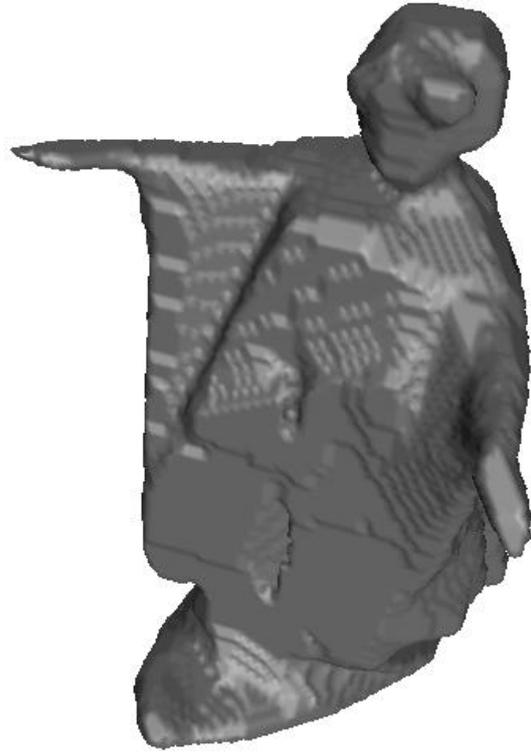


図 12 Visual hull

5)を導入した．このフレーム間変形においては，これを拡張して対象の動作に対するモデル化を行う．

まず基本となるアプローチは，フレーム内変形においてはその時刻における対象の形状を表現するようにメッシュの変形を行っていたところを，フレーム間変形ではある時刻 t の対象形状を表したメッシュを，次の時刻 $t+1$ における対象形状を表現するように変形することである．これにより，次の時刻 $t+1$ における形状とその間の動作が同時に得られる．つまり各頂点のフレーム間での移動ベクトルを推定することが，この変形の目的である．

ここで，図 17(a)において，左のメッシュの各頂点の移動ベクトルを求めて右のメッシュを得る場合を考えたとき，移動ベクトルを個別に推定したとしたり，同図 (b) のように移動ベクトルが互いに交差し，得られる形状もまた自己交差を起こしたのも許容されることになる．一方で全移動ベクトルは 1 つの回転・並進で記述可能としたならば，いわゆる剛体運動となり (同図 (c))，またいくつかのパーツごとに

分かれて剛体運動をするならば多関節運動に (同図 (d))，そして局所的な形状を保ちつつ自由に変形したならば warping 運動 (同図 (e)) となる [27]．

対象の運動に対するモデルの導入とは，まさにこの移動ベクトル間の制約条件のことであり，本研究では，対象の運動を warping と剛体運動の混合モデルであると仮定する．すなわちメッシュを構成する頂点を warping 領域と剛体領域に分類し，それぞれにその物理モデルに沿った変形過程を適用する．このようにメッシュを構成する頂点をそれぞれが持つ属性に応じて変形過程をコントロールすることから，これを heterogeneous deformation モデルと呼ぶことにする．

以下，まずフレーム間変形で使用する制約条件と，それらに対応する力の定義を行った後に，頂点をその属性に応じて分類する方法について述べる．

3.1 制約条件

フレーム間変形では，以下の 5 つの制約条件を使用する．なおここでは時刻 t から $t+1$ への変形であるとして記述する．

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 $t+1$ の視点の間で互いに一致しなくてはならない．

Silhouette 制約 メッシュを各視点に投影して得られるシルエットは，時刻 $t+1$ において観測され

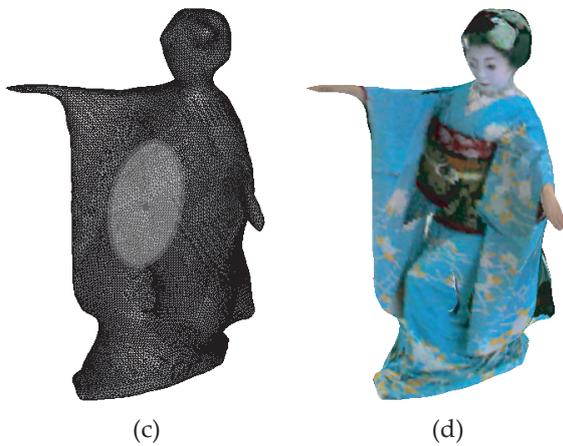
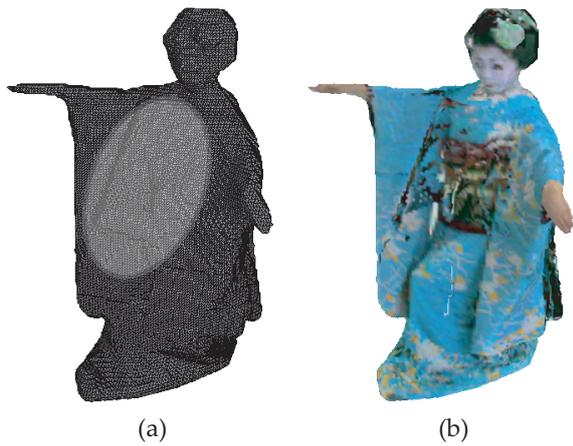


図 13 Rendering result. (a) and (b): the initial mesh and its rendering result, (c) and (d): deformed mesh and its rendering result.

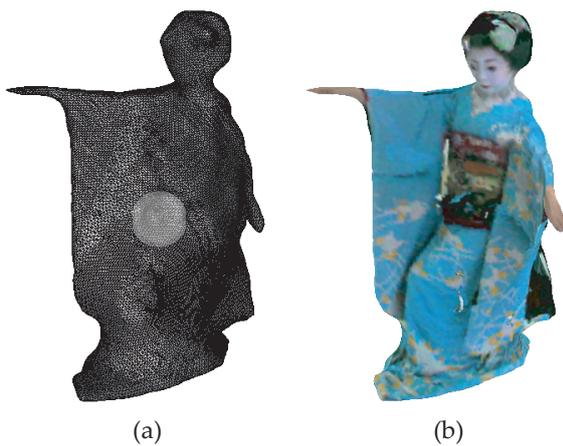


図 14 Deformed shape with adaptive coefficient control. (a) mesh model, (b) rendering result



図 15 Reference image (observed from CAM₁₂ in Figure 10)

	Deformation with fixed coefficient	Deformation with adaptive coefficient
Mesh model		
Rendering with CAM ₁₂		
Rendering without CAM ₁₂		

図 16 Rendering result

た対象のシルエットと一致してはならない。
 Smoothness 制約 メッシュは局所的に滑らかな形状をもち、自己交差を起こさない。
 Motion flow 制約 各頂点は visual hull から推定された motion flow に沿って変形を行う(後述)。
 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。

これらのうち最初の3つの制約条件は、前節同様に frame-and-skin モデルとして対象の形状をモデル化している。これに対して残る2つの制約条件は、頂点

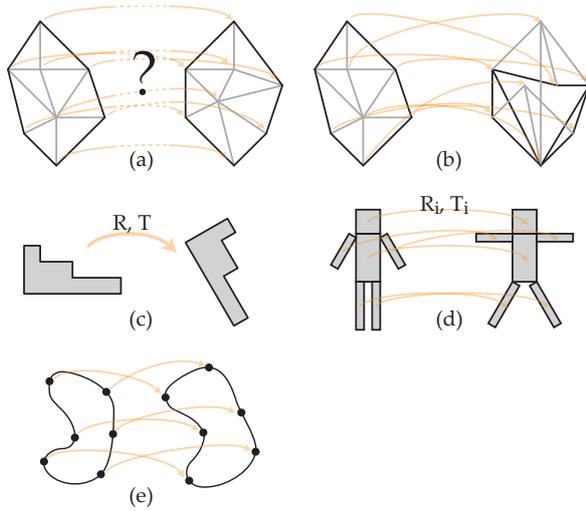


図 17 Object motion models

の移動に関する制約を表している．まず motion flow 制約とは，2 つの時刻の visual hull の間で計算された粗い運動推定結果である．これは以下のようにして計算される．

- Step 1. 2 つ時刻の多視点シルエットから，それぞれ visual hull を計算する．
- Step 2. それぞれの visual hull の境界 voxel の集合を V_t, V_{t+1} とする．
- Step 3. 3 次元空間中の点群 V_t から V_{t+1} への変形を計算する [28]．
- Step 4. 得られた遷移ベクトルを粗い運動推定結果とみなす．

一方 inertia 制約は対象運動の連続性への仮定である．

3.2 頂点に働く力

前述の 5 つの制約条件に対応する力の定義を行うにあたって，以下のように記号を定義する．まず時刻 $t+1$ における頂点を v ，そしてその位置を 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．

Photometric Force まず photometric force $F_e(v)$ を，2 つの時刻間でのテクスチャが一致するように定義する．

$$F_e(v) \equiv \nabla E_e(q_v), \quad (7)$$

ここで $E_e(q_v)$ は “photo-consistency” 関数 [16] [24] を表し，

$$E_e(q_v) \equiv \frac{\sum_{c \in C_{v_0}} \|p_{w_{v_0},c} - \overline{p_{w_{v_0},w_v}}\|^2 + \sum_{c \in C_v} \|p_{w_v,c} - \overline{p_{w_{v_0},v}}\|^2}{N(C_{v_0}) + N(C_v)}, \quad (8)$$

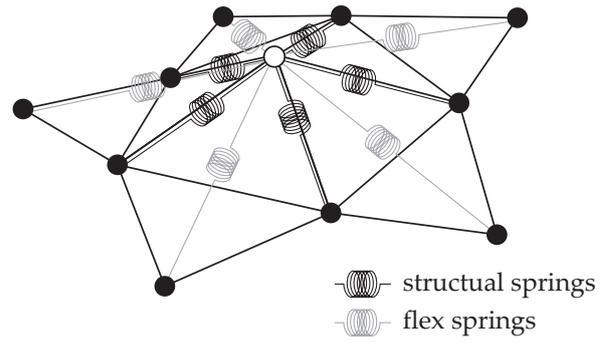


図 18 Internal force

とする．ただし c は C_v に含まれる視点を， $N(C_v)$ は C_v に含まれる視点の数を， v_0 は v の初期位置，すなわち時刻 t における頂点の位置を， $p_{w_v,c}$ は時刻 $t+1$ において視点 c で観測されるテクスチャを， $p_{w_{v_0},c}$ 時刻 t において視点 c で観測されるテクスチャを， $\overline{p_{w_{v_0},w_v}}$ は 2 つの時刻のテクスチャ $p_{w_{v_0},c}$ および $p_{w_v,c}$ 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．

Internal Force フレーム内変形における internal force の定義には，smoothness 制約と共に収縮という作用があった．これはフレーム内変形の初期メッシュが visual hull であることから有効な定義であったが，フレーム間変形では，初期状態すなわち時刻 t における形状が，時刻 $t+1$ の形状を内包するという仮定は成り立たないため，smoothness 制約のみを持つように再定義する必要がある．そこでフレーム間変形では，図 18 に示すようなバネモデルに基づく internal force を以下のように定義する．

$$F_i(v) \equiv \sum_j^n k_j (\|q_{v_j} - q_v\| - l_{v,v_j}) - d_v q_v \quad (9)$$

ここで v_j と n は v に隣接する頂点とその総数を， k_j は v と v_j の間のバネ定数を， l_{v,v_j} はバネの自然長を， d_v はバネのダンパ定数である．

Silhouette Force silhouette force に関しては，それが保持すべきシルエット輪郭が時刻 $t+1$ のものになるだけであるため，フレーム内変形における定義 (2.1 節) と同様である．

Drift Force 先に述べたように，我々は各時刻の多視点シルエットが得られることを仮定している．そしてこのシルエットから各時刻の visual hull が得られるため，簡単な頂点集合の変形アルゴリズム [28] によって粗い運動推定を行うことができる．これを時刻 t のメッシュを構成する頂点集合と，時刻 $t+1$ の visual hull の間で行い，推定結果を線群

$$L = \{l_i | i = 1, \dots, N(V)\}, \quad (10)$$

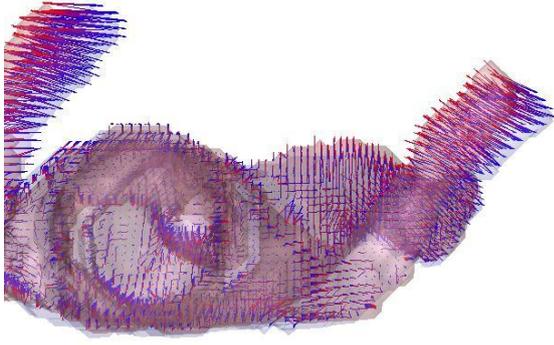


図 19 Roughly estimated motion flow lines

とする．ここで V と $N(V)$ は時刻 t におけるメッシュの頂点群とその頂点数を， l_i は V の i 番目の点を始点とする線を表す．図 19 に得られた推定結果の一例を示す．時刻 t 側が青，時刻 $t+1$ 側が赤で色付けされている．

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．

$$E_d(q_v) \equiv \|s_{l_v} - p_{l_v,v}\|^2 - \|q_v - p_{l_v,v}\|^2. \quad (11)$$

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

$$F_d(v) \equiv \nabla E_d(q_v) \quad (12)$$

と定義する．

Inertia Force 図 2 に示したように，フレーム間変形を逐次実行していた場合，時刻 $t-1$ から t へのフレーム間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．ここでフレーム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．そこでまずは外挿によって時刻 $t+1$ における推定位置 \hat{q}_v^{t+1} を計算する．

$$\begin{aligned} \hat{q}_v^{t+1} &= q_v^t + (q_v^t - q_v^{t-1}) \\ &= 2q_v^t - q_v^{t-1}. \end{aligned} \quad (13)$$

ただしフレーム間隔は一定であるとしている．

あとは drift force の定義と同様に，inertia force を

$$F_n(v) \equiv \nabla E_n(q_v), \quad (14)$$

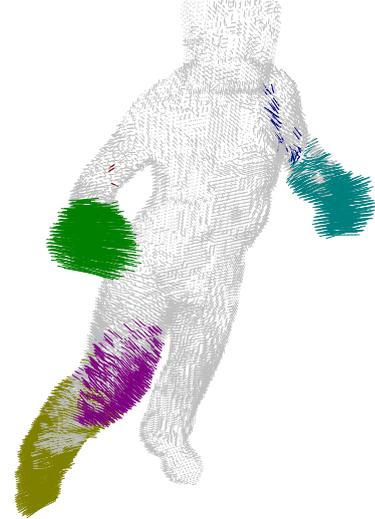


図 20 Clustered motion flow lines

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ことができる．

頂点に働く力 以上から頂点 v に働く力 $F(v)$ を係数 $\alpha, \beta, \gamma, \delta, \varepsilon$ を用いて以下のように定義する．

$$F(v) \equiv \alpha F_i(v) + \beta F_e(v) + \gamma F_s(v) + \delta F_d(v) + \varepsilon F_n(v). \quad (15)$$

3.3 Heterogeneous Deformation

先に述べたように，フレーム間変形では，対象が warping 運動領域と剛体運動領域からなるとモデル化し，その領域ごとに変形プロセスを変化させる．すなわちメッシュを構成する頂点をその特徴によって分類し，それに応じた変形プロセスを当てはめる．

ここでは warping と剛体という物理的な特徴に基づく分類と，頂点におけるテクスチャの有無という photometric な特徴に基づく分類という 2 つの観点から頂点を分類し，それらに応じた変形過程を定義する．

運動に基づく分類 先に述べたように，drift force と inertia force では時刻 $t+1$ における頂点位置を，各頂点ごとに推定している．そこで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 warping 部分に分類することにする．

- Step 1. 各頂点と推定された移動先を結んだ線群を得る．
- Step 2. このうち，長さが一定の閾値以下のものを除去する．除去されたものは warping 領域とみなす．
- Step 3. まず線群をその基点の空間的近接性に基づい

てクラスタリングする．
 Step 4. こうしてえ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クラスタリングを行う．こうして得られたクラスタをそれぞれ剛体部分とみなす．

得られたクラスタリング結果を図 20 に示す．灰色が warping 領域を表し，色付けられた部分がそれぞれ個別の剛体領域である．

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 [29] [30] [31] とともに，internal force におけるバネ定数を大きくする．

テクスチャに基づく分類 式 (8) で定義したように，photometric force を求める段階で我々は既に各頂点が有意なテクスチャを持つかどうかを判断することができる．ここで特に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ることで，有意なテクスチャを持つ頂点が周囲を導くようにする．

Heterogeneous Deformation 以上から heterogeneous deformation における変形アルゴリズムを以下のように定義する．

- Step 1. 時刻 t における対象の形状を表したものを初期状態とする．
- Step 2. drift force $F_d(v)$ と inertia force $F_n(v)$ の運動ベクトル推定を行う
- Step 3. 頂点を運動とテクスチャに基づいて分類する
 - Step 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**: 剛体部分に属するか，あるいは **Ca-2**: warping 部分に属するか分類する．
 - Step 3.2. **Ca-1** とされた頂点のバネ定数を大きくする．
- Step 4. 反復的に変形を行う．
 - Step 4.1. 各頂点に働く力 $F(v)$ を計算する
 - Step 4.2. 有意なテクスチャを持つとされた各頂点 v について，その周囲の頂点 v_j に働く力を以下のように修正する．

$$F(v_j) = \omega F(v) + (1 - \omega) F_{\text{prev}}(v_j),$$

$$\omega = e^{-D_g(v, v_j)},$$

(16)

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．

- Step 4.3. 計算された力に沿って，頂点を移動する．
- Step 4.4. 移動量が十分小さいときは変形終

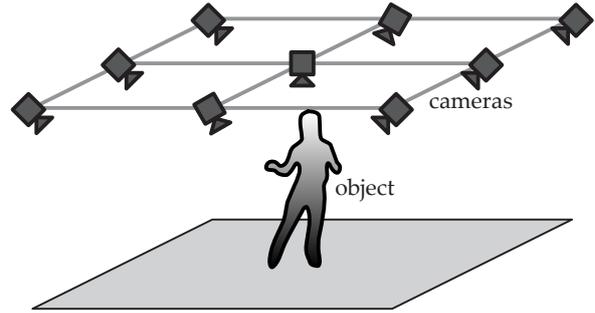


図 21 Camera arrangement

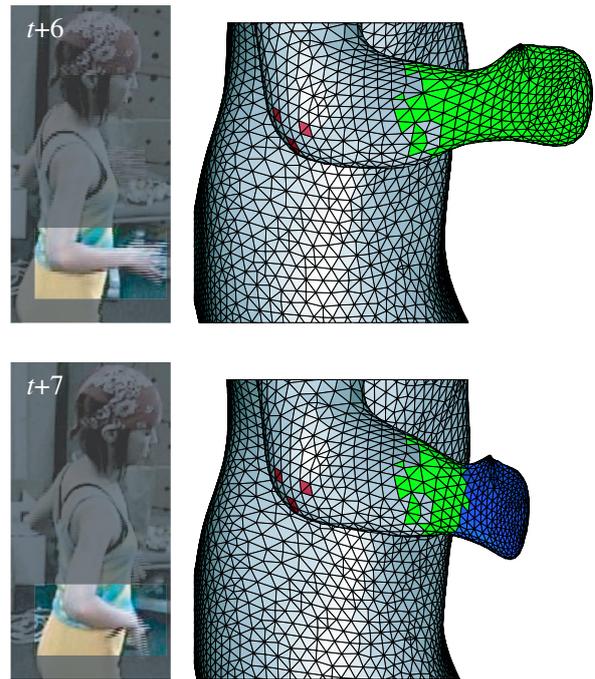


図 24 Successive deformation results (detailed, side view)

- 了，そうでなければ Step. 4.1 へ．
- Step 5. 得られたメッシュ形状を時刻 $t+1$ における対象の形状とみなし，変形ベクトルと t から $t+1$ への対象の運動とみなす．

3.4 評価実験

図 22, 23, 24 は 8 フレームにわたるフレーム間変形の結果である．実験は図 21 に示すように対象を囲むように置かれた 9 台のカメラからの映像を元に行い，図 22 は左から順に，観測画像，フレームごとに得られた visual hull，heterogeneous deformation によって得られたメッシュモデルである．またこれらの図において色付けられた部分は剛体領域とみなされた部分である．各メッシュはおよそ 12,000 頂点からなり，heterogeneous deformation による処理時間は 1 フレームあたり Xeon 3GHz の PC で 20 分であ

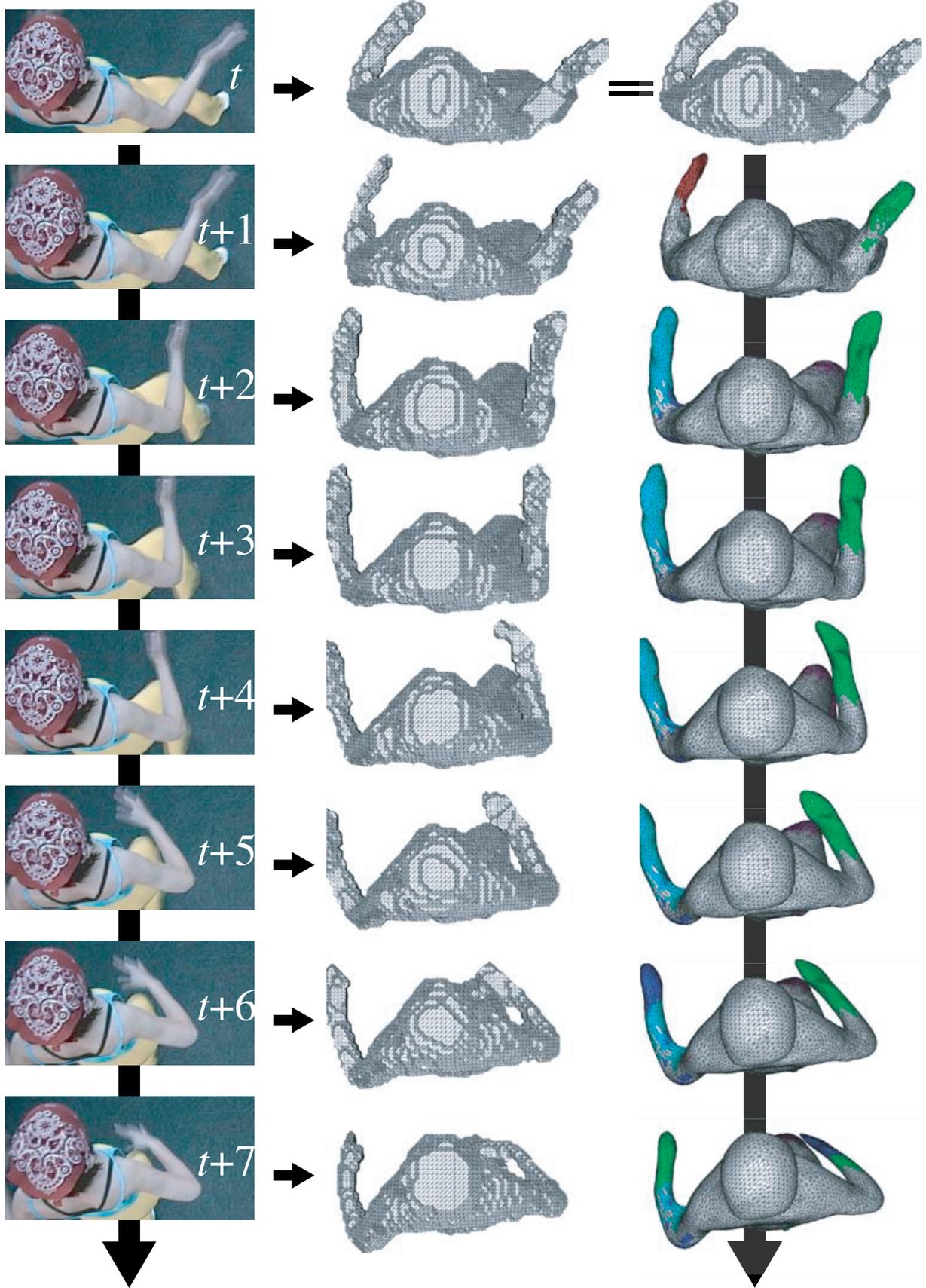


图 22 Successive deformation results

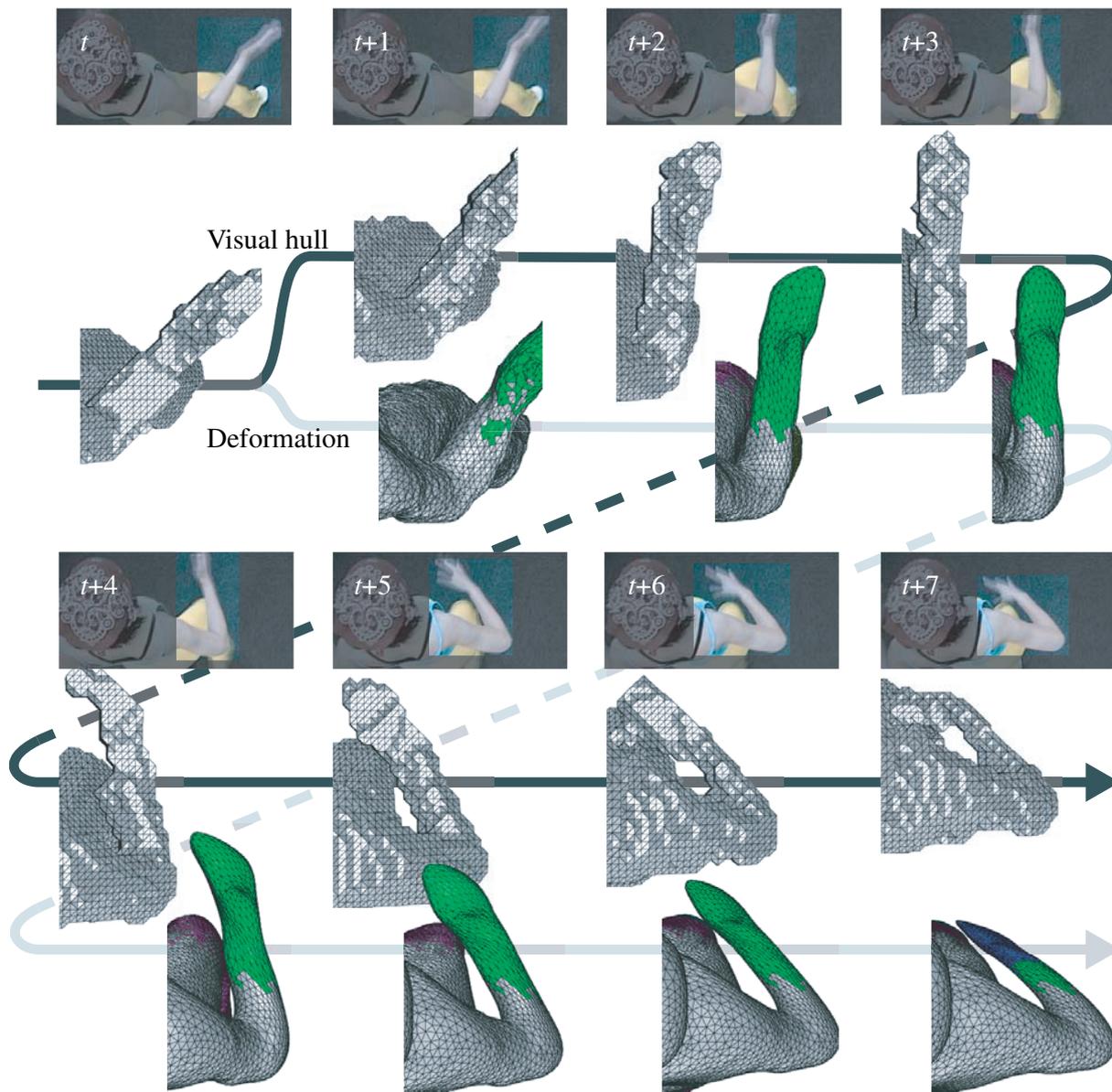


図 23 Successive deformation results (detailed)

る．また係数は固定値 $\alpha = 0.2, \beta = 0.2, \gamma = 0.2, \delta = 0.3, \epsilon = 0.1$ を使用している．

これらの結果から，

- 提案した弾性メッシュモデルは部分的に剛体とみなすことのできる対象の運動を表現することができる．例えば図 22 および 23 において，対象の腕がそれぞれ剛体領域として認識されている．
- 複数フレームに渡る変形を通して，我々の弾性メッシュモデルはメッシュの全体的および局所的トポロジーを共に保っている．すなわちあるフレームにおける 1 つの頂点に対応する頂点を，

別のどのフレームにおいても知ることができる．図 23 において，左上から右下へと変形が進む過程を見ると，提案するフレーム間変形ではフレーム間で各頂点の対応付けが容易に行うことができる一方で，各フレーム独立に求めたメッシュでは，そもそもメッシュの頂点数，そして頂点間の接続関係に一貫性が存在しないため，対応付けを行うことはできない．

- 図 24 において，対象の右手部分が $t+7$ において 2 つの剛体部分へと分割されたことがわかる．このことから，heterogeneous deformation 結果を元にした対象の関節位置の検出など，均質なメッシュから変形を開始して順次運動モデルを

学習していくことが可能であることがわかる。

4 トポロジー変化に対応した 3 次元形状・運動推定

最後に対象の形状トポロジーが変化した場合への拡張を行う。従来の研究では、対象のトポロジー変化に合わせてメッシュのトポロジーもまた変化させる手法なども提案されてきたが [32] [33]、本研究ではトポロジー変化によって隠れた面をそのまま保持するというアプローチを採る。これはフレーム間でトポロジーを保つことが運動解析やデータ圧縮の上で有用だからである。

このようなアプローチを採る上で heterogeneous deformation の定義を考えると、ここまでのすべての制約条件とそれに対応する力の定義は、すべて各頂点毎のメッシュの接続関係の意味で近接した頂点のみを考慮した、すなわち測地距離に基づいた定義であった。これに対して対象のトポロジーが変化するとき、例えば図 24 において対象がそのまま手を腰に当てる場合について考えると、測地距離的には遠くても、ユークリッド距離的には近い頂点同士が接することがわかる。

この節ではこのような状況にも有効に変形を行うことができるように、ユークリッド距離に基づいた力として global collision detection とそれに基づく repulsive force $F_r(v)$ を導入する。まずはじめに、我々は photometric force など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。ここで、他の頂点に接する可能性のある頂点では、 $C_v = \emptyset$ であることはカメラが対象を囲むように位置していることから明らかなので、collision detection はこのような頂点に対してのみ行えばよいといえる。そこで、

Step 1. すべての頂点において、repulsive force $F_r(v) = 0$ とする。

Step 2. $C_v = \emptyset$ である頂点の集合 V_0 に含まれる頂点それぞれについて、

Step 2.1. V_0 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 $l_{\min}(v)$ 以下である頂点を探す。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。

Step 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 $F_r(v)$ に加える。

$$f_r(v, v') = \frac{q_{v'} - q_v}{\|q_{v'} - q_v\|^3}. \quad (17)$$

という手順で repulsive force を計算し、これを $F(v)$ に加え、

$$F(v) \equiv \alpha F_i(v) + \beta F_e(v) + \gamma F_s(v) + \delta F_d(v) + \varepsilon F_n(v) + \zeta F_r(v). \quad (18)$$

と定義する。

4.1 評価実験

図 25 と 26 にトポロジー変化を伴う運動を行う対象を撮影した実画像を用いた実験結果を示す。カメラ配置などは前節と同様である。

図 25 は左から順に、観測画像、visual hull、フレーム内変形によって各フレーム独立に推定された対象形状、そして提案したフレーム間変形によって得られた変形結果をそれぞれ表している。メッシュはそれぞれ約 12,000 頂点から構成され、Xeon 3.0GHz の PC による処理時間は 1 フレームにつき約 2 時間である。なお係数は固定値 $\alpha = 0.15, \beta = 0.15, \gamma = 0.2, \delta = 0.2, \varepsilon = 0.1, \zeta = 0.2$ を使用した。

この結果から、

- 提案手法が対象のトポロジーが変化する状況においてもその形状および運動を推定することができること (図 25)
- トポロジーが変化した場合でも、図 26 において確認できるように、提案手法では全体的、および局所的トポロジーを保つことができ、1 つ 1 つの頂点の対応付けがフレーム間で可能である一方、visual hull やフレーム内変形では全体的・局所的トポロジーが共に変化してしまうこと

がわかる。

5 まとめ

本論文では、弾性メッシュモデルの変形という 1 つの枠組みを用いて、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。特に、対象の形状を表すメッシュを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 [1] [2] [3] への展開が可能であると共に、提案した heterogeneous deformation モデルでは、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。

一方で、本研究ではその入力として対象のシルエットを仮定したため、形状および運動推定の精度はシルエットの精度に大きく依存する。今後は、3 次元形状・運動を推定する過程でシルエットのエラーリカバリを行うなど、より頑健なアルゴリズムへの拡張が必要である。

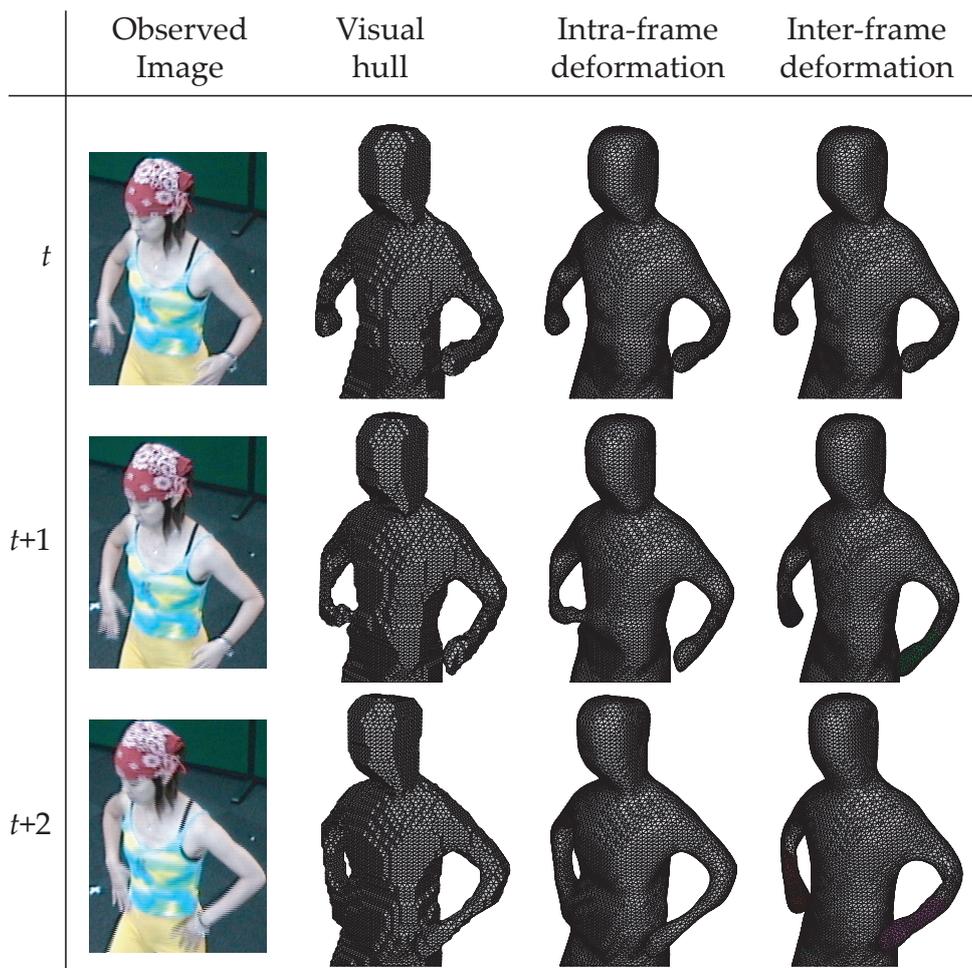


图 25 Successive deformation results

参考文献

- [1] H. M. Briceño, P. V. Sander, L. McMillan, S. Gortler, and H. Hoppe. Geometry videos: A new representation for 3d animations. In *Proc. of SIGGRAPH*, pp. 136–146, 2003.
- [2] L. Ibarria and J. Rossignac. Dynapack: space-time compression of the 3d animations of triangle meshes with fixed connectivity. In *Proc. of SIGGRAPH*, pp. 126–135, 2003.
- [3] H. Habe, Y. Katsura, and T. Matsuyama. Skin-off: Representation and compression scheme for 3d video. In *Proc. of Picture Coding Symposium*, San Francisco, Dec 2004.
- [4] S. Moezzi, L.-C. Tai, and P. Gerard. Virtual view generation for 3d digital video. *IEEE Multimedia*, pp. 18–26, 1997.
- [5] T. Kanade, P. Rander, and P. J. Narayanan. Virtualized reality: Constructing virtual worlds from real scenes. *IEEE Multimedia*, pp. 34–47, 1997.
- [6] T. Wada, X. Wu, S. Tokai, and T. Matsuyama. Homography based parallel volume intersection: Toward real-time reconstruction using active camera. In *Proc. of CAMP*, pp. 331–339, Padova, Italy, September 2000.
- [7] E. Borovikov and L. Davis. A distributed system for real-time volume reconstruction. In *Proc. of CAMP*, pp. 183–189, Padova, Italy, September 2000.
- [8] K. M. Cheung, T. Kanade, J.-Y. Bouguet, and M. Holler. A real time system for robust 3d voxel reconstruction of human motions. In *Proc. of CVPR*, pp. 714–720, South Carolina, USA, June 2000.
- [9] 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*, Vol. 96, pp. 393–434, De-

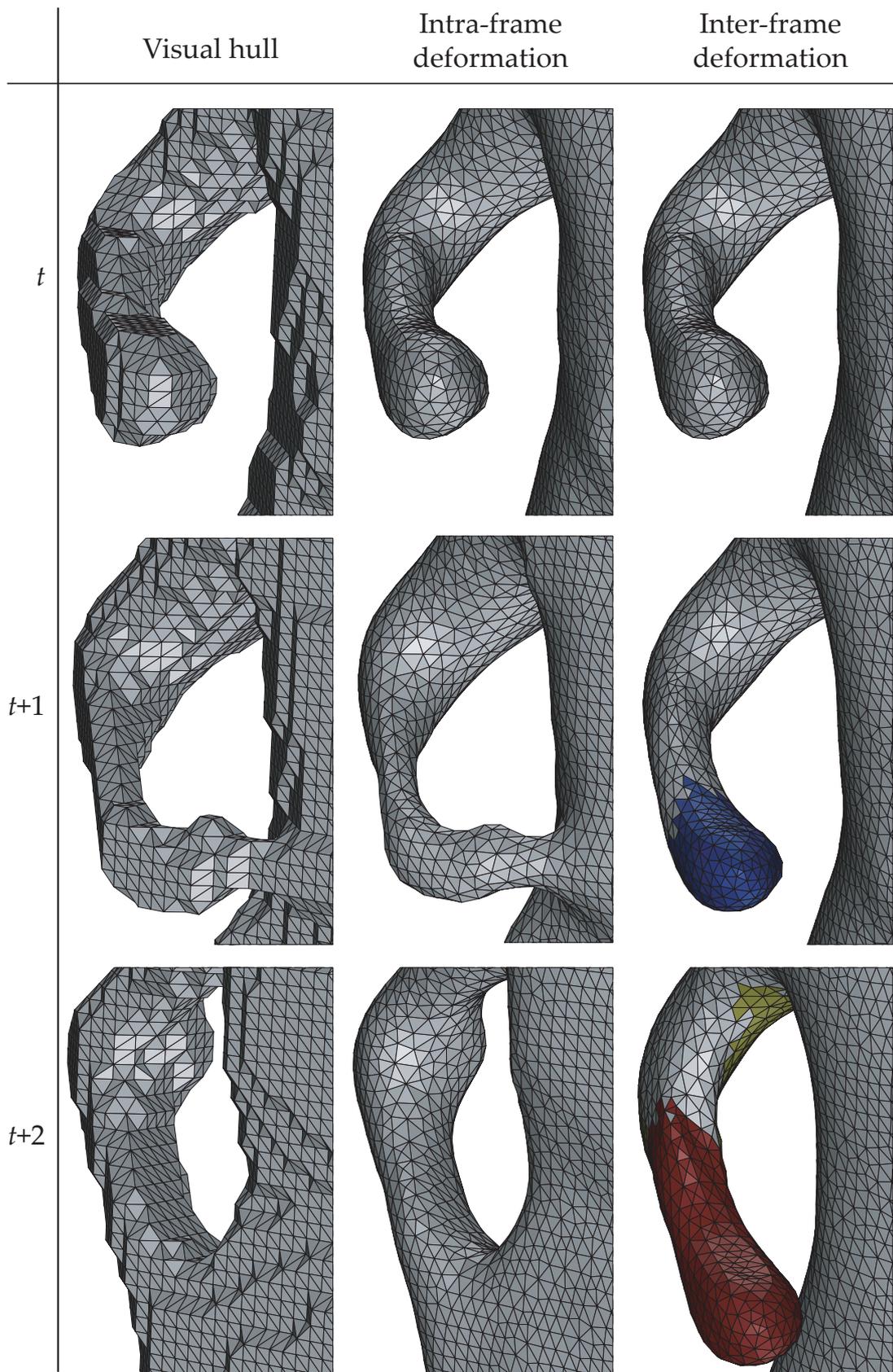


图 26 Successive deformation results (detailed)

- cember 2004.
- [10] K.-Y. K. Wong and R. Cipolla. Structure and motion from silhouettes. In *Proc. of ICCV*, Vol. II, pp. 217–222, Vancouver, Canada, July 2001.
- [11] T. Wada, H. Ukida, and T. Matsuyama. Shape from shading with interreflections under proximal light source, -3d shape reconstruction of unfolded book surface from a scanner image-. In *Proc. of ICCV*, pp. 66–71, Jul 1995.
- [12] S. Vedula, S. Baker, P. Rander, R. Collins, and T. Kanade. Three-dimensional scene flow. In *Proc. of ICCV*, Vol. 2, pp. 722 – 729, September 1999.
- [13] M. D. Wheeler, Y. Sato, and K. Ikeuchi. Consensus surfaces for modeling 3d objects from multiple range images. In *Proc. of ICCV*, p. 917. IEEE Computer Society, 1998.
- [14] A. Laurentini. How far 3d shapes can be understood from 2d silhouettes. *PAMI*, Vol. 17, No. 2, pp. 188–195, 1995.
- [15] W. B. Seales and O. D. Faugeras. Building three-dimensional object models from image sequences. *CVIU*, Vol. 61, No. 3, pp. 308–324, 1995.
- [16] K. N. Kutulakos and S. M. Seitz. A theory of shape by space carving. In *Proc. of ICCV*, pp. 307–314, September 1999.
- [17] P. Fua and Y. G. Leclerc. Using 3-dimensional meshes to combine image-based and geometry-based constraints. In *Proc. of ECCV*, pp. 281–291, 1994.
- [18] P. Fua. Reconstructing complex surfaces from multiple stereo views. In *Proc. of ICCV*, pp. 1078–1085, Jun 1995.
- [19] P. Fua and Y. G. Leclerc. Object-centered surface reconstruction: combining multi-image stereo and shading. *IJCV*, Vol. 16, No. 1, pp. 35–55, 1995.
- [20] G. Cross and A. Zisserman. Surface reconstruction from multiple views using apparent contours and surface texture. In *Proceedings of NATO Advanced Research Workshop on Confluence of Computer Vision and Computer Graphics*, pp. 25–47. Kluwer Academic Publishers, 2000.
- [21] C. H. Esteban and F. Schmitt. Multi-stereo 3d object reconstruction. In *Proc. of 3DPVT*, pp. 159–167, Padova, Italy, June 2002.
- [22] T. Heap and D. Hogg. Towards 3d hand tracking using a deformable model. In *Proceedings of the 2nd International Conference on Automatic Face and Gesture Recognition*, pp. 140–145, 1996.
- [23] R. Plänkers and P. Fua. Articulated soft objects for multiview shape and motion capture. *PAMI*, Vol. 25, No. 9, pp. 1182–1187, 2003.
- [24] S. Vedula, S. Baker, S. Seitz, and T. Kanade. Shape and motion carving in 6d. In *Proc. of CVPR*, June 2000.
- [25] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *IJCV*, Vol. 1, No. 4, pp. 321–331, 1988.
- [26] H.-Y. Shum, M. Hebert, K. Ikeuchi, and R. Reddy. An integral approach to free-formed object modeling. In *Proc. of ICCV*, pp. 870–875, Jun 1995.
- [27] A. J. Yezzi and S. Soatto. Deformation: Deforming motion, shape average and the joint registration and approximation of structures in images. *IJCV*, Vol. 53, No. 2, pp. 153–167, 2003.
- [28] D. Burr. A dynamic model for image registration. *Computer Graphics and Image Processing*, Vol. 15, pp. 102–112, 1981.
- [29] B. K. P. Horn. Closed-form solution of absolute orientation using unit quaternions. *Journal of the Optical Society of America. A*, Vol. 4, No. 4, pp. 629–642, Apr 1987.
- [30] 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.
- [31] N. Ohta and K. Kanatani. Optimal estimation of three-dimensional rotation and reliability evaluation. In *Proc. of ECCV*, pp. 175–187, Germany, Jun 1998.
- [32] A. J. Yezzi and S. Soatto. Stereoscopic segmentation. *IJCV*, Vol. 53, No. 1, pp. 31–43, 2003.
- [33] J. Brendo, T. M. Lehmann, and K. Spitzer. A general discrete contour model in two, three, and four dimensions for topology-adaptive multichannel segmentation. *PAMI*, Vol. 25, No. 5, pp. 550–563, 2003.