

欠損および過抽出を含む時系列シルエットからの三次元形状獲得

豊浦正広[†] 飯山将晃^{††} 船富卓哉^{†††}
角所 考^{†††} 美濃 導彦^{†††}

視体積交差法では、形状獲得に用いるシルエットの数が多いほど、実形状に近い獲得形状が得られる。しかし、実際に設置できるカメラ台数には限界があるので、用いることのできるシルエットの数にも限界がある。そこで、時系列で得られる視体積を統合する手法が提案されてきた。この手法では、視体積から三次元特徴点を抽出し、物体の剛体運動を推定して時系列の視体積を統合する。これにより、用いるシルエットの数を増やすことができる。このときに問題となるのが、得られるシルエットに欠損領域や過抽出領域が含まれるときには、異なるフレーム間で対応する特徴点が抽出されないことである。この結果、一部のフレーム間で運動推定誤差が生じ、獲得される形状に大きな欠損が起こることがある。

本研究では、獲得形状の突出部分に多く抽出される表出点に注目し、時系列視体積を統合する。視体積統合を行うフレーム間で、表出点がどれほど保存されるかを評価し、この評価に基づいて統合すべきフレームの選択を行う。これにより、物体形状の突出部分が保存され、実形状に近い形状を得られる。

3D Shape Reconstruction from Incomplete Silhouettes in Time Sequences

MASAHIRO TOYOURA,[†] MASAAKI IYAMA,^{††}
TAKUYA FUNATOMI,^{†††} KOH KAKUSHO^{†††}
and MICHHIKO MINOH^{†††}

In volume intersection method, 3D shapes are reconstructed from silhouettes which are obtained by multiple cameras. The more number of cameras are used, the more accurate shapes are reconstructed in the volume intersection method. Since the number of cameras is limited, the accuracy of reconstructed shapes is also limited. In recent works, silhouette integration methods have been proposed for shape reconstruction. By the silhouette integration, the number of silhouettes is increased. The silhouettes of a rigid object in time sequences are integrated by rigid object motion. The motion is estimated by using a kind of 3D feature points. It is known that the error of motion estimation causes the missing of reconstructed shape.

In this research, we discuss how the error of motion estimation affects to the incompleteness of extracted silhouettes. Based on the discussion, we propose an integration method for incomplete silhouettes. In the method, outcrop points, which are kinds of feature points for motion estimation, play an important role. By referring how many outcrop points are included in the reconstructed shape of different frame, estimated motion can be evaluated. Based on the evaluation, silhouettes in time sequences can be integrated with less missing parts.

1. はじめに

視体積交差法¹⁾²⁾は、複数のカメラで物体を観測し、観測画像から得られるシルエットから形状を獲得する

手法である。各シルエットから、物体が存在しうる錐体状の領域を計算し、すべてのシルエットから得られる領域の積領域を求めることで形状を獲得する。視体積交差法における獲得形状は、視体積と呼ばれる。視体積交差法では、原理的に、カメラ台数が多ければ多いほど視体積に含まれる実物体領域以外の余分な領域が減少し、実形状に近い形状を獲得することができる。しかし、現実には数十台を越えるカメラを設置することができない。ここで、対象物体の運動が剛体運動であるとすると、物体が運動することにより、カメラと

[†] 京都大学大学院 情報学研究所
Graduate School of Informatics, Kyoto University
^{††} 京都大学大学院 経済学研究所
Graduate School of Economics, Kyoto University
^{†††} 京都大学 学術情報メディアセンター
ACCMS, Kyoto University

物体の相対的な位置関係が変化し、各フレームで観測される視体積を新たな視点位置のカメラセットから得られる視体積とみなすことができる。このアイデアに基づいて、時系列で得られる視体積を統合することで、実形状に近い形状を獲得する手法がこれまでに提案されている。Cheungら³⁾は、各フレームの視体積と画像から Colored Surface Point(CSPs)を抽出し、物体の剛体運動を観測画像から推定した。時系列で変化するカメラと物体の位置関係は物体の剛体運動で記述できるので、この運動を基に時系列シルエットの統合を行うことができる。我々もこれまでに、色情報を用いることなく抽出できる特徴点として表出点を提案し、物体の剛体運動の推定に利用した⁴⁾。物体形状の突出した部分に対応したボクセルが CSPs や表出点として抽出されやすい。

時系列視体積統合手法の問題は、運動推定誤差が大きいフレームが存在するときに獲得形状が大きく欠損することである。時系列視体積統合においては、統合形状が各フレームで得られる視体積の積領域として求められるために、1フレームでも運動推定誤差の大きなフレームが存在すると、統合形状が大きく欠損する。欠損の原因となる運動推定誤差は、運動推定を行う2フレーム間で対応点を持つ特徴点を得られないことにより生じる。このことは、特徴点の抽出に用いるシルエットに欠損や過抽出があるために起こる。特に、物体形状の特徴を示す突出部分が欠損するときには、獲得形状の品質を大きく損なうことになる。しかしながら、従来の CSPs や表出点の特徴点抽出法では、シルエットに欠損や過抽出があることを想定していない。

本研究では、時系列視体積統合において、統合形状の欠損を抑えることを目指し、物体特徴を示す突出部分が保存されるような視体積統合手法を提案する。

我々は、物体の突出部分を表わしている三次元特徴点に着目した。物体形状の突出部分は、物体形状を特徴付ける点であるといえる。CSPs や表出点は視体積内部に抽出されるボクセルであるが、物体形状の突出した部分に多く現れる性質がある。この性質を利用した従来研究として、CSPs のアイデアの元になっている Frontier point を使う手法が提案されている。この手法では、Frontier point が物体形状を特徴的に示す点であることから、Frontier point を三次元的に繋いでなめらかな物体表面を構成する^{5)~8)}。本研究ではこの性質を、物体形状の突出部分を保存するようなフレームを選択することに利用する。特徴点を保存するようなフレームがどのフレームかが判定できれば、そのフレームで得られる視体積のみを統合することで、統合形状にも各フレームで得られる突出部分が含まれるようにすることができる。そこで、そのフレームを統合形状の獲得に利用したときに、特徴点がどれほど保存されるかという評価関数を設計し、その評価値が高いフレームの視体積のみを統合する。

2ではまず、本研究を進める上で前提となる視体積交差法、時系列シルエット統合および表出点抽出について説明する。3では、表出点がどのくらい統合形状に保存されるかを評価する評価関数を設計し、この評価関数を元に利用するフレームを選択する手法について述べる。4ではシミュレーション実験および実環境実験を通して、提案する評価関数および評価統合方法の正当性を検証する。最後に、5でまとめと今後の課題について述べる。

2. 時系列シルエットからの三次元形状獲得

本節では、議論を進める上で必要な視体積交差法、時系列視体積統合および表出点抽出について説明する。

2.1 視体積交差法

視体積交差法¹⁾²⁾では、複数のカメラで物体を観測し、三次元形状を獲得する。獲得される形状は視体積(Visual Hull, VH)と呼ばれる。

視体積は物体の全周モデルを獲得できる利点がある。形状獲得手法には、他に Shape-from-Motion^{9)~11)}や、多視点ステレオ¹²⁾¹³⁾、レーザスキャンを用いたステレオ視¹⁴⁾¹⁵⁾などがあるが、いずれも表面モデルを獲得する手法であり、全周モデルを獲得する工夫が別途必要となる。そのため、近年では、視体積を“物体が存在しうる領域の上限を与える”領域として用い、上述のような視体積交差法以外の形状獲得手法と組み合わせ最終的な形状を獲得する研究が盛んに行われている^{16)~21)}。

図1に時刻 $i(i = 1, \dots, M)$ において視体積交差法が適用される様子を示す。カメラは同期撮影しているものとし、時刻 i で得られる画像を第 i フレームにおける画像と呼ぶことにする。第 j 番目のカメラ ($j = 1, \dots, N$) を C_j と表す。各カメラでは対象物体 O が観測され、物体投影領域が抽出される。物体投影領域はシルエットと呼ばれ、第 i フレームで C_j によって得られるシルエットを S_{ij} と表す。 S_{ij} からは物体が存在しうる錐体状の領域が計算できる。この錐体状の領域を視錐体 V_{ij} と呼ぶ。 C_j における射影行列を P_j とすれば、離散ボクセル空間における V_{ij} は以下の式(1)のようにボクセルの集合として表される。

$$V_{ij} = \{ \mathbf{v} \mid P_j \mathbf{v} \subset S_{ij} \} \quad (1)$$

ただし、 \mathbf{v} はボクセルである。第 i フレームの視錐体から計算できる物体が存在しうる領域は、視錐体 $V_{ij}(j=1, \dots, N)$ の積領域として計算できる。この積領域を視体積 V_i と呼ぶ。 V_i は式(2)で表すことができる。

$$V_i = \{ \mathbf{v} \mid \forall j, P_j \mathbf{v} \subset S_{ij} \} \quad (2)$$

さらに、対象物体が剛体で運動しているときに、その剛体運動を推定することができれば、すべてのフレームの視体積 $V_i(i = 1, \dots, N)$ の積領域から、対象物体の存在しうる領域を求めることができる³⁾⁴⁾。第

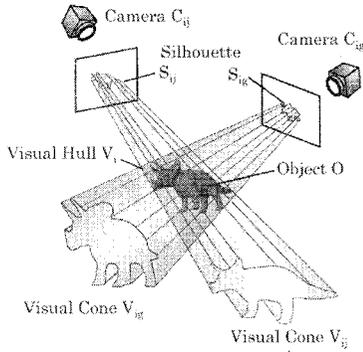


図1 第 i フレームにおける視体積獲得
Fig.1 Reconstructing a visual hull at i -th frame.

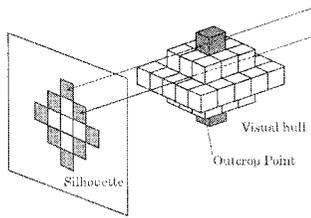


図2 表出点抽出
Fig.2 Extraction of outcrop points.

k フレームを基準として時系列シルエットを統合するものとしたとき、第 i フレーム ($i = 1, \dots, N$) と第 k フレームの間の物体の剛体運動 D_{ik} が推定できれば、時系列視体積から求められる物体が存在する領域 V^k は式 (3) で求められる。 V^k は統合形状と呼ばれる。

$$V^k = \{v \mid \forall i, D_{ik}v \subset V_i\} \quad (3)$$

2.2 表出点抽出

統合形状獲得のためには、物体の剛体運動を知る必要がある。表出点 (Outcrop Points, OPs)⁴⁾ は、複数カメラのシルエットから得られる三次元特徴点である。

シルエットが完全であれば、異なるフレームで得られる表出点間の運動から物体運動を推定し、時系列視体積を統合することができる。表出点は視体積に含まれるボクセルであり、1. 一つ以上のシルエット境界領域画素に投影され、2. 投影される画素には唯一そのボクセルが投影される、の2条件を満たすものであると定義されている。図2中の灰色のボクセルが表出点にあたる。

表出点は物体領域に含まれることが保証される。また、このようにして抽出される表出点は、形状の突出した部分に多く抽出される。図3(b)に示す点群は、図3(a)の物体を観測するときに得られる表出点である。表出点は、物体とカメラの位置関係が変化しても、安

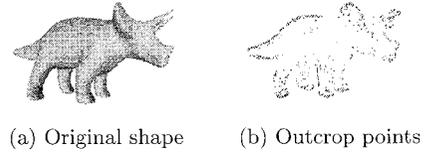


図3 表出点抽出例
Fig.3 An example of outcrop point extraction.

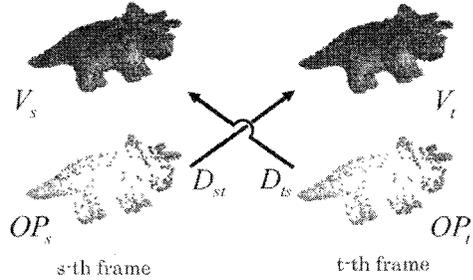


図4 視体積と表出点からの運動推定誤差評価
Fig.4 Motion estimation error evaluation from relationship between visual hulls and outcrop points.

定して抽出されやすいことが確かめられている。

ただし、上述の性質はシルエットが完全であるときに成り立つ性質である。本研究で扱うような、欠損および過抽出を含むようなシルエットから表出点を抽出する場合には、これらは完全には成り立たない。シルエットに欠損や過抽出があるときには、物体領域外のボクセルが表出点として抽出されたり、異なるフレームで対応する表出点が抽出されないことがある。

3. 表出点保存に基づく評価関数の設計と時系列視体積統合

第 s フレームと第 t フレームの間の剛体運動 D_{st} が正しく推定されており、かつ、それぞれのフレームでシルエットに欠損や過抽出がないものとするとき、図4に示すように、第 s フレームで得られる表出点群 OP_s は剛体運動 D_{ts} を作用させると視体積 V_t に含まれ、第 t フレームで得られる表出点群 OP_t は剛体運動 D_{st} を作用させると視体積 V_s に含まれる。

$$p_s \in OP_s, D_{st}p_s \in V_t \quad (4)$$

$$p_t \in OP_t, D_{ts}p_t \in V_s \quad (5)$$

式 (4), (5) が完全に成り立つときには、物体の特徴的な部分を示す表出点が視体積統合によって失われないことがいえる。しかし実際には、各フレームの視体積に欠損があったり、運動推定に誤差が含まれることによって、式 (4), (5) は完全には成り立たない。2フレームの視体積を統合するとき、表出点が保存

されるかどうかを評価するためには、式 (4), (5) を満たす表出点の割合を求めればよい。この割合を関数 $E_m(s, t)$ で表すとし、式 (6) で定義する。

$$E_m(s, t) = \frac{n_{st} + n_{ts}}{2} \quad (6)$$

$$n_{st} = \frac{n\{p_s | p_s \in OP_s, D_{st} p_s \in V_t\}}{n\{p_s | p_s \in OP_s\}}$$

$$n_{ts} = \frac{n\{p_t | p_t \in OP_t, D_{ts} p_t \in V_s\}}{n\{p_t | p_t \in OP_t\}}$$

ただし、 $n\{\cdot\}$ は集合に含まれるボクセルの個数であり、 OP_s, OP_t はそれぞれ s, t フレームで抽出される表出点の集合、 V_s, V_t はそれぞれ s, t フレームで構築される視体積に含まれるボクセルの集合である。評価関数 $E_m(s, t)$ は $[0, 1]$ の値を取り、値が大きいほど表出点が保存されるような視体積統合が行えることを表す。

$E_m(s, t)$ が小さな値をとるフレーム間では、統合形状の表出点が失われると考えられるので、閾値 E_m^{th} を設定し、 $E_m(s, t) < E_m^{th}$ であるようなフレームのシルエットは、視体積構築に利用しないこととする。具体的には、第 0 フレームの視体積をベースとして、 $E_m(0, i) < E_m^{th}$ であるようなフレームに対して、新たに第 i' ($i' = 1, \dots, M'$) フレームというラベルをつけ、ラベルのついたフレームの視体積のみを統合することで統合形状を得る。

4. 実験結果

実験では、シミュレーションデータと実環境データを処理対象として、表出点に基づく評価関数 $E_m(s, t)$ が機能し、統合形状に物体の特徴的な部分である突出部分が保存されるかどうかを検証する。

4.1 シミュレーション実験

恐竜モデルのシミュレーションデータに対して、カメラ 12 台で時系列のシルエットを抽出し、提案する時系列視体積統合手法を適用した。シルエット欠損率および過抽出率をそれぞれ A:1%, B:2%, C:5%, D:7.5% に設定した。恐竜モデルは、第 t フレームでは x, y, z 軸方向に $+0.5t$ ボクセル平行移動、 x, y, z 軸のまわりに $+t^\circ$ 回転している。

A, B, C, D のそれぞれについて、(a) は原形状、(b) はシルエット欠損率および過抽出率を設定したときに 1 フレームのみから得られる視体積、(c) は従来の時系列統合による 50 フレームのシルエットを統合した形状、(d) は $E_m^{th} = 0.97$ に設定したときの統合形状を図 5, 6, 7, 8 に示す。それぞれの形状は Marching cubes²²⁾ を用いて表面を構成し、表面にスムージング処理を施した。

それぞれの図の (b) の視体積では、得られる視体積が角張っていたり、物体領域から遠いところにボリュームが生成されたりして、余分な領域が多く含まれることがわかる。(c) の統合形状では、運動推定を誤るフ

レームの視体積を統合したことにより、獲得形状の一部が大きく欠損している。提案手法を用いて得られる (d) の統合形状では、原形状を包含するような形状を得ることができている。ただし、図 8(d) に示す D の提案手法による統合形状では、恐竜の角の部分が欠けている。この欠損は運動推定誤差によるものではなく、各フレームで得られる視体積の時点ですでに欠損が生じている (図 8(b)) ことが原因である。複数のフレームの視体積を統合することにより、欠損が顕著になっている。それでもなお、図 8(c) に示す通常の視体積統合手法結果に比べると、物体の突出部分の多くが保存されていることが確かめられる。A, B および C では、(b) の視体積に見られるような表面上の余分な領域もなく、(c) の統合形状に見られるような大きな欠損もない。それぞれの欠損率および過抽出率の設定の下で、(b) に示す 1 フレームでの視体積に対して、(d) の統合形状は突出部分を保存しながらも、原形状に近いめらかな形状を獲得できていくことがわかる。このことから、シルエットの欠損率および過抽出率が数% であるときには、提案手法が有効に作用し、突出部分を保存するような視体積統合ができていくことがわかる。

4.2 実環境実験評価

実環境下で 19 台のカメラにより恐竜を観測し、得られる時系列画像からシルエットを抽出した。提案手法により 50 フレームのシルエット統合を行った結果を図 9 に示す。シルエット抽出精度は、第 1 フレームの 19 枚のシルエットで、欠損率平均が 8.54%、過抽出率平均が 5.27% であった。閾値 $E_m^{th} = 0.95$ とした。図 9(c) に示す通常の時系列統合結果では、大きな欠損はないものの、恐竜モデルの足先の部分に欠損が見られる。図 9(d) に示す提案手法による獲得形状では、このような欠損は観察されず、物体形状の突出部分が保存されていることがわかる。

5. まとめ

本研究では、シルエットに欠損および過抽出がある場合にも、時系列視体積統合によって実形状に近い形状を獲得する手法を提案した。シルエット欠損および過抽出が表出点抽出に与える影響を明らかにし、物体の特徴的な部分である突出部分が保存されるかどうかの評価を与える関数を設計した。この関数の値を基に時系列視体積を統合することにより、物体形状の突出部分が統合形状に含まれるようになり、運動推定誤差によって統合形状が大きく欠損する問題を解決した。また、1 フレームで得られる視体積よりも誤差領域の少ない形状を獲得することができた。

今後の課題として、提案した評価関数に用いる閾値を設定環境、時系列統合の対象となるフレーム数などから自動的に与えることが考えられる。

参考文献

- 1) W.N. Martin, and J.K. Aggarwal, "Volumetric description of objects from multiple views," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.5, no.2, pp.150-158, 1983.
- 2) A.Laurentini, "How far 3d shapes can be understood from 2d silhouettes," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol.17, no.2, pp.188-195, 1995.
- 3) G.K.M. Cheung, S. Baker, and T. Kanade, "Shape-from-silhouette across time part i: Theory and algorithms," *International Journal of Computer Vision*, vol.62, no.3, pp.221-247, May 2005.
- 4) 豊浦正広, 飯山将晃, 角所考, 美濃導彦, "視体積交差法における時系列画像の統合による三次元復元形状の再現性の向上," *信学論*, vol.J88-D-II, no.8, pp.1549-1563, August 2005.
- 5) G. Cross, and A. Zisserman, "Surface reconstruction from multiple views using apparent contours and surface texture," *NATO Advanced Research Workshop on Confluence of Computer Vision and Computer Graphics*, Ljubljana, Slovenia, eds. A. Leonardis, F.Solina, and R.Bajcsy, pp.25-47, 2000.
- 6) J.S. Franco, and E. Boyer, "Exact polyhedral visual hulls," *Proceedings of the Fourteenth British Machine Vision Conference*, pp.329-338, September 2003, Norwich, UK.
- 7) S.Lazebnik, E. Boyer, and J.Ponce, "On computing exact visual hulls of solids bounded by smooth surfaces," *IEEE Conference on Computer Vision and Pattern Recognition*, pp.156-161, 2001.
- 8) M.Brand, K.Kang, and D.Cooper, "Algebraic solution for the visual hull," *IEEE Conference on Computer Vision and Pattern Recognition*, pp.30-35, 2004.
- 9) C. Tomasi, and T. Kanade, "Shape and motion from image streams under orthography - a factorization method," *International Journal of Computer Vision*, vol.9, no.2, pp.137-154, 1992.
- 10) 築澤宗太郎, 杉本晃宏, 松山隆司, "装着型視覚センサを用いた対象の3次元形状復元," *情報処理学会 CVIM 研究会資料*, vol.SIG-CVIM-133, pp.71-78, 2002.
- 11) S.Tsukizawa, K.Sumi, and T.Matsuyama, "3d digitization of a hand-held object with a wearable vision sensor," *International Workshop on Human-Computer Interaction (HCI)*, pp.129-141, May 2004.
- 12) S.Lazebnik, Y.Furukawa, and J.Ponce, "Projective visual hulls," *International Journal of Computer Vision*, vol.74, no.2, pp.137-165, August 2007.
- 13) C.Strecha, R.Fransens, and L.V. Gool, "Combined depth and outlier estimation in multi-view stereo," *IEEE Computer Vision and Pattern Recognition (CVPR)*, vol.2, pp.2394-2401, 2006.
- 14) S.Rusinkiewicz, O. Hall-Holt, and M. Levoy, "Real-time 3d model acquisition," *ACM Transactions on Graphics*, vol.21, no.3, pp.438-446, July 2002.
- 15) M.Levoy, K.Pulli, B.Curless, S.Rusinkiewicz, D.Koller, L.Pereira, M.Ginzton, S.Anderson, J.Davis, J.Ginsberg, J.Shade, and D.Fulk, "The digital michelangelo project: 3d scanning of large statues," *Proceedings of the ACM SIGGRAPH*, pp.131-144, July 2000.
- 16) S.Seitz, B.Curless, J.Diebel, D.Scharstein, and R.Szeliski, "A comparison and evaluation of multi-view stereo reconstruction algorithms," *IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, vol.1, pp.519-526, 2006.
- 17) J.Isidoro, and S.Sclaroff, "Stochastic refinement of the visual hull to satisfy photometric and silhouette consistency constraints," *IEEE International Conference on Computer Vision (ICCV)*, pp.1335-1342, 2003.
- 18) J. Starck, and A. Hilton, "Surface capture for performance-based animation," *IEEE Computer Graphics and Applications*, vol.27, pp.21-31, 2007.
- 19) Y.Furukawa, and J.Ponce, "Accurate, dense, and robust multi-view stereopsis," *IEEE Computer Vision and Pattern Recognition (CVPR)*, 2007.
- 20) C.H. Esteban, and F.Schmitt, "Silhouette and stereo fusion for 3d object modeling," *International Journal on Computer Vision and Image Understanding*, vol.96, no.3, pp.367-392, 2004.
- 21) 延原章平, 和田俊和, 松山隆司, "弾性メッシュモデルを用いた多視点画像からの高精度3次元形状復元," *情報処理学会 CVIM 研究会論文誌*, vol.43, no.SIG11(CVIM5), pp.53-63, December 2002.
- 22) W.E. Lorensen, and H.E. Cline, "Marching cubes: A high resolution 3d surface construction algorithm," *Proceeding of the ACM SIGGRAPH*, vol.21, no.4, pp.163-169, 1987.

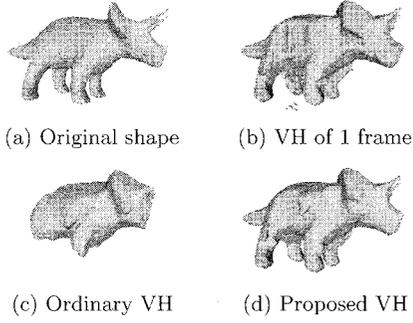


図 5 恐竜模型に対する獲得形状 (A : (missing rate, additional rate)=(1%, 1%))
 Fig. 5 Reconstructed shape for a toy of triceratops. (A : (missing rate, additional rate)=(1%, 1%))

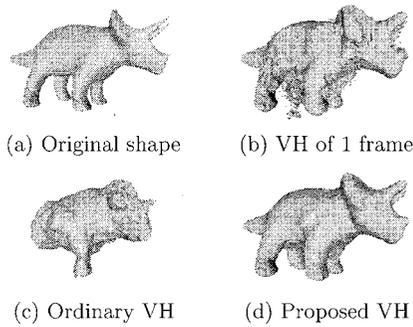


図 6 恐竜模型に対する獲得形状 (B : (missing rate, additional rate)=(2%, 2%))
 Fig. 6 Reconstructed shape for a toy of triceratops. (B : (missing rate, additional rate)=(2%, 2%))

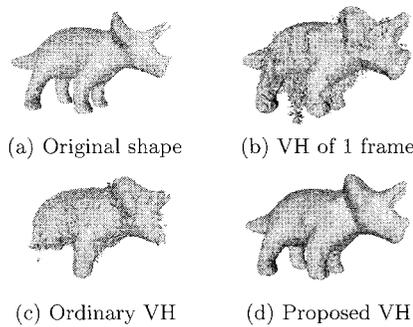


図 7 恐竜模型に対する獲得形状 (C : (missing rate, additional rate)=(5%, 5%))
 Fig. 7 Reconstructed shape for a toy of triceratops. (C : (missing rate, additional rate)=(5%, 5%))

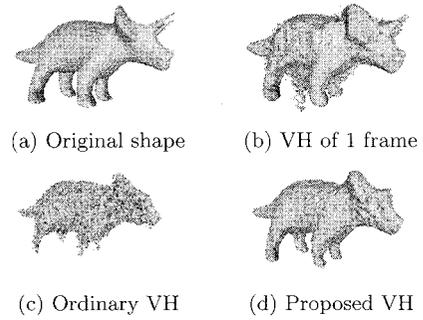


図 8 恐竜模型に対する獲得形状 (D : (missing rate, additional rate)=(7.5%, 7.5%))
 Fig. 8 Reconstructed shape for a toy of triceratops. (D : (missing rate, additional rate)=(7.5%, 7.5%))

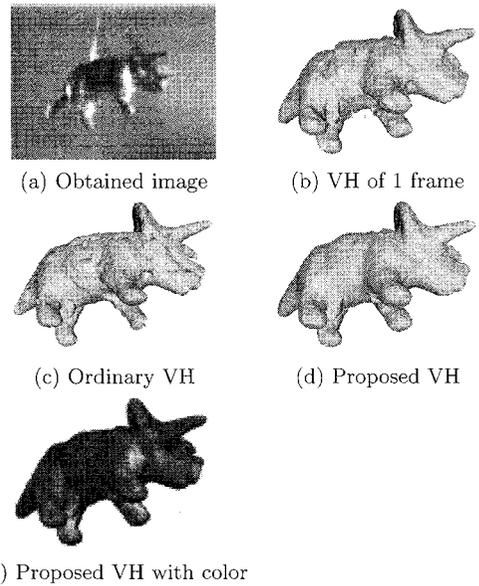


図 9 恐竜模型に対する獲得形状
 Fig. 9 Reconstructed shape for a toy of triceratops.