

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

豊浦 正広[†] 飯山 将晃^{††} 角所 考^{††} 美濃 導彦^{††}

An Accurate Shape Reconstruction Method by Integrating Visual Hulls
in Time Sequences

Masahiro TOYOURA[†], Masaaki IIYAMA^{††}, Koh KAKUSHO^{††}, and Michihiko MINOH^{††}

あらまし 視体積交差法では、対象物体を観測するカメラ台数が多ければ多いほど復元形状の再現性が向上するが、カメラを配置できる空間の制約によりカメラ台数を無制限に増やすことは難しい。これに対して、もし対象物体の剛体運動を各時刻の視体積から推定することができれば、複数の時刻でのカメラ画像を統合して、一つの時刻で得る視体積よりもより再現性の高い視体積を得ることができる。そこで我々は、この考えに基づき、運動する剛体物体を観測することで、時系列で得られる対象物体の画像を統合し、多くのカメラを用いたときと同等の復元形状を少ないカメラ台数で得ることを目指す。この際に対象物体の運動を推定する必要があるが、これは特徴点を各時刻の視体積から抽出することで行う。特徴点抽出で問題となるのは、視体積が物体が占める領域以外に余分な領域を含むことである。このために、物体表面に現れる特徴的な部分の特徴点として抽出しようとする、その特徴点が余分な領域に覆われて視体積表面上に現れないことが起こる。つまり、各時刻の視体積から時系列で安定して抽出される特徴点を得ることが難しいのである。我々は各時刻ごとのカメラ画像から視体積を求め、更に視体積をシルエットに投影し、余分な領域に覆われない点を抽出する手法を提案する。このような点の特徴点として用い、精度の良い運動推定を行うことができた。シミュレーション実験において、多くのカメラを用いたときと同等の復元形状を得られることを確認し、実画像を用いた実験でも復元形状の再現性の向上を確認した。

キーワード 視体積交差法, 特徴点抽出, 運動推定, 時系列データ, 標準化誤差

1. ま え が き

多種多様な物体の三次元形状復元の手法として視体積交差法 [1], [2] がある。視体積交差法は、複数のカメラで対象物体を観測して得られるシルエットから対象物体の形状を復元する手法である。視体積交差法は、博物学のように多様な物体を対象に形状復元を行う必要がある場合の手法として大変優れている。これは、視体積交差法が物体表面の色変化に乏しい物体やレーザを吸収する物体など、ステレオ視やレーザレン

ジファインダを用いた形状復元の適用が困難な物体に対しても形状復元が可能のためである。

視体積交差法では、原理的にはカメラ台数が増えれば増えるほど、より物体の最小凸包に近い復元形状を得ることができる。近年、カメラは低価格になり、大量カメラ環境の構築が容易に可能となったものの、カメラを配置できる空間の制約などからカメラ台数を無制限に増やすことは難しい。そこで、本研究では数十台程度の現実的に準備できる台数のカメラを用いた環境で、運動中の物体に対して時系列で撮影された画像を統合することで、カメラ台数を増加させたときと等価な視体積を得ることを目指す。

対象物体の運動が剛体運動である場合、物体が運動することにより、カメラと物体の相対的な位置関係が変化し、各フレームで撮影された画像を新たな視点位置からの画像とみなすことができる。これを“時系列

[†] 京都大学大学院情報学研究所, 京都市

Graduate School of Informatics, Kyoto University, Yoshidahonmachi, Sakyo-ku, Kyoto-shi, 606-8501 Japan

^{††} 京都大学学術情報メディアセンター, 京都市

Academic Center for Computing and Media Studies, Kyoto University, Yoshidahonmachi, Sakyo-ku, Kyoto-shi, 606-8501 Japan

画像の統合”と呼ぶ。時系列画像の統合により、もともとのカメラ台数に基づく復元形状よりも最小凸包に近い復元形状を得られることが期待できる。視体積交差法では、もし無限台数のカメラを用いて視体積を構築できた場合には、復元形状として物体の最小凸包が得られる。そこで、本論文では復元形状が物体の最小凸包に近づくことを再現性の向上と呼び、時系列画像の統合によってこれを実現することを目的とする。

なお、視体積交差法では物体の凹面を復元することは原理上不可能であるが、このような視体積交差法における凹面の復元問題は本稿では議論の対象としない。実際、この問題に対しては、既に Space carving [3] の考え方に基づいて、表面の色情報を用いて視体積交差法で得られた復元形状をもとに凹面も復元する手法 [4] が提案されている。このような Space carving においても、カメラ台数は多ければ多いほど復元される形状は再現性が高いという性質があるため、本研究はこれらの手法の性能の向上にも有効であると考えられる。

本研究と同様、カメラと物体の相対位置の変化により、再現性の高い形状を復元する手法としては、ターンテーブルを用いる手法 [5] ~ [7] がある。このような手法では、物体の運動方向がターンテーブルの回転軸周りの回転のみに限定され、回転軸の周り以外の視点方向からの情報を得ることが難しい。また、ターンテーブルに接する部分の形状復元も難しいという問題がある。

これに対し、Cheung ら [8] は画像上で物体の特徴点を追跡し、物体の運動を推定することによってカメラ台数を増加させた場合と同等の視体積を得る手法を提案している。また、築澤ら [9] は時系列画像統合により、手持ち物体の視体積構築を可能にしている。しかしながら、これらの手法では特徴点の追跡に物体の色情報を用いており、物体表面の色変化に乏しい物体への適用が難しいことから、表面の色変化に乏しい物体であってもシルエットさえ得られれば形状復元が可能であるという視体積交差法の本来の利点の一つが失われるという問題がある。

一方、画像のシルエットから得られる情報のみから、エピポーラ拘束条件を使って剛体運動推定を行う手法 [10] ~ [14] も提案されている。これらの手法では特徴点を対応付けして三次元の運動推定を行うが、特徴点はある二次元画像のシルエット境界領域に投影される点のうち、他の画像でもシルエット境界領域に投影される点を用いる。この場合には、特徴点が多く抽出

されすぎて運動推定がうまくいかない現象がしばしば発生する。また、画像のシルエットにノイズが含まれている場合にも運動推定に誤りが起こりやすい。これらの手法の問題点については、3.2 で詳しく述べる。

以上のような従来研究に対して、本研究では色情報を用いず、シルエットのみを利用し、かつ、そこから得られる各時刻の視体積の形状から三次元の特徴点を抽出して、運動推定を行う。

一般に各時刻の視体積から時系列で安定した特徴点を抽出することは難しい。なぜなら、視体積は物体が占める領域以外に余分な領域を含むため、物体表面に現れる特徴的な部分を特徴点として抽出しようとすると、観測時刻によってはその特徴点が余分な領域に覆われて、視体積表面上に現れないことがしばしば生じるためである。

この問題に対し、本論文では視体積をシルエットへ再投影することにより、各時刻の視体積から余分な領域に覆われない物体本来の表面上の点を抽出する手法を提案する。

2. ではまず、本研究の背景である視体積交差法について説明し、物体の運動を知ることによって仮想的なカメラ台数の増加を実現するアプローチの原理について説明する。3. では各時刻の視体積上で特徴点を求める手法を提案し、4. では特徴点からの物体の運動推定手法について説明する。5. では実験結果によって本手法による形状復元の再現性の向上の程度を示し、6. で今後の課題について議論する。

2. 復元形状の再現性の向上

本章では視体積交差法の原理を説明し、物体の運動による復元形状の再現性向上のためのアプローチについて述べる。

2.1 視体積交差法

視体積交差法では、複数台のカメラで対象となる物体を観測する。物体はすべてのカメラの観測範囲内に収まるものとする。以下ではカメラ n 台の場合の視体積交差法の原理を説明する。

図 1 に示すように、 m 番目のカメラ C_m ($m = 1, \dots, n$) の観測範囲に物体 O が存在するとき、物体 O は C_m の画像上に投影され、その投影像は二次元画像上のシルエット R_m として得られる。

このことから、逆にカメラ C_m における物体 O のシルエット R_m が得られたとき、 O は C_m のレンズ中心を端点とし、 R_m 上の任意の点を通る半直線の集

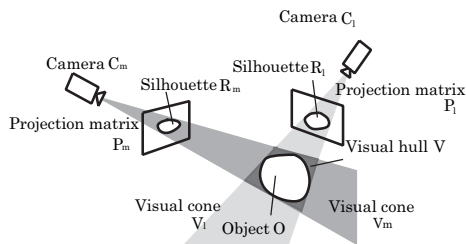


図 1 視体積交差法

Fig.1 Volume intersection method

合からなる錐体状の開空間に内接して存在することが分かる．この開空間をカメラ C_m における“視錐体”と呼び、 V_m で表す．これは C_m の画像から得られる物体 O の 3 次元形状に関する情報に相当する．このとき視錐体 V_1, V_2, \dots, V_n について、式 (1) のような積領域 V を考え、“視体積”と呼ぶ．

$$V = \bigcap_{m=1}^n V_m \quad (1)$$

この V を求めることで各カメラから得られる物体 O の 3 次元形状に関する情報を統合することができる．

視体積のもつ特徴として次の二つが挙げられる．

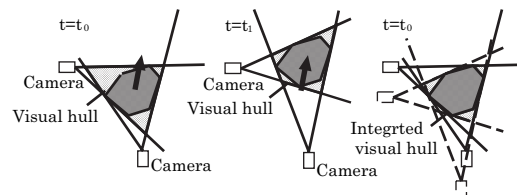
- (1) 対象物体は視体積に常に内接する
- (2) 視点が増加すればするほど視体積の体積は減少する

これらの特徴より、カメラ台数が増えれば多いほど視体積に含まれる物体領域以外の領域が減少し、視体積は物体形状に近づくことと結論づけられる．つまり、復元形状の再現性の向上を図るためにはカメラ台数を増やせばよい．また、このときの視体積に含まれる物体領域以外の領域を“残余領域”と呼ぶことにする．

本研究では、視体積 V の表現にボクセルモデルを用いる．ボクセルモデルでは、三次元空間中の各微小領域を 2 値ボクセルで表現し、視体積 V に含まれるボクセルの値を 1、そうでないものの値を 0 とする．

2.2 物体運動を利用した再現性の向上

前節でカメラ台数が増えれば増えるほど、復元形状の再現性の向上を図ることができることを述べたが、実際にはカメラを配置できる空間の制約によりカメラ台数を無制限に増やすことは難しい．そこで剛体を対象として複数の時刻での物体の姿勢と視体積から各時刻間での物体の運動を推定して視体積を統合することによって、カメラ台数を仮想的に増加させることを考



(a) VH before movement (b) VH after movement (c) Integrated VH

図 2 視体積の統合

Fig.2 Integration of visual hulls (VHs).

える．ここで運動推定は、物体の運動が小さな近接するフレーム間で行うという仮定を設けておく．

例として、2 台のカメラを設置し運動する物体を観察するとき、もとの位置での視体積と運動後の視体積が図 2 のように得られた場合を考える．

移動前の位置での視体積 (図 2(a)) は、運動後の位置にある物体 (図 2(b)) に対して、カメラを物体の運動分だけ差し引いた位置に設置したときに得られる視体積とみなすことができる．このようにして、運動後の視体積に運動前の位置の視体積の情報を統合することによって仮想的に、4 台のカメラによって得られる視体積 (図 2(c)) と同等のものが得られ、復元形状の再現性が向上する．

2.3 標本化誤差による再現性向上の限界

入力データである画像が空間的に離散的なデータであるために、標本化誤差に起因する再現性向上の限界が存在することがある．これは視体積交差法においてカメラ台数の増加による復元形状の再現性の向上を目指す場合に問題となる．したがって、時系列の視体積の統合に基づく仮想的なカメラ台数の増加によって復元形状の再現性の向上を図る場合には、標本化誤差による再現性向上の限界を明らかにしておく必要がある．この標本化による復元形状への影響に関しては、Niem [15] がまとめているが、誤差の最大値が大きく見積もられており実用的ではない．そこで本節ではまず、標本化誤差が与える復元形状の再現性の向上の限界を明らかにする．

まず、観測領域に含まれるすべてのボクセルのすべての辺を、すべてのカメラ画像上に投影したときに得られる投影幅を考え、その中の最大値が q ピクセル幅であるとする． q の値は、カメラと観測領域の位置関係、カメラの解像度、及び、ボクセルの大きさにより変化する． q が大きくなるほど、視体積に与えられる標本化誤差は小さくなる．以降では、 q を用いて標

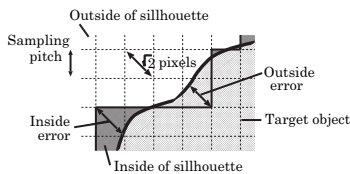


図3 標本化誤差による視体積表面上の誤差
Fig. 3 Error on surface of visual hull by sampling.

本化に起因する再現性の向上の限界について議論を進める。

カメラ画像中の一つの画素が物体領域に含まれるか、背景領域に含まれるかは画像中の各画素の値により決定される。このときに、ある画素に含まれるシルエットに注目し、その画素に収まる真のシルエット上の各点と画素を代表する座標との距離を考える。これが最大で k ピクセル幅であるとすると、視体積表面には最大で k/q ボクセル幅の形状誤差が生じる。

ここで標本化誤差の最大ピクセル幅 k について考える。ある物体が与えられるときのシルエットの境界は図3に示すように、各ピクセルの内部に収まる。

このとき、画像上において、標本化誤差によって間違っ物体領域にあると判定される背景領域の点は最大で $\sqrt{2}$ ピクセル幅だけ実際の物体領域の境界線から離れた位置に存在することになる。逆に観測空間中でも、標本化誤差によって間違っ背景領域と判定される物体領域の点は、最大で $\sqrt{2}/q$ ボクセル幅だけ実際の物体領域の境界線から離れた位置に存在することになる。つまり、画像の標本化誤差は物体領域の境界線から内外に最大で $\sqrt{2}$ ピクセル幅だけ生じ、これにより視体積表面には最大で $\sqrt{2}/q$ ボクセル幅の誤差が生じると考えることができる。これが画像の標本化に起因する再現性の向上の限界である。

実際には、画像の標本化によりこれほど大きな誤差が生じることは考えにくく、各カメラからの物体のシルエットを得る際に、物体領域と背景の輝度値が大きく変わらず、物体領域が画素の半分以上の面積を占めるときに物体領域、そうでないときに背景領域であると判定されるとすれば、画像の標本化誤差は最大で $\sqrt{2}/2$ ピクセル幅だけ生じることになる。

仮に対象物体が球体であるとし、その半径が r ボクセル幅であり $r \gg k/q$ であるとすると、表面に生じる誤差領域の体積の視体積全体に対する割合 α は以下の式で表される。

表1 各対象物体の変数値と α に達するときのカメラ台数
Table 1 Parameters of target objects.

	q	N_v	α	Num. of cameras
Sphere	1.679	1265791	1.88%	49
Cube	1.679	2424294	1.52%	406
Torus	1.679	268545	3.16%	103
Triceratops	1.679	438310	2.68%	242

$$\alpha = \frac{\frac{k}{q} \cdot 4\pi r^2}{\frac{4}{3}\pi r^3} = \frac{3k}{qr} \quad (2)$$

この式によれば、例えば $r = 50$ ボクセル幅、 $q = 0.75$ 、 $k = \sqrt{2}/2$ ピクセル幅とするときには、 $\alpha = 5.67\%$ が標本化に起因して生じる誤差であると計算できる。一般的に対象物体は球体ではないため、視体積の表面積は対象物体が球体であるときよりも大きく、 α の実際の値は式(2)よりも大きくなる。物体領域に含まれるボクセル数が N_v であるとき、半径 $(\frac{3N_v}{4\pi})^{\frac{1}{3}}$ ボクセル幅の球と同数のボクセルを含む。よって、視体積表面には $\alpha = \frac{3k}{q} \left(\frac{3N_v}{4\pi}\right)^{-\frac{1}{3}}$ 以上の割合で誤差が生じると計算できる。

この性質に基づき、再現性の向上が限界に達するカメラ台数をシミュレーション実験によって求めた。シミュレーションデータを用いて、カメラ位置をランダムに与え台数を増やしたときの、視体積に含まれる残余領域のボクセル数の変化を調べた。シミュレーション実験では、球のほかに立方体、トーラス環、5.1で用いる恐竜の形状データを用いた。それぞれの物体の q 、 N_v 、及び、これらから計算される α の値を表1に示した。

カメラ台数による残余領域のボクセル数の変化を図4に示す。ただし、縦軸は α を基準として表現されている。

球において残余領域に含まれるボクセルの数が視体積に含まれるボクセルの数に対して α 以下となるのは、カメラ台数が49台以上のときであった。球以外の物体についても、残余領域が視体積の α 以下となるには、球のとき以上のカメラ台数が必要であった。表面に突出する部分があるような物体では、より多くのカメラ台数が必要となることも確認できる。

ここまでの議論により、復元される視体積表面に生じる、全体の α 程度の体積を持つ誤差領域は、標本化誤差によるものであり、時系列情報の統合により解決されるものではない。このことから、時系列情報の統

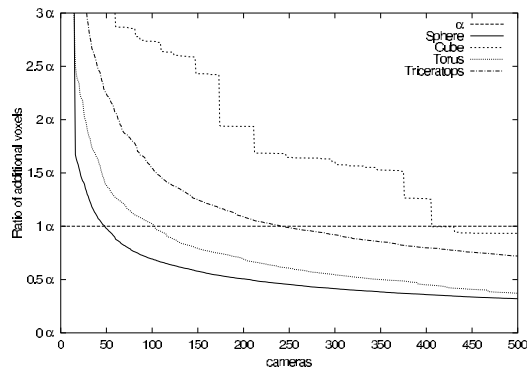


図 4 カメラ台数と残余領域の推移
Fig.4 Relation between camera number and extra region

合によりカメラ台数を仮想的に増やして復元形状の再現性の向上を目指す場合には、正しい形状から欠損する領域及び残余領域をとともに α 以下に抑えることを目標とする。

また、標準化誤差以外に再現性の向上の限界を与えるものとして、物体観測時に生じるカメラのノイズが挙げられる。視体積交差法では、カメラのノイズによりシルエットに欠損が生じると視錐体が欠損する。視体積 V はそれぞれのカメラで得られるシルエットから計算される視錐体の積領域で表現されるために、1台のカメラのシルエット欠損による影響は、視錐体を統合したものである視体積にまで及ぶ。このために、数十台もの多数のカメラを使って視体積交差法を行う際には、 n 台すべての視錐体に含まれていなくても $(n - N_{allow})$ 台のカメラの視錐体に含まれていれば視体積の要素に含めるという方法がよく用いられる [16]。この手法では例えば、 $N_{allow} = 2$ の場合、2.1 の式 (1) の代わりに以下の式を利用することになる。

$$V = \bigcup_{1 \leq m_1 < m_2 \leq n} \left(\left(\bigcap_{m=1}^n V_m \right) \cup V_{m_1} \cup V_{m_2} \right) \quad (3)$$

3. 特徴点抽出

本章では、対象物体の運動推定を行うための視体積からの特徴点抽出手法について述べる。

3.1 特徴点に要求される性質

提案手法では視体積はボクセルで表現されるため、視体積から抽出される特徴点もボクセルで表現される

ことに注意する。

ここで、運動追跡に用いる特徴点に求められる性質を考える。これには以下の二つが挙げられる。

- (1) 残余領域に覆われていない物体領域上の点である
- (2) 運動推定を行うフレーム間で対応するボクセルが得られやすい

このうちまず、(1) の性質をもたない点とは視体積の残余領域に含まれるボクセルである。このようなボクセルは、カメラと物体の位置関係が変化すると、視体積には含まれないことがあり、運動推定のための特徴点として適切でない。

しかし、残余領域に覆われていない物体領域上のボクセルを視体積のみに基づいて判別・抽出することは容易ではない。あるフレームで抽出された残余領域に覆われずに視体積表面上に現れていたボクセルが、次のフレームでは残余領域に覆われて視体積表面上に現れないことがあるためである。

また、(2) の性質を持つ特徴点を視体積形状のみから求めることは困難である。各フレームでの視体積は、それぞれ異なる形状の残余領域をもつためである。ただし、運動推定を行うフレーム間において、少数の特徴点に対応するボクセルを持たないときには、4. で示すロバスト推定を行うことにより運動推定が可能である。

そこでまず、各フレームの視体積に含まれる残余領域に含まれないような (1) の性質をもつボクセルを、視体積と画像の両方を利用して抽出することを考える。

3.2 従来的特徴点抽出手法との比較

物体のシルエット情報から抽出できる物体表面上の特徴点として、従来から frontier points (epipolar tangencies) と呼ばれる点が提案されている [11] [12] [13] [14]。この点は、2 台のカメラに対してエピポーラ面を考え、シルエットから得られる視体積領域とこのエピポーラ面が接するとき、その接点を残余領域に含まれていない物体領域に含まれると考えて、frontier points (epipolar tangencies) と呼んで、運動追跡や形状のスモーキングのための特徴点として採用している。図 5 は frontier points の抽出の様子である。

しかし、これらの点は物体領域に含まれないことがある。エピポーラ面上に複数の視体積領域の点が存在するときがこれにあたる。

図 6 に示すように、物体に対して 2 枚の画像が与えられると、白抜きの丸で示した点が frontier points

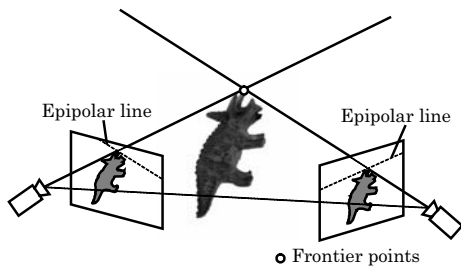


図 5 Frontier points
Fig.5 Frontier points

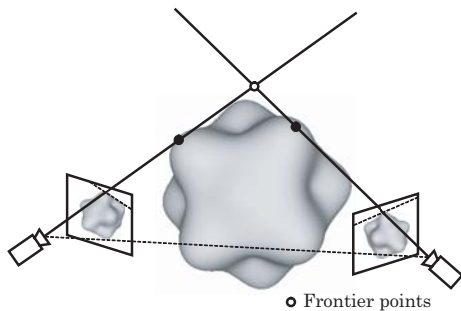


図 6 Frontier points が物体領域と一致しない場合
Fig.6 Missing points of frontier points

として抽出されてしまう。しかしこの点は物体領域に含まれない点である。2台のカメラそれぞれでのシルエット境界領域画素に対する視線が、物体領域以外で交差する場合にこの状況が起こる。もし物体領域に含まれない frontier points が視体積に含まれれば、特徴点に求められる性質 (1) を満たさない特徴点が見られる。対象物体が複雑になればなるほど、この問題が多く起こることが予想できる。そこで以下に述べる提案手法では、同じ視線に存在する点の数を考慮に入れ、複数の点在同一の視線に存在するときに、それらの点を特徴点として抽出しないことで、この問題を解決する。

3.3 再投影による特徴点抽出手法

各フレームでの画像からそのフレームでの視体積を構築した上で、図 7 のように視体積表面上のボクセルを画像へ再投影する。このときに、以下の二つの条件を満たすボクセル v を特徴点として採用し、表出ボクセルと呼ぶことにする。

- (a) v は一つ以上の画像のシルエットの境界領域画素に投影される
- (b) v が投影される画素には v 以外の視体積表面上

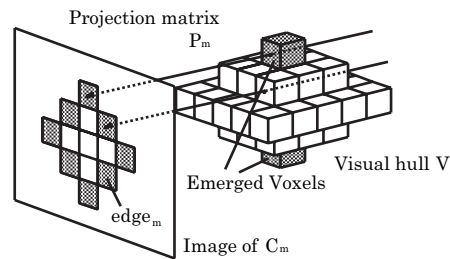


図 7 境界要素への投影
Fig.7 Projection to silhouette edge.

のボクセルが投影されない

表出ボクセルの抽出の様子を図 7 に示す。塗りつぶされたボクセルが表出ボクセルである。ただし、図 7 において、 P_m はカメラ C_m の射影行列、 $edge_m$ は C_m のシルエットの境界領域、 V は視体積である。

このようにして得られた表出ボクセルは、物体領域に含まれるボクセルであることが視体積交差法の原理により保証される。なぜなら、視体積交差法の原理からシルエットは視体積の投影像に一致することが保証されるためである。もし、表出ボクセルが物体領域に含まれないとすれば、(b) の条件から図 7 で示す視体積の投影像はシルエットに一致しないことになり、この保証が成り立たなくなる。

また、表出ボクセルは物体の運動が小さいとき、運動する前後で対応するボクセルが特徴点として抽出されやすい。つまり、3.1 で述べた特徴点に求められる性質であった、フレーム間で対応するボクセルが得られやすいことがいえる。このことは、物体形状で突出しているような部分が表出ボクセルとして抽出されやすいことから説明できる。図 8 にこれを示す。突出した形状をもつ物体の表出ボクセルを求めた図 8(a), (c) ではカメラと物体の相対的な位置が変化しても、その位置変化が小さければ、丸で示した点を含むボクセルが表出ボクセルとして抽出され続ける。一方、突出した形状を持たない物体の表出ボクセルを求める図 8(b), (d) ではカメラと物体の相対的な位置変化によって、その位置変化が小さくとも、丸で示した点を含むボクセルが表出ボクセルが抽出されないことがある。これは他のボクセルからの自己隠べいの起こりやすさに起因するものであり、物体上の突出した部分ほど表出ボクセルとして抽出されやすい。

結論として、物体の運動が小さいという仮定のもとで、表出ボクセルは前節に述べた特徴点に求められる

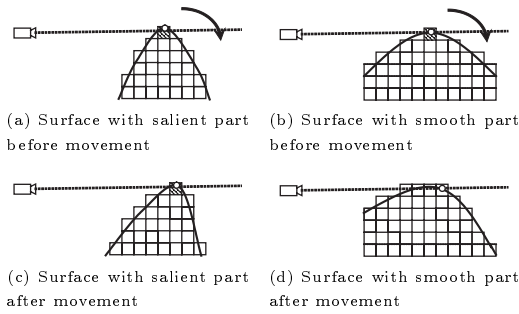


図 8 相対位置の変化による特徴点の変化

Fig. 8 Changing extracted points by related position among the object and cameras.

二つの性質をもつといえる。

しかし、すべての表出ボクセルが、突出した部分に対応するボクセルとはならない。カメラ配置によっては突出していない部分も表出ボクセルとして抽出され得るからである。そこで運動成分推定の際に、外れ値を扱うことのできるロバスト推定を行い、突出した部分を表現する表出ボクセルのみを推定に用いる。この手法については 4. で説明する。

また、突出部分をもたないような物体を観測するとき、運動推定に有効な表出ボクセルが得られず、フレーム間で対応するボクセルが得られない場合がある。どの程度の突出があれば適切な表出ボクセルが得られ、これを用いて運動推定を行うことができるかについては 5.2 の実験結果をもって考察を行う。

4. 表出ボクセルを用いた運動推定

特徴点を用いて剛体運動を推定するとき、この運動を表す剛体変換行列を用いる。運動推定はこの剛体変換行列を収束法を適用して求めることで行う。

4.1 剛体変換行列

剛体変換行列 $D^{t_i t_j}$ は、時刻 t_i と時刻 t_j の間の物体の運動変化を示す斉次座標行列で、 3×3 の回転移動行列 $R^{t_i t_j}$ と平行移動成分ベクトル $t^{t_i t_j}$ を用いて、次のように表すことができる。

$$D^{t_i t_j} = \begin{pmatrix} R^{t_i t_j} & t^{t_i t_j} \\ \mathbf{0}_{1 \times 3} & 1 \end{pmatrix} \quad (4)$$

ただし、 $\mathbf{0}_{1 \times 3}$ は 1×3 の零ベクトルである。剛体変換行列は回転、平行移動の変数を決定することで得られる。また、 $D^{t_i t_j} = (D^{t_j t_i})^{-1}$ が常に成り立つ。

ここで、回転行列には 4 元数表現を用いる。これに

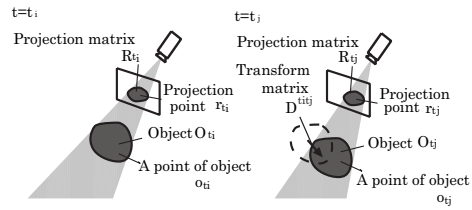


図 9 記号の定義

Fig. 9 Definition of symbols.

より、線形性が高まり、収束法を適用する際に局所解に陥りにくくなる。

回転行列 $R^{t_i t_j}$ は、4 元数 $\mathbf{q} = [\lambda_0, \lambda_1, \lambda_2, \lambda_3]^T$ を用いて表すと、以下のようになる。

$$\begin{pmatrix} \lambda_0^2 + \lambda_1^2 - \lambda_2^2 - \lambda_3^2 & 2(\lambda_1 \lambda_2 - \lambda_0 \lambda_3) & 2(\lambda_1 \lambda_3 + \lambda_0 \lambda_2) \\ 2(\lambda_1 \lambda_2 + \lambda_0 \lambda_3) & \lambda_0^2 - \lambda_1^2 + \lambda_2^2 - \lambda_3^2 & 2(\lambda_2 \lambda_3 - \lambda_0 \lambda_1) \\ 2(\lambda_1 \lambda_3 - \lambda_0 \lambda_2) & 2(\lambda_2 \lambda_3 + \lambda_0 \lambda_1) & \lambda_0^2 - \lambda_1^2 - \lambda_2^2 + \lambda_3^2 \end{pmatrix}$$

剛体変換行列を用いて異なる時刻の視体積を統合する手法を説明する。まず、物体の画像への投影を考える。図 9 に示すように記号を定義する。

まず、時刻 t_i で物体 O が占める領域を O_{t_i} とし、 o_{t_i} を O_{t_i} の要素とする。また、 P_n はカメラ C_n の射影行列であり、時間によって不変であるとする。更に、 r_{t_i} は o_{t_i} をカメラ n のスクリーンに投影した点であり、 r_{t_i} がなす投影領域を R_{t_i} とする。

このときに、 o_{t_j} を $(D^{t_i t_j})^{-1}$ により移動した点を P_n で画像に投影すると、これは R_{t_i} の要素となる。

これを式で表すと、

$$\begin{aligned} o_{t_j} \in O_{t_j} \text{ ならば、} \\ r_{t_i} = P_n((D^{t_i t_j})^{-1} \cdot o_{t_j}) \in R_{t_i} \end{aligned}$$

となる。同様にして、

$$\begin{aligned} o_{t_i} \in O_{t_i} \text{ ならば、} \quad r_{t_j} &= P_n((D^{t_j t_i})^{-1} \cdot o_{t_i}) \\ &= P_n(D^{t_i t_j} \cdot o_{t_i}) \in R_{t_j} \end{aligned}$$

であるといえる。この式は、 $D^{t_i t_j}$ と R_{t_j} が与えられれば O_{t_i} の要素である可能性をもつ o_{t_i} の集合が求められることを示す。この集合は時刻 t_i において、実際とは異なる視点配置で求めた視体積に等しい。つまり、時刻 t_i で求めた視体積と $D^{t_i t_j}$ と R_{t_j} から求めた視体積の積領域を利用することにより、時刻 t_i での視体積と時刻 t_j の視体積を統合することができる。

4.2 剛体変換行列の推定

誤差関数を最小にするような剛体変換行列 $D^{t_i t_j}$ を

Powell の収束法 [17] を用いて求める．Powell の収束法はその他の収束法に比べ、多くの変数を持つ関数に対して収束が速く、局所解に陥りにくいという性質をもつ．また、ロバスト推定を用いて、フレーム間で対応する点をもたない特徴点の推定結果への影響を小さくする．

時刻 t_i で求めた表出ボクセルは、理論的には剛体変換行列 $D^{t_i t_j}$ により移動させると、時刻 t_j で求めた表出ボクセルに一致する．しかし実際には、標準化誤差や表出ボクセルのうち物体の突出した部分に相当しないボクセルが抽出されることがあり、これらは完全には一致しない．そこで誤差関数 E を定め、これを最小化する剛体変換行列 $D^{t_i t_j}$ を Powell の収束法を用いて求める．時刻 t_i で求めた表出ボクセル $p_s^{t_i}$ を剛体変換行列 $D^{t_i t_j}$ により移動させたとき、時刻 t_j における表出ボクセルのうちで最も近くにある表出ボクセルを $p_s^{t_j}$ に対応する点と考え、この点との 2 乗距離を求める．誤差関数はこの 2 乗距離が M_{th} 以下であるものの平均をとった値に相当する．誤差関数 E を以下のように定める．

$$E = \sum_u f \left(\min_s (p_u^{t_j} - D^{t_i t_j} p_s^{t_i})^2 \right) \quad (5)$$

$$\text{ただし、} f(x) = \begin{cases} x & x \leq M_{th} \\ M_{th} & x > M_{th} \end{cases}$$

ここで、 $p_s^{t_i}$ 、 $p_u^{t_j}$ はそれぞれ時刻 t_i 、 t_j での表出ボクセルの座標を表し、 M_{th} は時刻 t_i 、 t_j の間の移動距離の最大量を定めたしきい値である． f のような外れ値の影響を小さくする関数を用いて収束法を適用するロバスト推定は、大きく離れた対応点をもつような表出ボクセルが収束に与える影響を小さくしている．

隣接するフレーム間で物体の運動量が微小であると仮定し、剛体変換行列をまず、隣接するフレーム間で求める．任意のフレーム間での剛体変換行列 $D^{t_i t_j}$ は以下のように求める．

$$D^{t_i t_j} = D^{t_j-1 t_j} D^{t_j-1 t_j-2} \dots D^{t_i+1 t_i+2} D^{t_i t_i+1}$$

5. 実験

はじめに、恐竜の形状を用いたシミュレーション実験によって手法の精度を評価した．次に、提案手法が有効に働くために必要な突出の程度を調べるために球

を基準として表面に正弦波の突出部を持つ物体を用いた実験を行い、手法が適用できる範囲について考察した．更に、現実に存在する物体の中で突出の程度が少ないと考えられる複数の物体に対して手法を適用して、提案手法の適用範囲を確かめた．最後に、実環境において実験を行った．

5.1 恐竜形状によるシミュレーション実験

まず、提案手法の精度を評価するために、シミュレーションデータを用いた実験を行った．三次元データには、形状が既知である恐竜の形状を用いた．

対象物体を各フレーム間で x 、 y 、 z 軸方向に 1 ボクセル平行移動、 z 軸の周りに 2° 回転移動させ、正 12 面体の頂点に配置した 20 台のカメラを用いて、提案手法を適用した．

物体を z 軸周りにのみ回転させるときに 72° 周期で同じ画像が現れること、また、カメラが物体に近接していない場合には原点に対して対称に位置するカメラからはほぼ同じ画像が得られると考えられることから、実験では $18(=72^\circ \div 2^\circ \div 2)$ フレーム分の画像情報を用いて、時系列画像を統合し、全周の情報が得られるようにした．

まず、提案手法で特徴点群から得られた運動推定結果を与えた運動とともに示す．表 2 は表出ボクセルから推定された各軸周りの回転運動成分であり、表 3 は各軸方向の平行移動成分である．2.3 で述べた q 、 N_v の値はそれぞれ、1.679 ピクセル幅、438310 ボクセルであった．

回転移動成分では絶対値の平均で 0.30° 、絶対値の最大で 0.93° の推定誤差が生じた．平行移動成分では絶対値の平均で 0.38 ボクセル、絶対値の最大で 1.24 ボクセルの推定誤差が生じた．これは突出した部分以外から抽出された表出ボクセルの影響によるものと考えられる．

平行移動成分については、ほとんど 1 ボクセル幅以内の誤差に収まり標準化誤差に吸収される程度の精度の運動推定が行われているといえる．回転移動成分については、数値だけでは標準化誤差で吸収される範囲にあるかどうかは判別できないため、統合視体積の結果によって判断する．

1 フレームで得られる視体積と統合視体積の差を図 10 を用いて説明する．このときには $360(= 20 \times 18)$ 台分のカメラ情報を用いて、視体積を構築したことになる．図 10(a) は再現性の向上の限界とみなすことのできるカメラ 40000 台を用いたときに構築される視体

表 2 回転移動成分
Table 2 Rotational parameters.

frame	x (degrees)		y (degrees)		z (degrees)	
	Truth	Diff	Truth	Diff	Truth	Diff
1	0.00	+0.02	0.00	-0.01	2.00	-0.01
2	0.00	+0.02	0.00	-0.02	4.00	-0.06
3	0.00	+0.03	0.00	-0.03	6.00	-0.11
4	0.00	-0.09	0.00	0.01	8.00	-0.14
5	0.00	-0.14	0.00	0.03	10.00	-0.19
6	0.00	-0.16	0.00	0.02	12.00	-0.24
7	0.00	-0.30	0.00	0.04	14.00	-0.24
8	0.00	-0.30	0.00	0.11	16.00	-0.27
9	0.00	-0.36	0.00	0.14	18.00	-0.32
10	0.00	-0.31	0.00	0.15	20.00	-0.39
11	0.00	-0.37	0.00	0.19	22.00	-0.46
12	0.00	-0.43	0.00	0.20	24.00	-0.52
13	0.00	-0.51	0.00	0.23	26.00	-0.56
14	0.00	-0.52	0.00	0.26	28.00	-0.64
15	0.00	-0.66	0.00	0.27	30.00	-0.74
16	0.00	-0.72	0.00	0.25	32.00	-0.79
17	0.00	-0.76	0.00	0.28	34.00	-0.83
18	0.00	-0.77	0.00	0.28	36.00	-0.93

表 3 平行移動成分
Table 3 Translational parameters.

frame	x (voxels)		y (voxels)		z (voxels)	
	Truth	Diff	Truth	Diff	Truth	Diff
1	1.00	-0.04	1.00	-0.02	1.00	-0.06
2	2.00	+0.08	2.00	+0.03	2.00	-0.10
3	3.00	+0.20	3.00	+0.13	3.00	-0.04
4	4.00	+0.35	4.00	+0.14	4.00	-0.04
5	5.00	+0.51	5.00	+0.13	5.00	-0.04
6	6.00	+0.56	6.00	+0.22	6.00	-0.08
7	7.00	+0.60	7.00	+0.24	7.00	-0.09
8	8.00	+0.62	8.00	+0.39	8.00	-0.05
9	9.00	+0.58	9.00	+0.45	9.00	-0.08
10	10.00	+0.60	10.00	+0.52	10.00	-0.09
11	11.00	+0.49	11.00	+0.50	11.00	-0.09
12	12.00	+0.56	12.00	+0.56	12.00	-0.00
13	13.00	+0.62	13.00	+0.60	13.00	-0.01
14	14.00	+0.79	14.00	+0.79	14.00	-0.04
15	15.00	+0.88	15.00	+0.86	15.00	+0.06
16	16.00	+1.01	16.00	+0.90	16.00	+0.13
17	17.00	+1.02	17.00	+1.03	17.00	+0.19
18	18.00	+1.23	18.00	+1.24	18.00	+0.28

積, 図 10(c) は 1 フレームの情報のみから構築される視体積, 図 10(b) は図 10(c) で示した視体積から得られる特徴点群, 更に, 図 10(e) は 18 フレーム統合時の視体積である.

図 10(b) では, 3.3 の図 8 で示した突出した部分に多く特徴点が抽出されていることが確認できる. 図 10(e), (f) では, 腹部の残余領域が減少し, 図 10(c), (d) に比べて背部の角度のある部分がなだらかになっていることが確認できる.

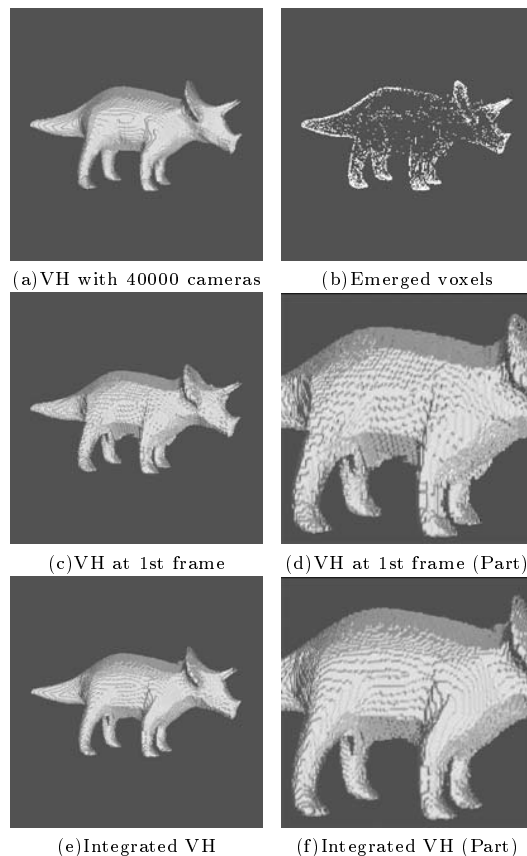


図 10 18 フレームでの視体積統合結果
Fig. 10 Result of integration of 18 visual hulls.

また, 数値的な評価を行うために, 40000 台を用いて得られた視体積を再現性の向上の限界とみなすことのできる視体積として, 正解の視体積とみなし, この正解の視体積と提案手法によって統合された視体積との違いを検証した. 評価には, 提案手法で得られた視体積に含まれ, 正解の視体積に含まれないボクセル (additional voxels), 逆に提案手法で得られた視体積に含まれず, 正解の視体積に含まれるボクセル (missing voxels), 及び, 両者の和 (additional + missing voxels) を用いた. これを図 11 に示す. ただし, 縦軸は α を基準に表現されている.

40000 台を用いて得られる視体積は 438310 ボクセルで表され, $q = 1.679$ ピクセル幅であるので, $k = \sqrt{2}/2$ とするときには, $\alpha = 2.68\%$ である. よって, 11747 ボクセルが additional voxels と missing voxels の目標値となる.

統合するフレームが増加するにつれ, 単調にエラー

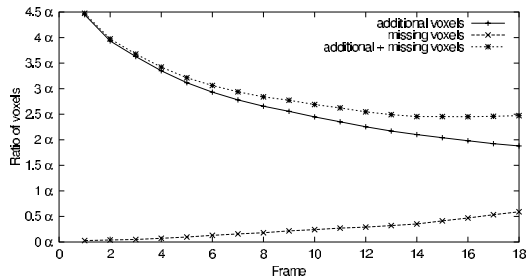


図 11 視体積の再現性の向上 ($q=1.679$)
Fig. 11 Relation between frames and voxels. ($q=1.679$)

表 4 q が異なる環境下での視体積の再現性の向上
Table 4 Relation between frames and voxels with multiple q

q	N_v	α	Add'l($\times\alpha$)	Miss($\times\alpha$)	Both($\times\alpha$)
1.926	363118	2.49%	1.83	0.94	2.77
1.679	438310	2.68%	1.95	0.62	2.57
1.564	541697	2.68%	1.60	0.32	1.92
1.398	677086	2.78%	1.54	0.60	2.14
1.291	860893	2.78%	1.55	0.37	1.92

である additional voxels と missing voxels の和が減少した。additional voxels の数は、1 フレームでの視体積では 4.62α (54358 ボクセル, 12.40%) であったのに対して、統合後の additional voxels は 1.95α (22975 ボクセル, 5.24%) となり、時系列画像の統合により標準化誤差の限界に再現性の向上が近づいていることが分かる。

提案手法で得られた missing voxels の増加は運動推定誤差によって引き起こされるものと考えられる。このボクセルの数の増加は、現実のカメラ増加によっては起こらない問題であり、提案手法の問題点であるといえる。しかし、missing voxels の数は 18 フレーム統合時において 0.62α (7237 ボクセル, 1.65%) であり、標準化誤差による再現性の向上の限界である α の範囲内にとどまっている。よって、再現性の向上を阻害するものではないと考えられる。

また、ボクセルの大きさを変えることで q の値を 5 段階に変化させて、同一の恐竜の形状に対する実験を行った。18 フレーム統合した時点での additional voxels, missing voxels, additional voxels と missing voxels の和を表 4 に示す。また図 12 に、 q の値及び統合に用いたフレーム数に対する additional voxels と missing voxels の和をグラフに示す。

q が異なるときには、18 フレーム統合時に得られる視体積の再現性にばらつきはあるものの、調べたすべ

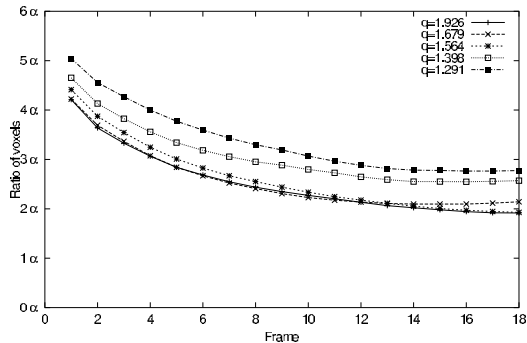


図 12 q の違いによる誤差減少の違い
Fig. 12 Relation between frames and voxels with different q .

ての条件下で、視体積の再現性の向上が確かめられた。

以上の結果より、提案手法で得られる表出ボクセルは画像情報の統合に十分な運動推定を可能にするといえる。また、得られた推定運動によって時系列の画像情報を統合し、視体積の再現性を向上することができた。

5.2 球に近い形状によるシミュレーション実験

次に、提案手法の適用のためにどの程度の大きさや数をもつ突出部分がどの程度の数だけ必要となるのかを確かめるために、半径 r の球を基準として、表面に正弦波の突出部をもつような物体を準備して実験を行った。以下の式に示すように、媒介変数 θ, ϕ を用い、 (x, y, z) で形状を表現する。

$$\begin{cases} x' = \cos \theta \cos \phi \\ y' = \sin \theta \cos \phi \\ z' = \sin \phi \end{cases} \quad (0 \leq \theta \leq 2\pi, -\pi/2 \leq \phi \leq \pi/2)$$

$$r' = A_l \cos F_l \pi x' \cdot \cos F_l \pi y' \cdot \cos F_l \pi z'$$

$$\begin{cases} x = (r + r') \cos \theta \cos \phi \\ y = (r + r') \sin \theta \cos \phi \\ z = (r + r') \sin \phi \end{cases}$$

A_l は振幅を表し、 F_l は周期を表す。 $r = 50$ として、 F_l, A_l を様々に変化させて提案手法を適用すると、 F_l, A_l が大きいほど表出ボクセルを特徴点とする運動推定は正確になり、再現性の向上が行われた。ここでは、 $F_l = 4, 6, A_l = 2, 4$ にそれぞれ設定した形状に対して提案手法を適用した結果を示す。それぞれの形状を図

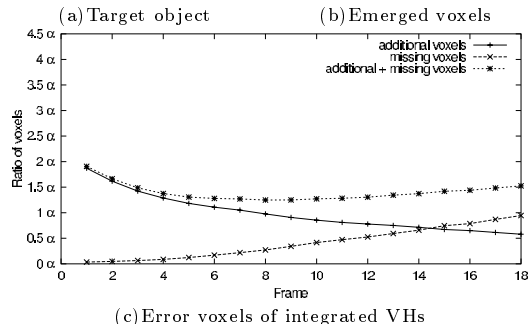
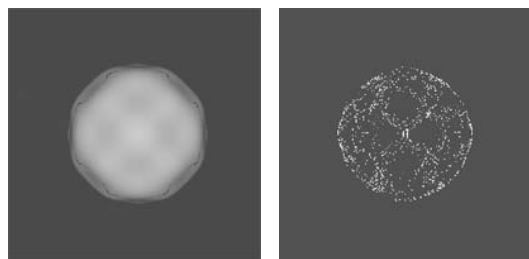


図 13 18 フレームでの視体積統合結果 ($F_l=4, A_l=2$)
Fig.13 Result of integration of 18 visual hulls.
($F_l=4, A_l=2$)

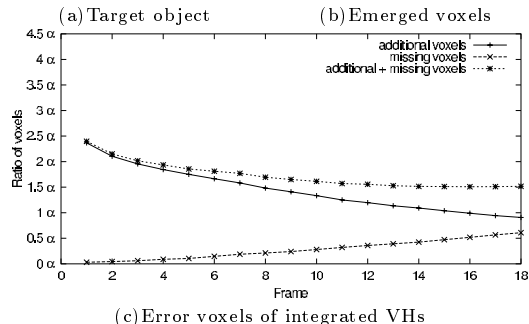
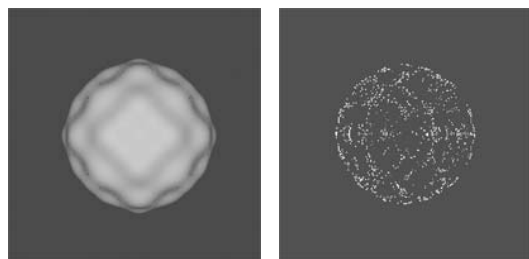


図 15 18 フレームでの視体積統合結果 ($F_l=6, A_l=2$)
Fig.15 Result of integration of 18 visual hulls.
($F_l=6, A_l=2$)

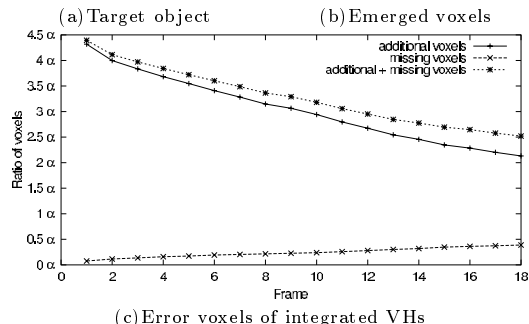
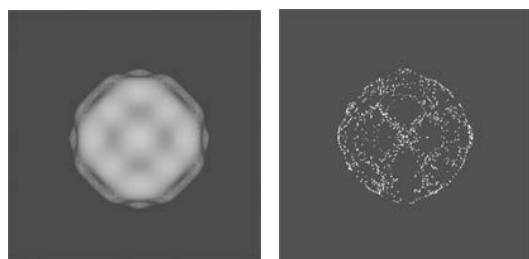


図 14 18 フレームでの視体積統合結果 ($F_l=4, A_l=4$)
Fig.14 Result of integration of 18 visual hulls.
($F_l=4, A_l=4$)

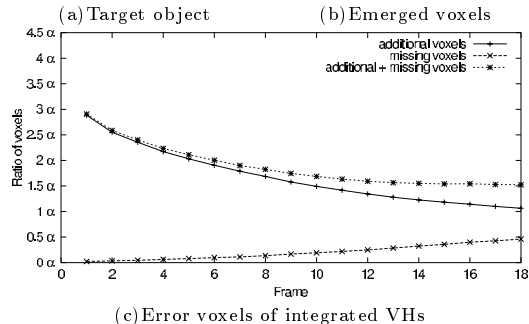
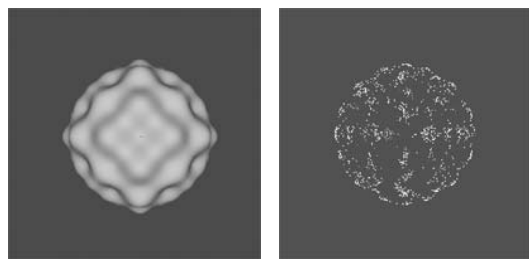


図 16 18 フレームでの視体積統合結果 ($F_l=6, A_l=4$)
Fig.16 Result of integration of 18 visual hulls.
($F_l=6, A_l=4$)

13(a), 図 14(a), 図 15(a), 図 16(a) に示し, 視体積から得られる表出ボクセルを図 13(b), 図 14(b), 図 15(b), 図 16(b) に示す.

q, N_v 及び α の値は表 5 のようになった. なお,

$k = \sqrt{2}/2$ とした.

5.1 の恐竜と同様の運動を与えて, 評価を行った. これを図 13(c), 図 14(c), 図 15(c), 図 16(c) に示す.

$F_l = 4, A_l = 2$ の形状を除いて, 18 フレーム目

表 5 各対象物体の変数値
Table 5 Parameters of target objects.

	q	N_v	α
$F_l = 4, A_l = 2$	1.679	1193405	1.92%
$F_l = 4, A_l = 4$	1.679	1146325	1.95%
$F_l = 6, A_l = 2$	1.679	1208889	1.91%
$F_l = 6, A_l = 4$	1.679	1223334	1.90%

で additional voxels と missing voxels の和は単調に減少し、提案手法が有効に働くことが確認できた。 $F_l = 4, A_l = 2$ の形状では、突出部の幅と数とともに十分でないために、9 フレーム目以降で additional voxels と missing voxels の和が増加した。今回実験した環境下では、 $F_l = 6$ 以上または $A_l = 4$ 以上であることが提案手法の適用のために必要であることが分かった。

なお、上で用いた $q = 1.560, r = 50$ の物体の場合には、 $A_l = 4$ のときの突出の幅が全体の 8% を占め、最大で 6.24 ピクセル幅として投影される。 $F_l = 6$ の場合には、おおよそ 30° おきに突出部分が存在する計算になる。

5.3 様々な形状によるシミュレーション実験

次に、形状復元の対象として選択されることが考え得る現実的な物体のうちで突出部の小さい形状に対して提案手法を適用し、どのような形状に対して提案手法が有効に働くのかについて調べた。対象物体として、バナナ、ティーポット及びチェスの駒を選択し、シミュレーション実験を行った。それぞれの物体は、突出した部分が少なく表出ボクセルの抽出が難しいと予想される形状として選択した。図 17、図 18、図 19 はそれぞれバナナ、ティーポット及びチェスの駒に対して提案手法の適用を行った結果である。それぞれの形状に対して、5.1 の恐竜と同様の運動を与えて、評価を行った。

図 17 (c) に示す統合結果のグラフでは、7 フレーム目以降で additional voxels と missing voxels の和が増加している。よって、図 17 に示したバナナの形状に対しては提案手法が有効に働かないといえる。

また、図 17(c) のグラフは、図 13(c) に示した $F_l = 4, A_l = 2$ の表面に正弦波の突出部を持つ球について得られたグラフとよく似ており、1 フレーム目で 2α 程度の additional voxels が残り、それぞれ 7 フレーム目と 9 フレーム目以降で additional voxels と missing voxels の和が増加している。これは、図 17 に示したバナナの形状は、図 13 に示した、 $F_l = 4,$

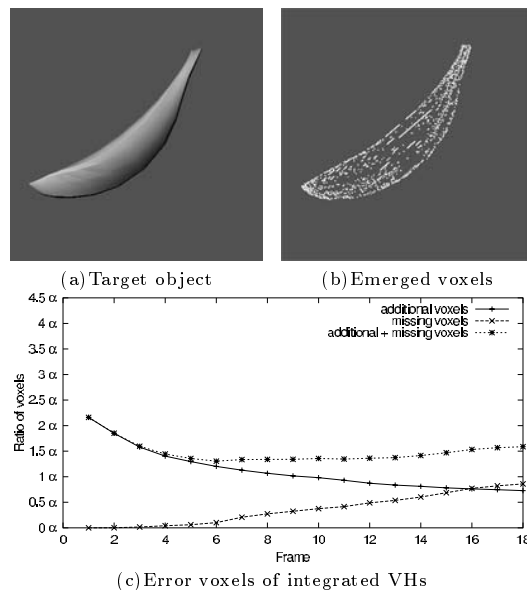


図 17 18 フレームでの視体積統合結果 (バナナ)
Fig. 17 Result of integration of 18 visual hulls of banana.

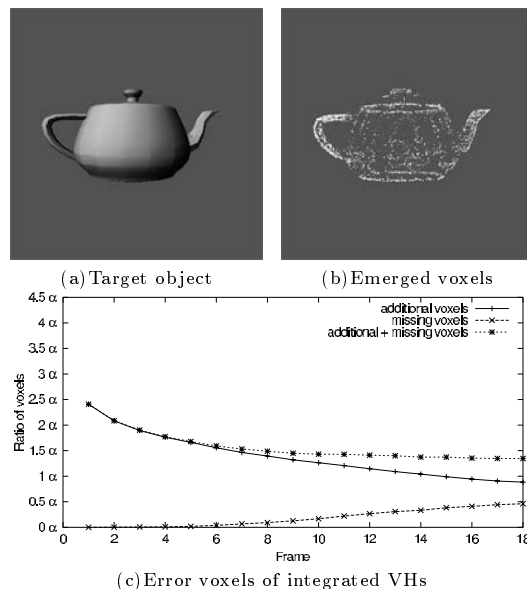


図 18 18 フレームでの視体積統合結果 (ティーポット)
Fig. 18 Result of integration of 18 visual hulls of teapot.

$A_l = 2$ の表面に正弦波の突出部をもつ球と同様、表出ボクセルの抽出に十分な大きさと数をもつ突出がない形状であるためであるといえる。

一方、図 18(c)、図 19 (c) にそれぞれ示した統合結果

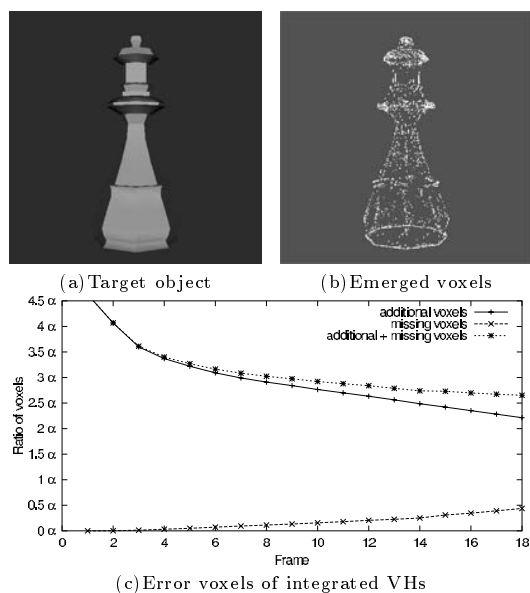


図 19 18 フレームでの視体積統合結果 (チェスの駒)
Fig. 19 Result of integration of 18 visual hulls of queen of chess.

では、18 フレームまでで additional voxels と missing voxels の和が単調に減少しており、ティーポットとチェスの駒のそれぞれの形状では、提案手法が有効に働くといえる。また、図 18(c)、図 19(c) のグラフは、図 15(c)、図 14(c) にそれぞれ示した $F_l = 6$, $A_l = 2$ 及び $F_l = 4$, $A_l = 4$ の表面に正弦波の突出部をもつ球について得られたグラフと、1 フレーム目と 18 フレーム目での additional voxels と missing voxels の和の値がよく似ている。図 18 及び図 19 に示したそれぞれの形状では、図 15(c)、図 14(c) にそれぞれ示した $F_l = 6$, $A_l = 2$ および $F_l = 4$, $A_l = 4$ の表面に正弦波の突出部をもつ球と同様、表出ボクセルの抽出に十分な大きさや数をもつ突出がある形状であるといえる。このことは、図 18(b) および図 19(b) に示した表出ボクセルが、突出部分に対応するところから得られていることからいえる。

このように、突出の大きさや数が少ないと考えられる物体でも、形状によっては表出ボクセルの抽出が可能で、提案手法が有効に機能するものがあることから、提案手法が有効に機能する物体は、必ずしも突出が顕著である物体のみに限らないことが確認できた。

また、突出の大きさや数が十分でないバナナや表面に $F_l = 4$, $A_l = 2$ の正弦波の突出部をもつ球で

は、1 フレーム分の情報のみから additional voxels と missing voxels の和が 2α 前後であるような視体積が得られていた。この和は突出の大きさや数が十分であるような他の物体に対し、相対的に少ないといえる。更に、2.3 の図 4 に示した物体ごとのカメラ台数と残余領域の関係からも、突出部の少ない形状である球やトーラス環では、少ないカメラ台数から正解の視体積に近い視体積を得られることが分かる。これらのことから、物体表面に十分な大きさや数の突出部がないような物体では、もともと、提案手法の適用があまり必要ではなく、逆に、物体表面に十分な大きさや数の突出部があるものほど、提案手法の適用が必要であるといえる。

5.4 実環境実験

最後に、提案手法が実環境でも有効であることを示すために、実環境で計測した画像を用いて時系列画像の統合を行った。空間中に 19 台のカメラを設置し、対象物体を時系列で撮影した。対象物体は上から細い糸で吊るし、揺らすことで微小な運動を与えた。

図 21(b) から、実物体に対しても、突出した部分に対応する表出ボクセルが得られることが確かめられた。

しかし、図 21(e) に示した 5 フレーム統合した視体積では復元形状に欠損が見られる。これは、画像からシルエットを抽出するときに欠損が生じたためである。画像のシルエット欠損による視体積の欠損は、2.1 で述べたように視体積交差法がすべてのカメラ画像から得られる視錐体の積領域を求めるために起こる問題であり、実環境でカメラ台数を増加させる場合にも起こる。シルエット抽出例を図 20 に示す。

シルエットの抽出は取得画像 (図 20(a)) と背景画像 (図 20(b)) をそれぞれ YUV 空間に変換して差分値の大きい領域をシルエット (図 20(c)) としたが、欠損は主に物体と背景色の近い部分で多く起こっていた。この問題に対し、 N_{allow} を 1, 2 に設定したものをそれぞれ図 21(g), (i) に示した。図 21(g) では、統合により 1 フレームで得られる視体積よりも再現性の高い形状を復元できている。しかし、図 21(i) では、 N_{allow} を大きく設定しすぎて統合の優位性が失われている。この問題の解決のためには、画像からのシルエット抽出の精度を上げることが必要である。

6. む す び

本研究では、視体積交差法において時系列で得られる画像を統合し、現実的なカメラ台数で理論的限界の

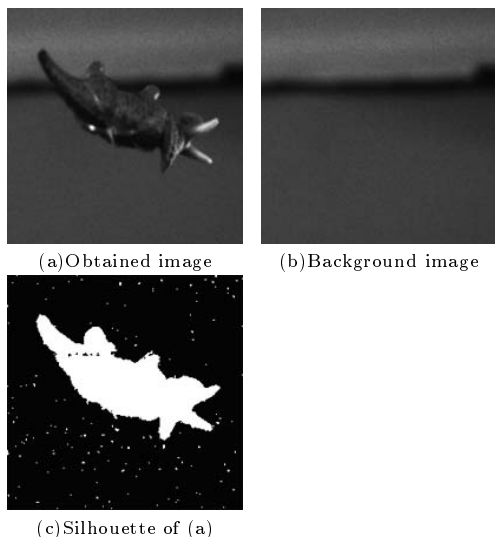


図 20 シルエット欠損
Fig.20 Chip on silhouettes.

再現性を示す台数のカメラを用いたときと同様の復元形状を得ることを目指した。各時刻で表出ボクセルを求め、これを追跡することにより対象物体の運動推定を行い、別の時刻での画像の情報を統合して視体積の構築を行った。これにより、理論的限界に近い再現性をもつ復元形状を得ることができた。

今後の課題として、時系列の画像から背景差分を精度良くとることが挙げられる。実環境では、画像上にノイズが生じ、多くの画像情報を統合するときに問題となった。ノイズによるシルエット欠損の解消は、時系列に並んだ画像と視体積を用いて行うことで実現できるのではないかと考えている。

文 献

[1] W.N. Martin and J.K. Aggarwal, "Volumetric description of objects from multiple views," IEEE Trans. Pattern Anal. Mach. Intell., vol.5, no.2, pp.150-158, 1983.
 [2] A. Laurentini, "How far 3D shapes can be understood from 2D silhouettes," IEEE Trans. Pattern Anal. Mach. Intell., vol.17, no.2, pp.188-195, 1995.
 [3] K.N. Kutulakos and S.M. Seitz, "A theory of shape by space carving," Int. J. Comput. Vis., vol.38, no.3, pp.199-218, 2000.
 [4] A. Y. Müayim, U. Yilmaz and V. Atalay, "Silhouette-based 3D model reconstruction from multiple images," IEEE Trans. Syst. Man Cybern. B, vol.33, no.4, pp.582-591, August 2003.
 [5] S. Sullivan and J. Ponce, "Automatic model construction and pose estimation from photographs us-

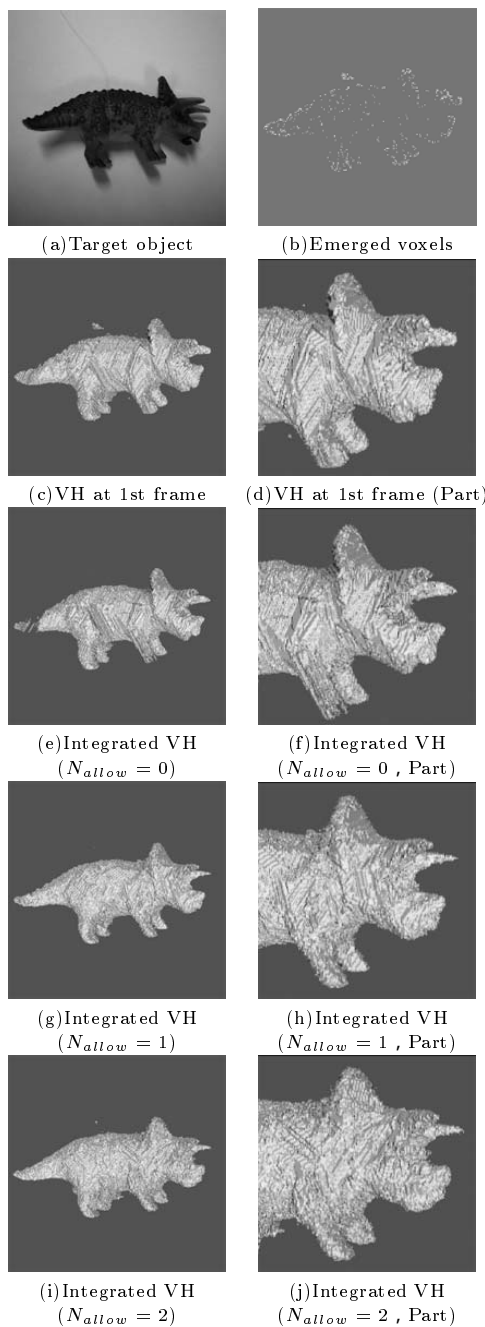


図 21 5フレームでの視体積統合結果
Fig.21 Result of integratation of 5 visual hulls.

ing triangular splines," IEEE Trans. Pattern Anal. Mach. Intell., vol.20, no.10, pp.1091-1096, Oct. 1998.
 [6] J.Y. Zheng, "Acquiring 3-D models from a sequence of contours," IEEE Trans. Pattern Anal. Mach. In-

- tell., vol.16, no.2, pp.163-178, Feb. 1994.
- [7] S. Rusinkiewicz, O. Hall-Holt, and Marc Levoy, "Real-time 3D model acquisition," Trans. Graphics (SIGGRAPH proceedings), pp.438-446, 2002.
- [8] G.K.M. Cheung, S. Baker, and T. Kanade, "Visual hull alignment and refinement across time: A 3D reconstruction algorithm combining shape-from-silhouette with stereo," Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp.77-84, 2003.
- [9] 築澤 宗太郎, 鷲見 和彦, 松山 隆司, "装着型視覚センサを用いた手持ち物体の3次元デジタル化," 情処学 CVIM 研報, no.143-4, March 2004.
- [10] S. Vedula, S. Baker, S. Seitz, and T. Kanade, "Shape and motion carving in 6D," Proc. IEEE Conference on Computer Vision and Pattern Recognition, vol.2, pp.592-598, 2000.
- [11] K. Forbes, A. Voigt, and N. Bodika, "Using silhouette consistency constraints to build 3D models," Proc. the Fourteenth Annual Symposium of the Pattern Recognition Association of South Africa (PRASA 2003), Nov. 2003.
- [12] C.H. Esteban and F. Schmitt, "Using silhouette coherence for 3D image-based object modeling under circular motion," Technical Report 2003D011, Ecole Nationale Supérieure des Télécommunications, Sept. 2003.
- [13] R. Cipolla, K.E. Astrom, and P.J. Giblin, "Motion from the frontier of curved surfaces," Proc. 5th International Conference on Computer Vision, pp.269-275, 1995.
- [14] 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, 2000.
- [15] W. Niem, "Error analysis for silhouette-based 3D shape estimation from multiple views," Proc. International Workshop on Synthetic - Natural Hybrid Coding and Three Dimensional Imaging (IWSNHC3DI'97), pp.6-9, Rhodos, Sept. 1997.
- [16] D. Snow, P. Viola, and R. Zabih, "Exact voxel occupancy with graph cuts," Proc. IEEE Conference on Computer Vision and Pattern Recognition, vol.1, pp.345-352, 2000.
- [17] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, 丹慶勝市, 奥村晴彦, 佐藤俊郎, 小林誠訳, NUMERICAL RECIPES in C, 技術評論社, 1993.

(平成 16 年 10 月 8 日受付, 17 年 2 月 7 日再受付,
6 月 16 日最終受付)



豊浦 正広 (学生員)

平 15 京大・工・情報工学卒。平 17 同大大学院情報学研究科修士課程了。現在、同大学院・情報学研究科博士後期課程在学中。コンピュータビジョン・三次元モデルの研究に従事。



飯山 将晃 (正員)

平 10 京大・工・情報工学卒。平 15 同大大学院情報学研究科博士後期研究指導認定退学。同年同大学術情報メディアセンター助手。コンピュータビジョン・三次元データ処理の研究に従事。情報処理学会会員。



角所 考 (正員)

昭 63 名大・工・電気卒。平 5 阪大大学院工学研究科通信工学専攻博士課程了。平 4~6 日本学術振興会特別研究員。平 5~6 スタンフォード大学ロボティクス研究所客員。平 6 大阪大学産業科学研究所助手。平 9 京都大学総合情報メディアセンター助教授。平 14 同大学術情報メディアセンター助教授。視覚メディア処理, コミュニケーション, インタラクションに関する研究に従事。博士(工学)。IEEE, ACM, 情報処理学会, 人工知能学会各会員。



美濃 導彦 (正員)

昭 53 京大・工・情報工学卒。昭 58 同大大学院博士課程了。同年工学部助手, 昭 62-63 マサチューセッツ州立大学客員研究員, 平元京都大学工学部附属高度情報開発実験施設助教授, 平 7 同教授, 平 9 京都大学総合情報メディアセンター教授, 平 14 京都大学学術情報メディアセンター教授。画像処理, 人工知能, 知的コミュニケーション関係の研究に従事。工博。IEEE, ACM, 情報処理学会, 画像電子学会, 日本ロボット学会各会員。

Abstract In the volume intersection method, the more cameras are used, the more accurate shapes can be reconstructed. Generally, setting many cameras up in a small space is difficult due to physical restrictions. In the case when the rigid object motion of the target is estimated, images in time sequences can be integrated as if they were obtained from the different positions in the space. The motion is estimated with the feature points of a visual hull at each time. Since the visual hull includes target object region and additional region, to extract feature points which are corresponded each other between frames is difficult. The proposed method is that the feature points are extracted by reprojection of voxels of the visual hull. As a result, we can get more accurate shapes than those by the previous methods by integrating the visual hulls from the images in time sequences in both of simulation and actual environment.

Key words Volume intersection method, extracting feature points, motion tracking, time series data, sampling error