

運動推定を用いた3次元復元形状の再現性の向上

An Accurate Shape Reconstruction Method by Motion Tracking

豊浦 正広¹⁾, 飯山 将晃²⁾, 角所 考²⁾, 美濃 導彦²⁾

Masahiro TOYOURA, Masaaki IYAMA, Koh KAKUSHO, Michihiko MINOH

1) 京都大学大学院 情報学研究科

(〒606-8501 京都市左京区吉田本町, toyo@mm.media.kyoto-u.ac.jp)

2) 京都大学学術情報メディアセンター

(〒606-8501 京都市左京区吉田本町, {iyama, kakusho, minoh}@media.kyoto-u.ac.jp)

Abstract : In the volume intersection method, the more accurate shapes can be reconstructed with the more cameras. Generally, using many cameras is difficult due to physical restrictions and difficulty of calibration. In the case when the rigid object motion of the target is estimated, images can be integrated in time sequences. The motion is estimated with the feature points of a visual hull at each time. The visual hull includes target object region and additional region. The additional region makes it difficult to extract feature points. Feature points are extracted by back projection of voxels of the visual hull. As a result, we can get more accurate shapes from images in time sequences.

Key Words: *Volume intersection method, feature points, motion tracking, time series data.*

1. はじめに

多種多様な物体の3次元形状の復元手法として視体積交差法 [1] がある。視体積交差法は、複数のカメラから対象物体を観測し、それぞれのカメラで得られるシルエット画像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、広範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に対しての形状復元に優れているためである。ただし、物体の凹面を復元することは原理上不可能であり、復元形状として物体の凸包が得られる。

視体積交差法では、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。これに伴い、復元形状を原形状により近づけることができるようになってきた。大量カメラが物理的には入手可能になったが、空間的に設置できるカメラ台数には限界がある。そのため、視体積交差法で得られる復元形状の再現性の向上にもまた限界がある。本稿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、これを目標とする。

そこで、数十台程度の現実的な台数のカメラを用いた環境で時系列に並んだ画像情報を統合することで、カメラ台数を増加させた場合と等価な視体積を得る手法を考える。対

象物体が剛体であり、観測している間にその形状が変化しないときには、物体が運動することによる対象とカメラの相対位置の時間的変化をカメラ位置の空間的な変化とみなすことができ、複数の時刻での視体積の情報を統合することで、1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。

カメラの相対位置の時間的変化をカメラ位置の空間的な変化とみなす手法として、ターンテーブルを用いる手法が一般的に用いられる。ターンテーブルを用いる手法では、物体の運動方向がターンテーブルの回転軸まわ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を復元することも難しい。

これに対し、Cheungら [2] 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た。しかし、この手法では特徴点の追跡に物体の色情報を用いており、対象物体として物体表面の色変化に乏しい物体を選ぶときには、この手法は有効ではない。

これらの従来研究に対して、我々は色情報を用いず、各時刻の視体積の形状情報のみから特徴点をとらえて、運動推定を行う。このときに、各時刻の視体積から異なる時刻間で対応する点を抽出することが難しい。なぜなら、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により、

物体表面に現れる部分の特徴点として抽出しようとすると、ある時刻では余分な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。

そこで、視体積表面に現れる物体表面上の点に注目する。視体積表面に現れる物体表面上の点の中には、カメラといふ物体の位置関係に依らず安定して抽出されるものが多数存在することに着目し、これらの特徴点として抽出することによって、運動推定を行う。

2節ではまず、本研究の背景である視体積交差法について説明する。3節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節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節では形状復元を含めた実験結果を示し、その精度について考察を行い、6節で全体についてまとめる。

2. 視体積の再現性の向上

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。

カメラ C_m 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2次元画像平面上の領域として得られる。この領域を C_m に対する物体投影領域と呼び、 R_m で表す。

物体 O はカメラ C_m のレンズ中心を端点とし、物体投影領域 R_m 上の任意の点を通る半直線の集合からなる錐体状の開空間に内接して存在する。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。ここで、カメラ C_1, \dots, C_n における画像から得られる視錐体 V_1, V_2, \dots, V_n について、式(1)のような V を考え、視体積と呼ぶ。

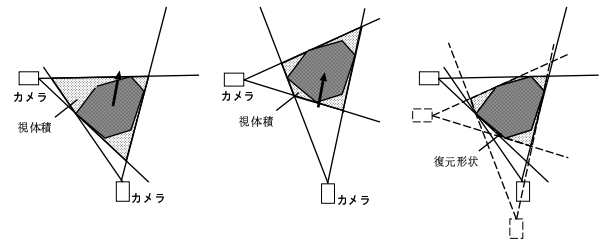
$$V = \bigcap_{m=1}^n V_m \quad (1)$$

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。

ここで1節に述べたように、数十台程度の現実的な台数以上のカメラ数の増加を実現するために、時系列で得られた視体積を統合することを考える。複数の時刻での物体の姿勢と物体の視体積が得られ、かつ、物体が剛体ならば、各時刻間の物体の運動を推定し、カメラ台数を仮想的に増加させることができる。

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、元の位置での視体積と運動後の視体積が図1のように得られた状態を考える。

元の位置での視体積は、運動後の位置にある物体を物体の運動分だけ差し引いた位置にカメラを設置したときに復元できる視体積(図1(c))とみなすことができる。つまり、運動後の視体積に元の位置の視体積の情報を統合して、2台



(a) 移動前の視体積 (b) 移動後の視体積 (c) 統合視体積

図 1: 視体積の統合

のカメラで4台のカメラによって得られる視体積と同等のものが得られる。

3. 特徴点抽出

本研究では、各時刻で求められる視体積から特徴点を抽出し、隣接する時刻間に対応する特徴点の運動を考えることにより、対象物体の運動推定を行う。本節ではまず、各時刻での視体積からの特徴点抽出を考える。

3.1 特徴点に要求される性質

特徴点として物体とカメラの位置関係が変化しても、安定して抽出されるような点を用いることが必要となる。ここで、各時刻の視体積からの特徴点抽出では、視体積が余分な領域を含むことが問題となる。時系列で安定した特徴点抽出のためには、各時刻の余分な領域による物体表面の隠蔽を考慮に入れた特徴点抽出が必要となる。

また、視体積交差法の特長である物体表面の詳細な色情報を必要としないという性質を生かすことを考えると、与えられる情報は画像の物体投影領域情報、および、そこから得られる各時刻の視体積形状のみであるといえる。我々はこれらの情報のみを用いる特徴点抽出手法を提案する。

3.2 逆投影による特徴点抽出手法

特徴点を求めるために、図2のように視体積表面上のボクセルを画像へ逆投影する。物体投影境界領域の画素を通る視線に注目し、その視線上に視体積であるボクセルが1つだけ存在するとき、この視体積表面上のボクセルを特徴点に採用する。これを対象となる時刻の視体積を構築するすべての画像について行う。

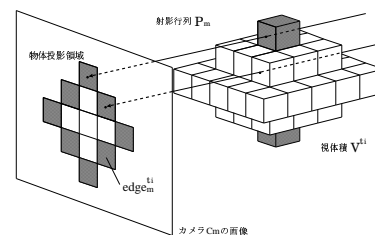


図 2: 境界要素の再投影

つまり、時刻 t_i における特徴点 F^{t_i} は以下の式で表される。

$$F^{t_i} = \{v \mid \forall m, P_m v \in \text{edge}_m^{t_i}, v \in V^{t_i}\} \quad (2)$$

ただし、 P_m はカメラ C_m の射影行列、 $\text{edge}_m^{t_i}$ は時刻 t_i におけるカメラ C_m の物体投影境界領域、 V^{t_i} は時刻 t_i の視体積である。

このようにして得られた特徴点は視体積表面上のボクセルであることはもちろんであるが、物体表面上に含まれるボクセルであることも保証ができる。もし、斜線で示したボクセルが物体領域に含まれないとすれば、図 2 で示す視体積の投影像は物体投影領域に一致しないことになり矛盾が生じる。よって、斜線で示したボクセルが物体領域に含まれるボクセルであることがいえる。

特徴点は、各時刻で対応するものが抽出される。なぜなら、物体形状で突出しているような部分が特徴点に抽出されやすいためである。図 3 にこれを示す。図 3(a), (c) のように突出した物体形状を持つような部分ではカメラと物体の相対的な位置が変化しても、特徴点となるボクセルが変化しない。一方、図 3(b), (d) のように突出しない物体形状を持つような部分では、カメラと物体の相対的な位置変化によって、特徴点となるボクセルが変化する。これは他のボクセルからの自己隠蔽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各時刻で対応する点が特徴点として抽出されやすい。

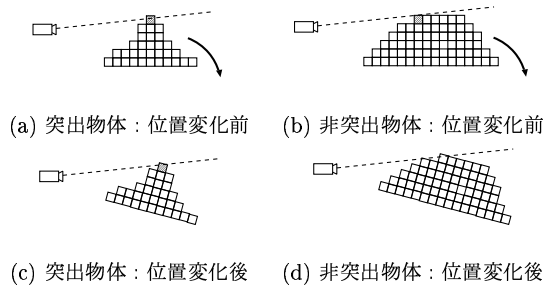


図 3: 相対位置の変化による特徴点の変化

運動推定を行う時刻間の物体の運動が小さい場合には、物体表面上で対応するボクセルが多数抽出されることが予想される。これは、物体の運動変化が小さい場合には、視体積に含まれる余分な領域が物体表面上の同様の部分に存在するために、ある時刻で現れる特徴点が次の時刻でも観測されやすいと考えることができるためである。

しかし、すべての特徴点が、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特徴点として抽出され得るからである。これは視体積が余分な領域を含むためである。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特徴点のみを推定に用いる工夫を行っている。これは 4 節で扱う。

4. 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。運動追跡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 $\mathbf{t}^{t_i t_j}$ を用いて、次のように表すことができる。

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

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。回転行列には 4 元数表示を用い、線形性を高め収束法の適用をより有効にした。また、 $D^{t_i t_j} = (D^{t_j t_i})^{-1}$ が常に成り立つ。

4.2 運動成分の求め方

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 [3] 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持つ。また、ロバスト推定を用いて、時刻 t_i と時刻 t_j で対応する点を持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 で求めた特徴点は、剛体変換行列 $D^{t_i t_j}$ により移動させると、時刻 t_j で求めた特徴点に一致するはずである。しかし実際には、量子化誤差や特徴点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 で求めた特徴点 $p_s^{t_i}$ 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 における特徴点のうちで最も近くにある特徴点を $p_u^{t_j}$ に対応する点と考え、この点との二乗距離を求める。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。

$$E = \sum_u f \left(\min_s (p_u^{t_j} - D^{t_i t_j} p_s^{t_i})^2 \right) \quad (4)$$

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

ここで、 $p_s^{t_i}$ 、 $p_u^{t_j}$ はそれぞれ時刻 t_i 、 t_j での特徴点の座標を表し、 M_{th} は時刻 t_i 、 t_j の間の移動距離の最大値を定めた閾値である。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持つような特徴点が収束に与える影響を小さくしている。

任意の時刻間での剛体変換行列 $D^{t_i t_j}$ を求めるには、まず、隣接する時刻間で剛体変換行列を求める。これを用いて、任意の時刻間での剛体変換行列 $D^{t_i t_j}$ を以下のように

求める.

$$D^{t_i t_j} = D^{t_j-1 t_j} D^{t_j-1 t_j-2} \dots D^{t_{i+1} t_{i+2}} D^{t_i t_{i+1}} \quad (5)$$

5. 実験結果

実験はシュミレーションデータを用いて行った. 3次元データはVRML形式で与えられ, 仮想空間に配置したカメラから撮影を行い, 視体積の構築を行った. VRML形式のデータを座標値調整により物体を仮想的に移動させ, それぞれの位置に対して撮影した画像をテストデータとして用いた. 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面体の頂点上に配置した20台のカメラを用いて, 提案手法を適用した.

物体をz軸周りにも回転させるときに72°ごとに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照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°÷2°÷2)フレーム分の画像情報を用いて, 統合視体積を構築した.

1フレームで得られる視体積と統合視体積の差を図4を用いて説明する. このときには360(=20×18)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4(a)は正解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積, 図4(b)は1フレームの情報のみから構築される視体積, 図4(c)は図4(b)で示した視体積から得られる特徴点群, さらに, 図4(d)は18フレーム統合時の視体積のそれぞれ一部分である. 太線で1フレームの情報のみから構築される視体積の輪郭の一部を示す.

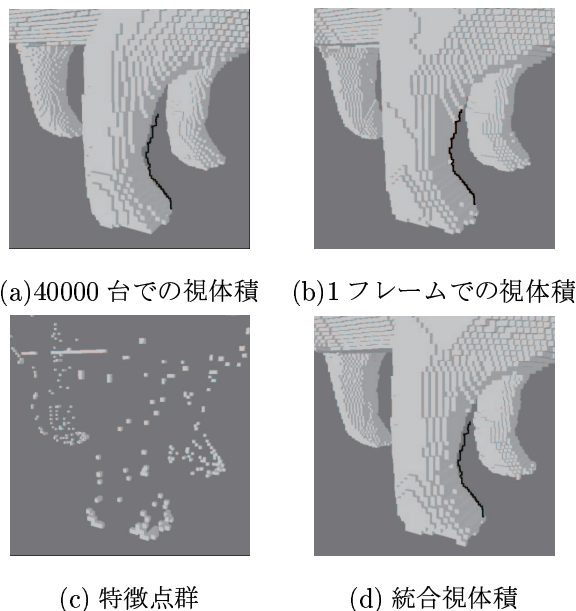


図 4: 18 フレームでの視体積統合結果

図4(c)では, 物体上の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4(d)で示す足の前後のくびれの部分について, 特に再現性の向上が顕著に見られる.

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正解の視体積とし, 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 (extra voxels), 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), 両者の和 (extra + missing voxels) を用いた. これを図5に示す.

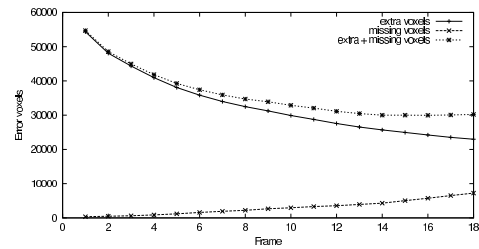


図 5: 視体積の再現性の向上

統合するフレームが増加するにつれ, 単調に誤差である統合視体積に含まれる余分なボクセルと統合視体積に含まれない正解の視体積中のボクセルの和が減少した.

以上の結果より, 提案手法で得られる特徴点群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, 視体積の再現性を確認することができた.

6. おわりに

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の情報を統合し, 現実的なカメラ台数で多くの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. 各時刻で特徴点を求め, 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.

今後の課題として, 実環境での適用を試みることを挙げる. 実環境下では, 物体のシルエットが正確に得られない場合がある. このような場合の特徴点抽出方法について, 研究を進めていく.

参考文献

- [1] Aldo Laurentini, "How Far 3D Shapes Can Be Understood from 2D Silhouettes," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.17, No.2, pp.188-195, 1995
- [2] German K.M. Cheung, Simon Baker, Takeo Kanade, "Visual Hull Alignment and Refinement Across Time: A 3D Reconstruction Algorithm Combining Shape-From-Silhouette with Stereo," Proc. CVPR, pp.77-84, 2003.
- [3] NUMERICAL RECIPES in C, 技術評論社, William H.Press, Saul A.Teukolsky, William T.Vetterling, Brian P. Flannery