
多視点画像からの3次元情報抽出

Three Dimensional Information Extraction from Multiple-View Images

黄 英傑* 坂本 拓之** 西田 広文***

Ying-jieh HUANG Takuji SAKAMOTO Hirobumi NISHIDA

要 旨

動画像からの対象物体の3次元形状推定は、コンピュータ・ビジョンの分野の中に一つの大きな研究課題であり、応用面、例えば、ロボット・ビジョン、自動走行車、モデルベースの画像符号化、3次元モデリングなどでも大きな関心が持たれている。因子分解法 (factorization) は線形射影モデルに基づいて、推定パラメータ (運動・形状) について線形となるように定式化し、数値誤差に対して頑健な行列の特異値分解を用いたため、他の手法に比べて、解は極めて安定であることが特徴である。因子分解法においては、動画像の最初の画像フレーム上で選ばれたすべての特徴点が、その後のすべてのフレームにおいて追跡できることを前提としている。本論文では、特定の画像フレーム上で隠れて見えないような特徴点について、その画像面での投影位置を、実際のカメラモデルである中心射影による投影にできるだけ近くなるように、高い精度で推定する方法を提案する。合成データを使って復元精度の評価を行うとともに、実写画像から復元された3次元情報の結果をも示す。

ABSTRACT

The recovery of 3-D information from an image sequence has been an important research subject in computer vision, as well as in such application areas as robot vision, autonomous vehicles, model-based image/video coding, and 3-D modeling. Based on the linear projection model, the factorization method formalizes the problem as solving a set of linear equations in terms of the shape and motion parameters with the Singular Value Decomposition, which is known to be robust against numerical errors. The factorization method assumes that all the feature points selected on the first frame of the image sequence can be tracked throughout the sequence. This assumption is violated, however, if some feature points observed on the previous frame disappear or new features are introduced on a later frame. In this paper, based on the factorization method with the paraperspective projection model, we present a method for accurate inference of the projection of an occluded feature point onto the image plane so that the estimated location can be close to the mapping by the perspective projection. The results of estimation are evaluated by using synthetic data and some results of 3-D information extraction are shown by using real image sequences.

* 研究開発本部 情報通信研究所

Information and Communication R&D Center, Research and Development Group

** リコーシステム開発(株)

Ricoh System Kaihatsu Company, Ltd.

*** 研究開発本部 ソフトウェア研究所

Software Research Center, Research and Development Group

1. 背景と目的

最近、デジタルカメラをはじめ、デジタルビデオカメラも市場に出回ってきている。これらのデジタルメディアは従来のアナログメディアに比べて、編集・加工・処理が容易にできるという大きな利点があるため、デジタルメディアの再利用、あるいは、デジタルメディアから有用な情報を取り出す技術が重要になってきている。取り込んだ一連の静止画像、あるいは、動画像からの対象物体の3次元形状推定は、コンピュータ・ビジョンの分野の中に一つの大きな研究課題であり、応用面、例えば、ロボット・ビジョン、自動走行車、モデルベースの画像符号化、3次元モデリングなどでも大きな関心が持たれている。

時系列で撮影された2次元動画像からの3次元情報の抽出問題において、いわゆる「運動からの構造推定 (Structure From Motion)」という方法では、「運動距離 形状」という手順にしたがって、まず、カメラ運動を計算してから、物体上の特徴点のカメラ中心からの距離を求めることにより、形状を推定する。しかし、動画像ではフレーム間での動きが小さいので、運動を平行移動あるいは回転によって特定するのはほぼ不可能に近い。結局、解として求められた距離が数値的に不安定になり、形状の推定精度が上がらないことになる。逆に時間のサンプリング間隔を大きく取る場合には、フレーム間での動きが大きくなるので、特徴点の対応付けの信頼性が低下してしまう。

運動と形状を同時に計算することにより解を安定させる方法として、カーネギー・メロン大学のTomasi & Kanade¹⁾により提案された方法が、「因子分解法 (factorization)」である。正射影 (平行投影) モデルに基づいて、推定パラメータ (運動・形状) について線形となるように定式化し、数値誤差に対して頑健な行列の特異値分解を用いたため、他の手法に比べて、解は極めて安定であることが特徴である。そして、定式化の線型性を保ちながら、中心射影に対する近似精度がより高い、疑似中心射影 (paraperspective) モデルを利用する因子分解法を、Poelman & Kanade²⁾が提案した。

因子分解法においては、動画像の最初の画像フレーム上で選ばれたすべての特徴点が、その後のすべてのフレームにおいて追跡できることを前提としている。すなわち、最初の画像フレームで見えている特徴点が途中で隠れて見えなくなったり、あるいは、途中の画像フレームで新しい特徴点を導入しないことを仮定している。しかしながら、カメラが物体の周りを一回りするような状況では、最初の画像フレームで見えている特徴点は途中で見えなくなるので、このような仮定は長時間に渡って撮影されたような動画像には適用できない。この問題への

対処として、Tomasi & Kanade¹⁾は、特定の画像フレーム上で隠れて見えないような特徴点について、その画像面での投影位置を推定する方法を提案している。すなわち、物体が透明な場合に、その特徴点はその画像面上で観測されるべき位置を推定するわけである。この方法では、推定対象の近傍の画像フレームの部分集合と特徴点の部分集合に因子分解法を適用して得られる、画像フレームと特徴点の位置情報の部分推定を使って、隠れて見えない特徴点の画像面での投影位置を最小二乗近似によって求める。ところが、この方法で得られる推定値は、本来の投影条件である中心射影によるものではなく、因子分解法で仮定する投影モデル (すなわち、正射影や疑似中心射影) により投影されたものになってしまう。したがって、場合によっては、推定値が、実際のカメラモデルである中心射影による投影から大きくずれ、誤差が大きくなることがある。このような誤差が大きい推定値を使うことは、全体の形状と運動の復元の精度に影響する。

さらに、因子分解法ばかりでなく、動画像から物体の3次元形状やカメラの運動を推定する方法では、特徴点をすべての画像フレームで対応付けることが基本である。したがって、画像フレームで見え隠れするような特徴点については、その点が隠れて見えなくなるような画像面への投影位置を推定しなければならない。本論文では、特定の画像フレーム上で隠れて見えないような特徴点について、その画像面での投影位置を、実際のカメラモデルである中心射影による投影にできるだけ近くなるように、高い精度で推定する方法を提案する。提案する推定方式により、カメラが大きく運動するような場合や、長時間に渡って撮影された動画像からにおいても、隠れて見えなくなるような特徴点の画像面上での位置を高い精度で推定できるようになるので、特徴点をすべての画像フレームに渡って、追跡することが可能になる。この結果として得られる画像フレーム間での特徴点の対応をもとに、物体の3次元形状やカメラ運動の高精度の復元が可能になる。

以後、第2節では因子分解法で仮定するカメラ投影モデルである、疑似中心射影モデルについて述べる。第3節では、因子分解法の基本アルゴリズムの概要を述べる。第4節では、特定の画像フレーム上で隠れて見えないような特徴点について、その画像面での投影位置を、実際のカメラモデルである中心射影による投影にできるだけ近くなるように、高い精度で推定する方法を考案する。第5節では、動画像と複数の静止画像からの、物体の3次元構造とカメラ運動の推定の実験例を示す。第6節は結びである。

2. 擬似中心射影モデル

因子分解法で仮定するカメラ投影モデルである，擬似中心射影 (paraperspective) モデルについて述べる．このモデルは，中心射影のスケーリング効果 (近くの物体が遠くのものより大きく見える) と位置効果 (画像の端に写っている物体が，投影中心の近くのものとは違った角度で見える) を考慮しながら，正射影モデルのもつ線形性を保持している．擬似中心射影モデルによる物体の画像面への投影は次のようなステップからなる：

- (1) 物体の重心を通り，画像面に平行な平面を定義する (「仮想面」と呼ぶ)．
- (2) カメラ中心と物体の重心を結ぶ直線に沿って，物体上の点を仮想面に投影する．
- (3) 中心射影によって，その仮想面上の点を画像面に投影する．

この操作を具体的な形に表現する．Fig.1のように，ワールド座標系の原点を物体の重心Cに採り，特徴点 p のワールド座標系での3次元座標を s_p とする．フレーム f のカメラ中心の座標を t_f ，画像面の2次元ローカル座標系の基底ベクトルを i_f, j_f (R^3 (これら2つのベクトルは直交し，長さは1)，カメラの光軸方向を $k_f = i_f \times j_f$ とする．フレーム f において，画像面とベクトル k_f の交点 O_f を原点に採り，単位直交ベクトルの組 (i_f, j_f) により，2次元ローカル座標系 $_f = (O_f; i_f, j_f)$ を定義する．点 p の画像面への擬似中心射影の $_f = (O_f; i_f, j_f)$ での座標を u_{fp}, v_{fp} ，また，中心射影の座標を (u_{fp}, v_{fp}) とする．簡単のため，カメラの焦点距離を $l=1$ とする．

詳細な導出は省略するが，擬似中心射影モデルでは， (u_{fp}, v_{fp}) は次のように表現される²⁾．

$$\begin{aligned} u_{fp} &= \mathbf{m}_f \cdot \mathbf{s}_p + x_f \\ v_{fp} &= \mathbf{n}_f \cdot \mathbf{s}_p + y_f \end{aligned} \quad (1)$$

ただし，

$$\begin{aligned} z_f &= (-\mathbf{t}_f) \cdot \mathbf{k}_f, \\ x_f &= \frac{(-\mathbf{t}_f) \cdot \mathbf{i}_f}{z_f}, \quad y_f = \frac{(-\mathbf{t}_f) \cdot \mathbf{j}_f}{z_f} \\ \mathbf{m}_f &= \frac{\mathbf{i}_f - x_f \mathbf{k}_f}{z_f}, \quad \mathbf{n}_f = \frac{\mathbf{j}_f - y_f \mathbf{k}_f}{z_f} \end{aligned} \quad (2)$$

ここで， z_f はカメラ中心から仮想面までの距離 (x_f, y_f) は物体の重心 α (ワールド座標系の原点) の中心射影による画像面への投影となる．さらに，中心射影

$$\begin{aligned} U_{fp} &= \frac{\mathbf{i}_f \cdot (\mathbf{s}_p - \mathbf{t}_f)}{z_{fp}}, \quad V_{fp} = \frac{\mathbf{j}_f \cdot (\mathbf{s}_p - \mathbf{t}_f)}{z_{fp}}, \\ z_{fp} &= \mathbf{k}_f \cdot (\mathbf{s}_p - \mathbf{t}_f) \end{aligned} \quad (3)$$

において (u_{fp}, v_{fp}) を z_f のまわりでテーラー展開することにより，擬似中心射影モデルは，中心射影モデルを

$$|\mathbf{s}_p|^2 / z_f^2 \cong 0 \quad (4)$$

の仮定のもとで近似したもの (すなわち，1次近似) であることを示すことができる．

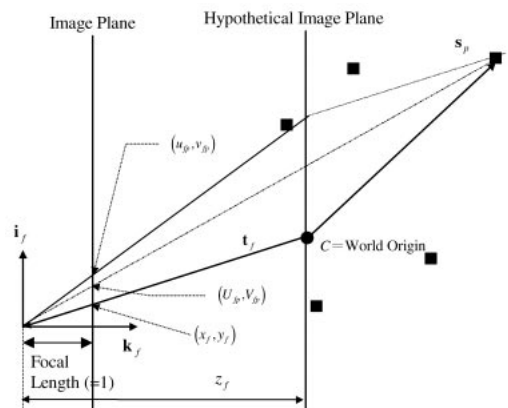


Fig.1 Paraperspective projection model .

3. 因子分解法

本節では，擬似中心射影モデルに基づく因子分解法の基本アルゴリズムの概要を述べる．前提条件として，1個の物体 (剛体) を撮影した，動画像，あるいは，複数枚の静止画像が与えられていると仮定する．動画像の場合には，フレームから自動的に特徴点を抽出して，フレーム間での特徴点の動きを自動的に追跡することにより，画像間での点の対応付けを自動的に行うことができる．一方，静止画像の場合には，一般に，完全の自動的な対応付けは難しいので，人間とのインタラクションにより，画像間での点の対応付けを行う．

3-1 計測行列とその因子分解

物体上の点 p の画像フレーム f での投影座標を (u_{fp}, v_{fp}) とする． P 点の特徴点を， F 枚の画像に渡って追跡した結果として得られる，投影点の画像面上での2次元ローカル座標 (u_{fp}, v_{fp}) ， $p=1, 2, \dots, P$; $f=1, 2, \dots, F$ ，を式 (5) のように並べた行列 W を「追跡行列」と定義する．

$$\mathbf{W} = \begin{bmatrix} u_{11} & \cdots & u_{1P} \\ \vdots & & \vdots \\ u_{F1} & \cdots & u_{FP} \\ v_{11} & \cdots & v_{1P} \\ \vdots & & \vdots \\ v_{F1} & \cdots & v_{FP} \end{bmatrix} \quad (5)$$

追跡行列 \mathbf{W} は $2F \times P$ の行列で、上半分は特徴点の x 座標値 u_{fp} 、下半分は特徴点の y 座標値 v_{fp} を表す。 \mathbf{W} の各列は一つの特徴点に対する追跡結果、各行は単一フレーム内の全特徴点の x 座標値、または、 y 座標値に対応する。

次に、各フレームに対して、全特徴点の x 座標値の平均値 x_f と y 座標値の平均値 y_f を求める。すなわち、

$$\frac{1}{P} \sum_{p=1}^P u_{fp} = x_f, \quad \frac{1}{P} \sum_{p=1}^P v_{fp} = y_f \quad (6)$$

である。そして、 \mathbf{W} の各要素から式(7)のように x_f, y_f を差し引いて行列 \mathbf{W}^* を作る。

$$\mathbf{W}^* = \mathbf{W} - \begin{bmatrix} x_1 \\ \vdots \\ x_f \\ y_1 \\ \vdots \\ y_f \end{bmatrix} \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix} \quad (7)$$

ここで、 \mathbf{W}^* を「計測行列」と定義する。

計測行列 \mathbf{W}^* は、フレームの数 F と特徴点の数 P をいくら増やしても、階数が高々3であるので、以下のように行列を分解できることが示される。

$$\mathbf{W}^*_{(2F \times P)} = \mathbf{R}_{(2F \times 3)} \mathbf{S}_{(3 \times P)} \quad (8)$$

式(1)と比べてみると、 $2F \times 3$ の行列 \mathbf{R} はカメラの姿勢ベクトル $\mathbf{r}_f = (m_f, n_f): f=1, 2, \dots, F$ 、 $3 \times P$ の行列 \mathbf{S} は特徴点の位置ベクトル $\mathbf{s}_p = (x_p, y_p): p=1, 2, \dots, P$ を、それぞれ並べた行列であることがわかる。

実際には計測行列にノイズが含まれるため、行列の階数が3とは限らない。この場合でも特異値分解を用いて、大きな3つの特異値だけを保つように分解すると、二乗誤差の最小化の意味での最適分解が得られる。このような計測行列に対する分解法を「因子分解法 (factorization)」と呼ぶ。

3-2 因子分解法のアルゴリズム

計測行列に対する因子分解のアルゴリズムを述べる。まず、特異値分解を用いて、計測行列 \mathbf{W}^* を以下のように3つの行列の積に分解する。

$$\mathbf{W}^* = \mathbf{U}_{(2F \times P)} \mathbf{\Sigma}_{(P \times P)} \mathbf{V}_{(P \times P)} \quad (9)$$

ただし、 \mathbf{U} は $2F \times P$ の直交行列、 $\mathbf{\Sigma}$ は計測行列の特異値

$$(\sigma_1, \sigma_2, \dots, \sigma_P)$$

から成る $P \times P$ の対角行列、 \mathbf{V} は $P \times P$ の直交行列である。もし計測行列の階数が3であれば、 σ_4 以後は0に近い、小さい値になる。ここで、 σ_4 以後を0と置いて、計測行列を分解してみると、

$$\mathbf{W}^* = \mathbf{U}_{(2F \times 3)} \mathbf{\Sigma}_{(3 \times 3)} \mathbf{V}_{(3 \times P)} \quad (10)$$

となる。そこで、

$$\mathbf{U} = \mathbf{R}, \quad \mathbf{\Sigma} \mathbf{V} = \mathbf{S} \quad (11)$$

とおけば、一つの分解

$$\mathbf{W}^* = \mathbf{R} \mathbf{S} \quad (12)$$

が得られる。しかし、式(12)の分解は一意ではない。実際、任意の正則行列 \mathbf{Q} により、

$$\begin{aligned} \mathbf{W}^* &= \mathbf{R} \mathbf{S} = \mathbf{R} \mathbf{Q} \mathbf{Q}^{-1} \mathbf{S} \\ &= (\mathbf{R} \mathbf{Q}) (\mathbf{Q}^{-1} \mathbf{S}) = \tilde{\mathbf{R}} \tilde{\mathbf{S}} = \tilde{\mathbf{W}}^* \end{aligned} \quad (13)$$

のように、無数の解が存在する。そこで、次のような拘束条件を導入し、拘束条件を満たす \mathbf{Q} を求める。

$$\frac{\|\mathbf{m}_f^T\|^2}{f_x^2 + x_f^2} = \frac{\|\mathbf{n}_f^T\|^2}{f_y^2 + y_f^2} \left(= \frac{1}{z_f^2} \right) \quad (f=1, 2, \dots, F) \quad (14)$$

$$\mathbf{m}_f^T \cdot \mathbf{n}_f = \frac{x_f y_f}{2} \left(\frac{\|\mathbf{m}_f^T\|^2}{f_x^2 + x_f^2} + \frac{\|\mathbf{n}_f^T\|^2}{f_y^2 + y_f^2} \right) \quad (f=1, 2, \dots, F) \quad (15)$$

$$\|\mathbf{m}_1^T\| = 1 \quad (16)$$

そして、 \mathbf{Q} を使って、

$$\mathbf{R} = \hat{\mathbf{R}} \mathbf{Q}, \mathbf{S} = \mathbf{Q}^{-1} \hat{\mathbf{S}} \quad (17)$$

とおけば、

$$\mathbf{W} = \mathbf{R} \mathbf{S} \quad (18)$$

に一意に分解できる。 \mathbf{R} は $2F \times 3$ の行列で、カメラの姿勢を示す。 \mathbf{S} は $3 \times P$ の行列で、各特徴点の3次元座標値を示す。詳細は、Poelman & Kanade²⁾に譲るが、行列 \mathbf{R} 、すなわち $\{(m_f, n_f): f=1, 2, \dots, F\}$ と、式(6)で計算される (x_f, y_f) から、カメラ方向 $\{(i_f, j_f, k_f): f=1, 2, \dots, F\}$ が求められる。次に、式(14)から z_f 、式(2)からカメラの位置 t_f が計算できる。

4. 遮蔽特徴点の投影位置の推定

因子分解法の入力として、まず F 枚のフレームに渡って、 P 点の特徴点を追跡した結果から、 $2F \times P$ の追跡行列を作る必要があるが、オクルージョンや照明変動などの影響で、各特徴点が必ずしも全フレームに渡って追跡できるとは限らない。例えば、Fig.2に示したように、特徴点の追跡が失敗して、位置情報が得られなくなってしまう(「欠測値」と呼ぶ)ことがある。このために、計測行列の要素が欠測値を含む場合には、部分的な特徴点の3次元情報あるいは部分的なカメラの姿勢情報が復元できない。全ての特徴点の3次元情報と全てのフレームにおけるカメラの姿勢情報を求めるには、追跡が不完全な特徴点に対して、何らかの方法で2次元位置情報を推測し、計測行列を補う必要がある。本節では、特定の画像フレーム上で隠れて見えないような特徴点について、その画像面での投影位置を、実際のカメラモデルである中心射影による投影にできるだけ近くなるように、高精度で推定する方法を述べる。

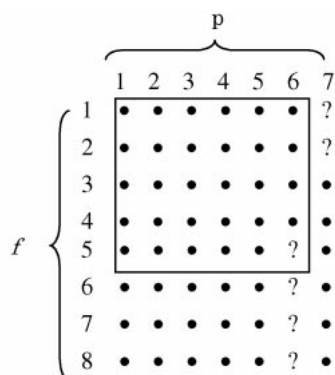


Fig.2 A measurement matrix with absent values showing the result of the tracking for the feature points 1-7 up to the eighth frame. The mark "?" denotes an absent value for which the tracking has failed. The sub-matrix for inferring the element $(f, p) = (5, 6)$ is denoted by the rectangular frame.

4.1 推定行列

因子分解法は冗長な情報を利用する優決定系(over constrained)の方法の一種なので、全特徴点の形状情報、あるいは全フレームにおけるカメラの姿勢情報が分かれば、欠測値の推定が可能である¹⁾。一つの欠測値を推定するには、まず追跡行列からFig.3のように、その欠測値を含む推定行列を作る必要がある。欠測値 W_{fp} 以外、推定行列の各要素が既知なので、欠測値所在の行を抜きにして(Fig.3のA)、全特徴点の形状情報($S_1 \cdots S_p$)を因子分解法で計算できる。同様に、欠測値所在の列を抜きにして(Fig.3のB)、全フレームにおけるカメラの姿勢情報($R_1 \cdots R_f$)が求められる。これらを用いて、欠測値 W_{fp} の推定値は式(1)に代入して計算できる。

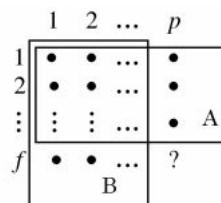


Fig.3 The sub-matrix for inferring the unknown projection position of point p at frame f (W_{fp}).

推定行列を構築するとき、Fig.3に示したように、推定対象要素、すなわち欠測値以外は全て既知であることが必要である。推定行列のサイズについて考えると、大きい行列の方が推定精度が上がる反面、1つの欠測特徴点あたりの計算コストが高くなってしまいます。そこで、まず適当なサイズの推定行列を作って、もし推定が失敗したら、行列のサイズを増やしていくことが妥当である。したがって、効率よく推定行列を構築するために、推定しながら、以下で述べるような整形処理を、追跡行列に施す必要がある。

特徴点を一連のフレームに渡って追跡した結果の一例をFig.4(a)に示す。その中に、白色の部分は追跡失敗、灰色の部分は追跡成功を表している。各特徴点を特徴点位置の存在するフレームによって、Fig.4(b)のように、次のような四つのグループに分けることができる：

- A 全部のフレームに存在する特徴点。
- B 最初のフレームに存在するが、途中で消えた特徴点。
- C 途中で出現するが、暫く経ったら消えた特徴点。
- D 途中で出現して、最後まで存在する特徴点。

Aグループの点を使って、まずBグループとCグループの各欠測値に対して推定を行う。一旦BグループとCグループ推定が終了(必ず全部推定できるとは限らないが)したら、全体の追跡行列を逆さにして、またBグループとCグループの欠測値に対して推定を行う。このようにある予め決めた終了条件まで、追跡行列の欠測値に対する推定を繰り返す。

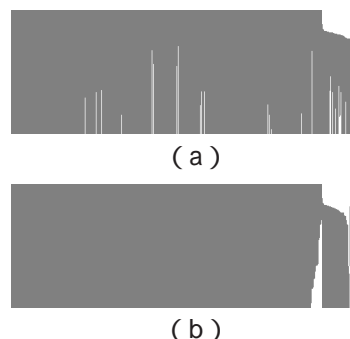


Fig.4 An example of feature tracking result for 509 points over 182 frames (a) and the result of rearranging the feature points according to their existence duration in frames (b)

ここで、推定行列の各計測値が中心射影のモデルによって投影された座標値であるのに対して、欠測値の推定には中心射影の近似モデルである正射影、あるいは疑似中心射影を用いることに注意する。そのために、近似条件が悪い場合には欠測値の推定値の誤差が大きくなる可能性がある。このような大きな推定誤差を含む計測行列を使うと、最後の全体特徴点の形状情報とカメラの姿勢情報の復元精度に影響をもたらす。この問題の解決方法を次に述べる。

4-2 推定行列の構築

疑似中心射影モデルは、中心射影モデルにおける、カメラの光学中心から物体の重心までの距離のまわりで展開し、一次近似を取ったモデルである。すなわち、次の二つの近似条件を仮定した上で、疑似中心射影モデルが成立する。

$$\mathbf{k}_f \cdot \mathbf{s}_p \approx 0 \quad (19)$$

$$\left| \mathbf{s}_p \right|^2 / z_f^2 \approx 0 \quad (20)$$

式(20)は、撮影距離に比べて、物体のサイズが小さいことを仮定して、疑似中心射影モデルの固有誤差である。式(19)には、各特徴点のカメラの光軸への投影量が非常に小さいことを仮定している。もし、各特徴点がモデルの仮定面上にあれば、疑似中心射影が中心射影に一致することが言える。これは、追跡行列から推定行列を構築する際に、式(19)が良い推定行列の構築条件を示唆することを意味している。しかし、仮定面が常にカメラの画像面に平行であるので、カメラ運動が平行運動以外の場合には、各フレームにおけるカメラの姿勢も異なる。特にカメラの運動が激しい場合、式(19)の仮定条件は保ち難いが、ビデオカメラで撮った動画画像の場合、連続するフレーム間のカメラの姿勢変化が小さいため、局所的に集中している特徴点の集合が式(19)を満たすことを仮定できる。

この仮定をもとに、追跡行列から、まず欠測特徴点と最も近い特徴点を選択して推定行列を構築する。しかし、欠測特徴点の位置、すなわち推定対象の欠測値が未知という問題点がある。ここで、連続的に隣り合うフレーム間で、特徴点の運動が平行運動と仮定し、欠測値の代わりに、欠測特徴点の直前のフレームにある位置情報(以下「仮定欠測値」と)と最も近い特徴点を選択して推定行列を構築する。特に動画画像の場合、フレーム間の特徴点の運動が小さいので、この仮定を満たす。

5. 実験結果

3次元情報の復元精度を定量的に解析するために、合成データを使って評価を行う。予め用意した3次元点

群データ468点を中心射影に基づいて、物体サイズの約60倍の撮影距離で、カメラを1度ずつ回転し、60フレーム分の追跡行列を作る。追跡行列の各要素にガウシアンノイズを加えて、特徴点の追跡誤差をシミュレートする。一つのノイズに対して、五つの異なる乱数初期値(seed)を使って計算した結果の平均値を結果とする。

5-1 合成データを用いた復元結果

特徴点の追跡度合いを表すパラメータとして計測率を定義する。追跡行列の各要素が全部計測できた、あるいは追跡成功した場合には、計測率が1で、半分の要素が計測した場合には、計測率が0.5である。提案した推定行列構築手法を検証するために、計測率を1から0.65まで変化させる。

Fig.5に、欠測値の推定により得られた追跡行列の推定誤差(単位:画素)を、要素のRMSE (root-mean-square)により各計測率について示す。計測率が1の各誤差値は、追跡誤差をシミュレートするために与えたガウシアンノイズの標準偏差値に一致する。ノイズの標準偏差が0.1以下の場合には計測率の減少にしたがって追跡行列の推定誤差がわずかに増加する。これに対して、ノイズの標準偏差値が0.5以上の場合には、計測率の減少にしたがって、追跡行列の誤差がかえって減っていく現象が起きた。これは本手法固有の計測値の推定誤差が少なくとも0.5画素以下であることを示している。計測率が0.7以下に下がると追跡行列の誤差が少し跳ね上がるが、これは推定行列内の計測値が元々の欠測値の推定値を含んでいるためであると考えられる。

Fig.6とFig.7は各計測率における推定した追跡行列から抽出した形状誤差と回転誤差である。形状誤差は抽出した点群の3次元データと用意した合成データとの差のRMSで表す。なお合成データを1に正規化した。回転誤差は復元したカメラの姿勢をある回転軸の周りに真の姿勢に合わせるための回転角(単位:ラジアン)である。ノイズの高い追跡行列が、計測率の減少、すなわち欠測値の推定数の増加に伴って全体の3次元情報の復元誤差を減らしていく。

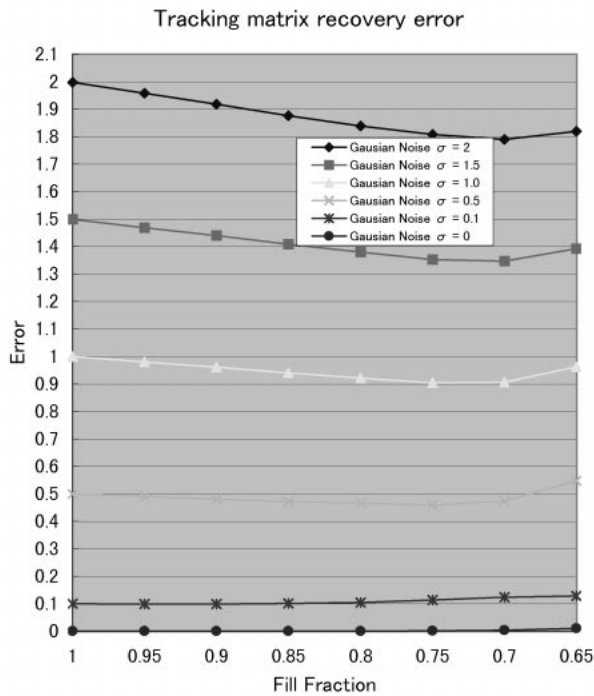


Fig.5 Recovered tracking matrix error according to the noise level.

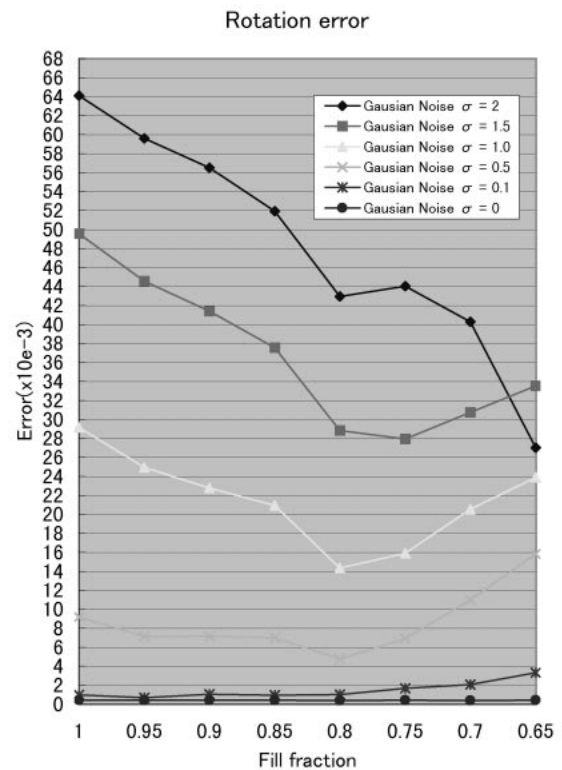


Fig.7 Rotation error according to the noise level.

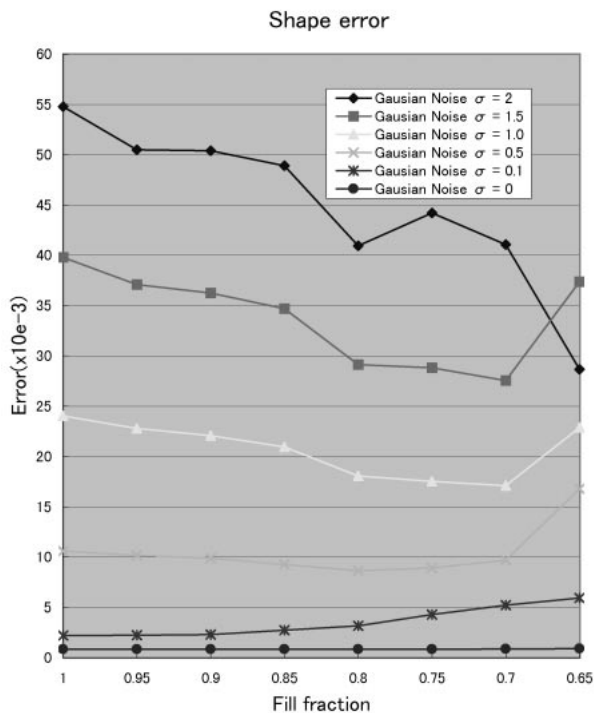


Fig.6 Recovered shape error according to the noise level.

提案した、仮定欠測値と近い計測値を選択して構築した推定行列を検証するために、任意の計測値を選択して構築した推定行列の復元結果をノイズの標準偏差値を1.5に固定した場合に比較した。Fig.8に、追跡行列の推定誤差の比較を示すが、任意計測選択手法に比べて、本手法がノイズを減らすに効果があることが分かる。Fig.9とFig.10は復元した形状と姿勢の誤差の比較であるが、計測率0.7以上の場合、本手法はで形状誤差と回転誤差ともに下回るが、計測率が0.65になると、逆転の現象が起こった。これは推定行列の計測値が、任意計測値選択より、仮定欠測値と近い計測値を選択した方が推定した計測値を選択した確率が高いと考えられる。

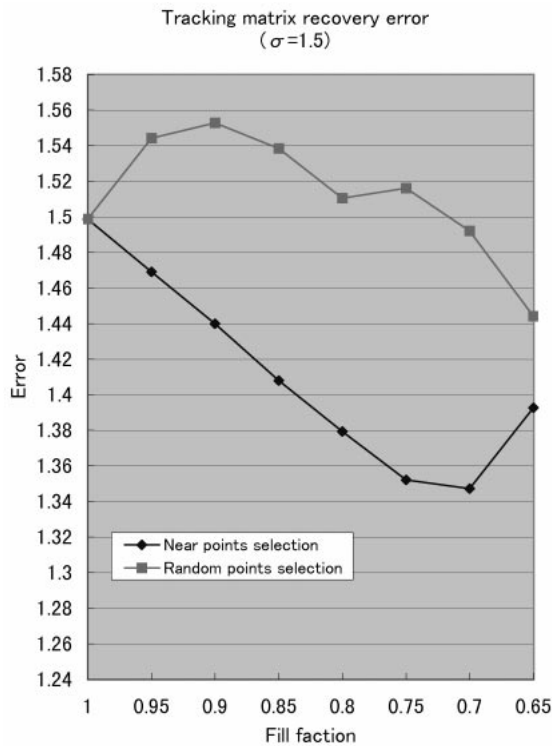


Fig.8 Comparison of methods with respect to the recovered tracking matrix.

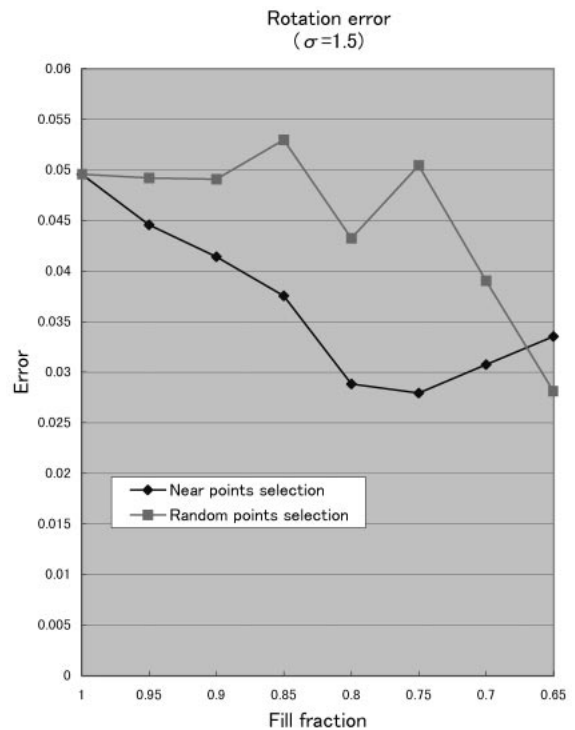


Fig.10 Comparison of methods with respect to the rotation error.

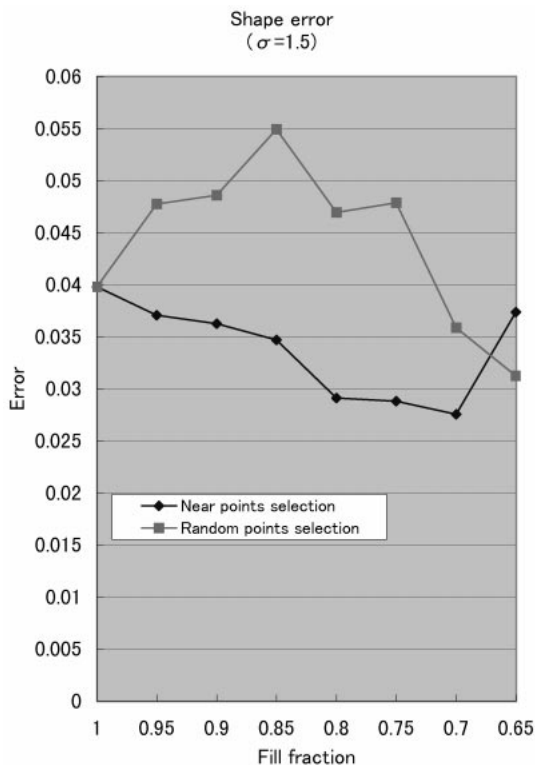


Fig.9 Comparison of methods with respect to the recovered shape error.

5-2 実画像データの抽出結果

実物体を普通の室内照明環境の下に撮影して、その物体3次元情報を抽出してみた。撮影の対象物体として Fig.11(a) に示したモアイ像 (約180×70×50 mm.) を回転テーブルに乗せて、約3メートルの距離で60枚分の動画データを撮った。撮影に使ったカメラはSONY製民生用デジタルビデオカメラ (DCR-VX9000) である。撮影した動画データに対し、KLT³⁾ という特徴点追跡法で約3000点の特徴点を追跡して、追跡行列を作る。追跡行列内要素の計測率は約0.94である。Fig.11(b) には、一枚目のフレームにおける特徴点 (灰色部分) 抽出の様子を示す。復元した特徴点の3次元データを二つの方向から見た様子を Fig.11(c) と Fig.11(d) に示した。

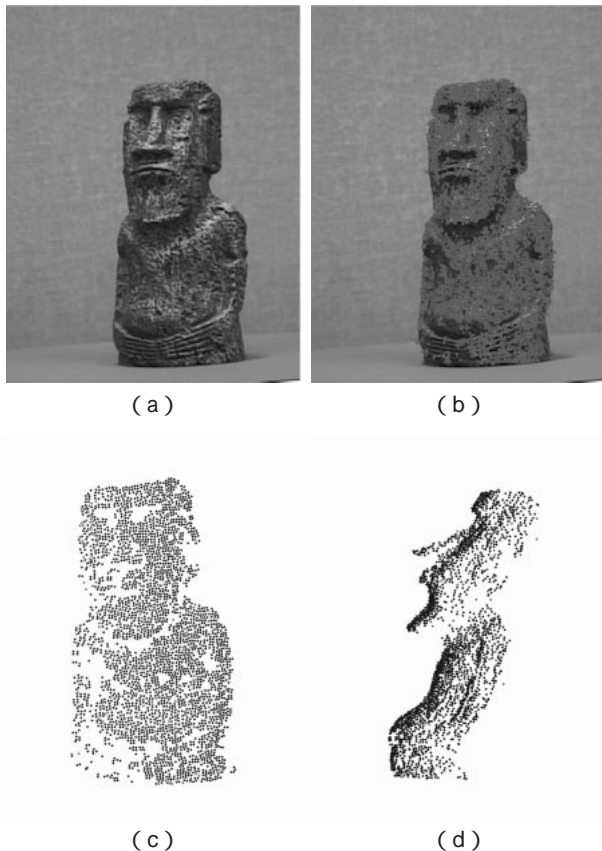


Fig.11 First frame of MOAI image sequence(a) feature selection results(b) and two views of recovered shape (c , d)

動画像に比べて、幾つかの方向から撮影した静止画の場合には、完全な自動追跡が難しいので、人間とのインタラクションにより、画像間の特徴点の対応付けを行う必要がある。Fig.12 (a)~(c)に示したのは、(株)リコーのデジタルカメラDC-1を用いて18度ずつ回転しながらPHS電話機を撮影した20枚画像の中の3枚である。特徴点の数は40で、復元した形状情報に基づいて、76個のポリゴンを生成し、テクスチャーマッピングした結果をFig.12 (d)に示した。

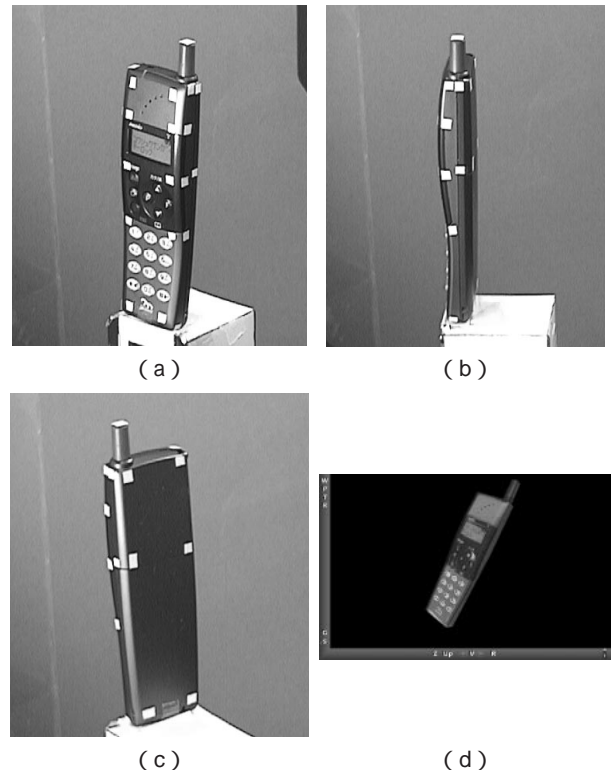


Fig.12 Three frames of PHS image(a, b, c), and its recovered shape with texture mapped(d).

6. 結び

マルチメディア情報の重要な特徴は、すべての情報がデジタルで表されていることであり、従来のアナログ情報と比べて、編集・加工・処理などが容易にできるという大きな利点を持っている。そして、入力機器も、デジタルカメラをはじめ、一般のユーザでも簡単に手に入るため、デジタル情報の再利用も重要になってくる。本稿で説明した多視点画像からの3次元情報抽出も、このようなデジタルデータの応用の可能性を示している。

参考文献

- 1) C.Tomasi and T.Kanade, Shape and motion from image streams under orthography: A factorization method, International Journal of Computer Vision, vol.9, 1992, pp.137-154.
- 2) C.J.Poelman and T.Kanade, A paraperspective factorization method for shape and motion recovery, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.19, no.3, 1997, pp.206-218.
- 3) B.D.Lucas and T.Kanade, An iterative image registration technique with an application to stereo vision, Proceedings of Seventh International Conference on Artificial Intelligence, 1981.