

3次元向きセンサを取付けたカメラを用いた 投票によるカメラの並進運動の推定

岡谷 貴之 出口 光一郎
東北大学大学院情報科学研究科

概要: カメラを自由に動かして撮影したシーンや物体の2枚の画像に対して、画像内の点の対応が未知の場合に、カメラの位置と姿勢を推定する問題について述べる。カメラの姿勢は、カメラに3次元向きセンサを取付けてそこから得る。カメラの内部パラメータを既知とし、カメラの位置すなわち視点間の並進運動だけを、画像とセンサから得た姿勢情報から推定する。このカメラの並進は、対応の可能性がある2点の組を投票することにより推定される。これは点の対応関係を陽に決めずに行われる。この種の問題には、従来よりランダムサンプリングに基づく手法が適用され、その有効性が知られている。提案する投票による推定方法と頑健さについて比較する。

Estimating camera translation based on a voting method using a camera with a 3D orientation sensor

Takayuki Okatani and Koichiro Deguchi
Graduate School of Information Sciences, Tohoku University

Abstract: This paper is concerned with the problem of estimating pose of a camera in space from a pair of images of a scene or an object that is taken with the camera in a case where the point correspondences between the images are unknown. The camera orientation is obtained by a 3D orientation sensor rigidly attached to the camera. Assuming that the intrinsic parameters of the camera are known, we estimate only the camera position, i.e. translation between viewpoints, from the images and the camera orientation obtained from the sensor. This camera translation is estimated by voting pairs of points that likely correspond to each other. This is done without explicitly determining the point correspondences. Several methods using random samples have been applied to these kinds of problems and their effectiveness has been already known. We compare these methods with our voting-based method in terms of robustness.

1 はじめに

自由に持ち運べるカメラを用いて物体やシーンを撮影し、画像からその3次元形状を復元することはCVの中心的テーマである。structure and motion (SAM) ではもちろんのこと、その他の様々な3次元モデル構築や画像合成のための手法、例えば space carving や light field rendering、あるいは shape from silhouette などにおいて、正確なカメラの位置と姿勢の推定は不可欠である。本報告では、カメラに3次元向きセンサを取り付け、カメラの姿勢はセンサから得て、姿勢情報とそれぞれの画像から、空間における位置を計算する手法について述べる。2枚の画像間の特徴点の対応関係をロバストに定めつつ位置を

推定することがテーマである。

自由にカメラを動かして物体やシーンを撮影した画像の間で、特徴点を正しく対応付けることは難しい問題である。ビデオカメラで連続的に撮影した画像系列を対象とし、カメラの運動が滑らかで画像の差異が少ない場合 [1, 2, 3] と、離散的に撮影された2枚以上の画像を対象とし、画像間で見え方が大きく変わるような場合 [4, 5, 6] とで、問題の性格は変わる。本研究は、見え方が大きく変わる場合を対象とし、その場合に特徴点の対応と画像間の関係を正しく推定することを目指す。

これまでに提案されている方法 [4, 5] は主にランダムサンプリングに基づく方法である。その基本的な考え方は、次の通りである。まず、それ

ぞれの画像において別々に抽出した特徴点について、そのまわり濃淡に関する相関などを用いて、画像をまたいだ特徴点間の類似度を算出し対応付ける。このようにして得た対応には、考えられる画像の誤差の範囲を大きくそれら対応の誤り（外れ値、outlierと呼ばれる）が多数含まれる。これに対処するため、特徴点の対応の組からランダムに少ない組み合わせを選び、そこから2枚の画像間に成り立つエピポラ条件を求める。求めたエピポラ条件が、残りの対応の組をどれだけ矛盾無く説明するかを調べる。ランダムにサンプルを選び直してこれを多数回繰り返す、最適なものを解とする。

この方法はある程度はうまく働くが、それほど悪条件と思えない画像を対象にしても、うまくいかないことがある（後で示す図12参照）。例えば対象の点が平面上に近くなってしまふとき（退化の問題）などいくつかの要因があり、その対策も考えられてきているが、本質的には、ランダムサンプルの性質—少ないサンプルから自由度が多いエピポラ条件を推定する—に、困難さがあると思われる。

本研究では、カメラを校正済とし、3次元向きセンサをカメラに取付けて回転情報を得ることで、推定すべき画像間の関係を並進だけとした。高い自由度をもつ画像間の関係を、画像自体からすべて決めることをあきらめれば、困難さが軽減される可能性がある。ここで、カメラは校正済つまり内部パラメータは既知とするが、画像中心や skewness は予め定めておくとしても、焦点距離については、最近のデジタルカメラでは画像データとともにその情報が得られるものが多く、それを利用する。これにより撮影中も焦点距離を変化させることができる。向きセンサを使うことと同じく、画像以外から得られるものは最大限使う立場をとる。

3次元向きセンサは、空間における向きをロール、ピッチ、ヨーなどの角度で出力するセンサである。加速度センサ、ジャイロセンサや磁気センサなどの組み合わせで構成される。このような向きセンサをカメラに取付けても、カメラの可搬性は何ら損なわれず、通常のカメラを使う場合と全く同じように撮影できる。

3次元向きセンサをカメラに取付けてその姿勢情報を得、様々な問題を解決しようとする試みは従来より存在する。ビデオカメラにおいて手振れなどの振動を除去し安定した画像を得ようとするものは言うに及ばず、因子分解法にお

いて、形状とカメラの運動の不定性をカメラに向きセンサを取付けて取り除く方法などがある。また、特徴点の対応付け問題を、姿勢を知ることである程度自動化しようとしたものもある。ただしこれは、本報告で議論しているようなロバストな手法ではなかった。また、SAMにおいて、カメラの並進と回転の分離が不安定であるとの認識から、ビデオカメラにジャイロセンサを取付けて並進のみを推定すべきパラメータとした[7]がある。そこでは特徴点の対応（追跡）は済んでいるものとしていた。

このような既存の方法に対し、ここでは、画像間で推定すべき関係が並進だけであることを最大限利用して、特徴点の対応付けをよりロバストに行うことを考える。この場合、推定すべきパラメータの自由度はたったの2である。このように自由度が減ることだけでも、対応付けの問題は容易になるかもしれない。しかし、並進の方向しか自由度がなくともカメラの配置には（当然だが）無数の可能性があり、特徴点の対応を求めることに関しては、残念ながら単純にはメリットを得られない。全パラメータを画像から推定するのと同様にランダムサンプリングの手法で並進を定めると、向きセンサを付けなかった場合に比べて、計算量が減る程度の効果しかない。

そこで本研究では、自由度が2と少ないことを利用し、投票を用いることを考えた。一般に推定するパラメータの自由度が3以上になると、投票という枠組みを用いてパラメータの推定を行うことは困難であるが、今の場合は2しか自由度がなく好都合である。具体的には、特徴点の画像間における対応の候補を類似度を用いてある程度まで限定した後、これを投票に利用し、集積点として並進パラメータを得ようとするものである。その結果、特徴点の対応付けは陽には行われることなく、並進が推定される。このことからいくつかの利点が生まれる。

2節で画像間の関係について簡単にまとめ、3節でランダムサンプリングに基づく対応付けの手法について述べ、4節で投票を用いる方法を述べる。5節で実験結果を示し、最後にまとめる。

2 画像間の関係と特徴点の対応

2.1 画像間の関係

カメラの位置と姿勢を変えて同一シーンの画像を2枚取得したとする。シーンの空間にお

るある 1 点の座標を，それぞれの画像を取得したカメラの座標系で \mathbf{X}, \mathbf{X}' と表す．さらにその点が画像上に占める座標を \mathbf{x}, \mathbf{x}' とする．これらは齊次座標で表し， $\mathbf{X} = [x, y, z, 1]^T$ ， $\mathbf{x} = [u, v, 1]^T$ とする． \mathbf{x} と \mathbf{x}' の間には

$$\mathbf{x}'^T \mathbf{F} \mathbf{x} = 0 \quad (1)$$

の関係がある．カメラが未校正の場合には，この基礎行列 \mathbf{F} を推定しつつ特徴点の正しい対応を探すことが問題となる． \mathbf{F} は，最低 7 点の対応が与えられ，点の空間における分布が平面上にないことや，カメラの運動が回転のみでないなど，退化していなければ，3 つの不定性を残して決定できる．また 8 点あれば線形計算により一意に決定できる．

カメラの座標系間の回転，並進を \mathbf{R}, \mathbf{t} とし，カメラの内部パラメータを保持した行列を \mathbf{A} とすると， $\mathbf{x} \propto \mathbf{A}[\mathbf{R} | \mathbf{t}]\mathbf{X}$ と書ける．それぞれの画像におけるカメラの行列を \mathbf{A}, \mathbf{A}' とすると，

$$\mathbf{F} = \mathbf{A}'^{-T} \mathbf{T}_\times \mathbf{R} \mathbf{A}^{-1}$$

の関係がある． \mathbf{T}_\times は，任意の 3 次元ベクトル \mathbf{p} に対して $\mathbf{T}_\times \mathbf{p} = \mathbf{t} \times \mathbf{p}$ となるような 3×3 行列である．

カメラの内部パラメータ \mathbf{A} が既知で，さらに座標系間の回転 \mathbf{R} も既知であれば，並進 \mathbf{t} は次のようにして簡単に計算できる．(1) 式を書き換え， \mathbf{t} についてまとめると

$$\mathbf{a}(\mathbf{x}, \mathbf{x}'; \mathbf{A}, \mathbf{A}', \mathbf{R})^T \mathbf{t} = 0 \quad (2)$$

の形の式を得る．単眼視におけるスケールの不定性より並進 \mathbf{t} の大きさは原理的に不定であり，その意味で自由度は 2 しかない．そこで $|\mathbf{t}| = 1$ を拘束とすることで，2 点以上の点の対応により \mathbf{t} を決定できる．

2.2 特徴点の対応付け

特徴点はコーナー検出を行って抽出する [8]．2 枚の画像に対し別々に行い点を決める．2 枚の画像間において抽出した特徴点は，点まわりの画像の濃淡の類似度や画像に占める位置の近さなどを基準に，対応の候補を決める．類似度は局所的な相関や，回転に不変な微分特徴量 (local jets)[9] などを用いて行う．これらを使って，1 枚目の特徴点のそれぞれに対し，最も類似度の高い 2 枚目の特徴点を唯一，あるいは複数個選ぶことができる．

1. あらかじめ定めた回数だけ次を行う．
 - 1.1 対応の組から n 個のサンプルをランダムに選ぶ．
 - 1.2 選んだサンプルを元にパラメータ ($\mathbf{F}, \mathbf{R}, \mathbf{t}$ など) を推定．
 - 1.3 推定したエピソード条件が，その他の対応の組と矛盾がないかを調べる．LMS では，各組のエピソード条件からの誤差を求めその中央値を評価に用い，RANSAC では各組のうち正しいもの (inlier) の数を評価に用いる．
2. データに最も矛盾無く適合するサンプルと，そのときのパラメータを選んで終了．

図 1: ランダムサンプリングによる手法の流れ

上にも述べたように，類似度の他に，特徴点が画像に占める位置の近さも考慮することができる．[5] や [4] ではそれを利用している．しかし，カメラを大きく動かすことを想定した場合，制限するのは得策ではない．また [4] では，対応候補の決定後，緩和法を利用して 1 枚目と 2 枚目の画像の特徴点同士を 1 対 1 に対応させている．これも有効な場合はあるが万能ではないし，かえって頑健さを損なう方向に働くこともあると思われる．

3 ランダムサンプリング

前節のようにして，1 枚目の画像の 1 つの特徴点に対し 2 枚目の画像の特徴点を 1 つ選んだとする．この場合，画像の類似度だけが頼りにしたので，この特徴点の対の集合には多くの誤りが含まれる．正しい対応については，2 枚の画像間に成り立つエピソード条件を満すから，これを正誤判定の手がかりとできる．least median of squares (LMS) を用いる [4] と，random sample consensus (RANSAC) を用いる [5] などが代表的である．これらの手法についてまとめる．

3.1 基本的な考え方

通常，カメラが未校正の場合が対象とされる．[5] では \mathbf{F} を求めるのに最低必要な 7 点を，[4] では線形アルゴリズムが使える 8 点をランダムに選んでいる．ランダムに選んだ点を元に \mathbf{F} を計算し，残りの対応点の組について計算された \mathbf{F} が妥当であるかどうかを調べる．これを一度の試

表 1: 誤対応が多い場合に求められるランダムサンプルの総数

| p | Proportion of outliers, ε | | | |
|-----|---------------------------------------|-------------------|-------------------|----------------------|
| | 0.5 | 0.8 | 0.9 | 0.95 |
| 2 | 11 | 74 | 3.0×10^2 | 1.2×10^3 |
| 5 | 94 | 9.3×10^3 | 3.0×10^5 | 9.6×10^6 |
| 7 | 3.8×10^2 | 2.3×10^5 | 3.0×10^7 | 3.8×10^9 |
| 8 | 7.7×10^2 | 1.2×10^6 | 3.0×10^8 | 7.7×10^{10} |

1. 投票空間の範囲を決め、初期化する.
2. すべての対応候補の組について以下を計算する.
 - 2.1 現在の組について \mathbf{a} を求める.
 - 2.2 $\mathbf{a}^T \mathbf{t} = 0$ を満たす \mathbf{t} に投票する. つまり投票空間に直線を加算的に引く.
3. 投票空間のピークを検出し、並進 \mathbf{t} を復元する.

図 2: 投票による並進の推定

行として、何度も特徴点をランダムに選び直す。得られている特徴点の組を最もよく説明する \mathbf{F} が、最終的に望むものとして決定される。この判断にはいくつかのやり方がある。RANSAC では、エピソード条件を満たす「正しい」対応 (inlier) の数が最も多くなるようなものを解とする。LMS では、すべての対応の組にわたる、各組のエピソード条件に対する「誤差」の中央値を、最も小さくするようなものを解とする。アルゴリズムを図 1 にまとめる。

何回の試行を行えばよいか、つまり何回のランダムなサンプルを行うと正しい組を少なくとも 1 回は引き当てることができるかは、次の式によって見積もれる。1 組 p 対からなるサンプルを m 組選んだとき、その中に 1 組以上すべて正しい対応を与えるものが入っている確率を γ とすると、これは

$$\gamma = (1 - (1 - (1 - \varepsilon)^p)^m) \quad (3)$$

で与えられる。 ε は対応の組に含まれる誤対応の割合である。したがって、 $\gamma = 0.95$ とすると、 $p = 8$ で、データの半分が誤対応 ($\varepsilon = 0.5$) であったとき、800 組程度のランダムサンプルを必要とする。

3.2 誤対応が非常に多い場合

ほとんど誤対応というような場合には、取るべきランダムサンプルの数は急激に増加する。いくつかの例を表 1 に示す。表中 $\varepsilon = 0.95$ は、特徴点の組のわずか 5% しか正しいものがない場合を指す。このような例があり得ない極端な例かという決してそうではない。画像間のカメラの移動量を大きくし画像の差異が大きくなるにつれより高い頻度で誤対応が起きてくる。そのような場合、 \mathbf{F} (7 組の対応を要する) はもちろん \mathbf{R} と \mathbf{t} (5 組を要する) を推定する場合で

も、選ぶべきサンプルの数は大変多くなる。一方で、推定すべきが並進 \mathbf{t} だけであれば 2 組あればよく、このような厳しい場合でも十分実行可能な領域にあることが表 1 から分かる。そして実際に、それほど誤った対応があっても、ランダムサンプリングによってある程度までうまく推定できる (これは 5 節で示す)。

このように、推定すべきが並進だけであることで、条件がかなり悪い (ほとんどが誤対応である) 場合でも、計算量の点から見て十分実行可能な領域に問題を落とすことができることが分かる。しかしながら、やはり少ないデータから推定したパラメータを使って全体をよく説明しようとするところから、5 節に示すように、その性能は限定的である。

4 投票による並進の推定

推定すべきが並進だけのときは、ランダムサンプリングでなく、投票を用いることが選択肢になりうる。まず原理について述べ、ランダムサンプリングに比した利点を述べる。

4.1 並進の推定

回転行列が与えられているとき、特徴点の組が 1 つ与えられると (2) 式のように \mathbf{t} の拘束を 1 つ得る。

$$\mathbf{a}(\mathbf{x}, \mathbf{x}'; \mathbf{A}, \mathbf{A}', \mathbf{R})^T \mathbf{t} = 0$$

並進の大きさはどのみち不定なので \mathbf{t} の大きさには意味がない。そこで例えば $\mathbf{t} = [1, t_2, t_3]^T$ と置くと、上の式は $t_2 t_3$ 平面上の直線を表す。特徴点の組が複数あるとき、 $t_2 t_3$ 平面にこの直線を描いて行くと、特徴点の組が正しければ、並進の真の値を座標とする点 (t_2, t_3) で、すべての直線は交差するはずである。誤った対応が含まれているとき、 $t_2 t_3$ 空間に投票を行えば、真の値

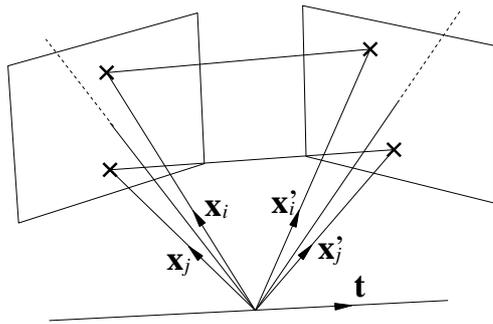


図 3: 両カメラの投影中心が一致するように平行移動させると、特徴点の 1 対が規定するエピポーラ面の交線が並進（方向）になる。

(t_2, t_3) 付近で、最も高い投票数を得ることが期待できる。並進の推定アルゴリズムを図 2 にまとめる。

4.2 幾何学的解釈

上で述べた投票の過程を、幾何学的に解釈すると次のように考えることができる。今、カメラ間の回転が既知であり、内部パラメータも既知である。図 3 のように、両カメラの投影中心が一致するように平行移動させる。このとき、任意の空間の点に対するエピポーラ面は、一致させた投影中心とさらにその点の両画像上への投影点を通る。したがって、空間の任意の点について、画像の位置が分かれば（対応が分かれば）エピポーラ面が決まる。推定すべき並進（の方向）は、このような任意の空間の点に対するエピポーラ面同士の交線として決まる。対応が 1 つ与えられるごとにエピポーラ面を作る作業が投票に相当し、その交線を集積点として選ぶ作業が並進パラメータの推定ということになる。

4.3 点の対応について

ランダムサンプリングでは点の対応は固定され、対応の正誤を問題にした。つまり、1 つの点には 1 つの点だけが対応し、ランダムサンプリング以前の段階でその対応関係は固定され、別の点に対応する可能性を考えることはできなかった。

これに対し、投票を用いると、1 つの点に、可能性を有する複数の点に対応させて、パラメータの推定に有効に利用できる。ある点に対応する点を探すとき、対応点を 1 つに決めるのは難しいが、一定の大きさの集合を与えてその中の 1 つには当りがあるようにするのはずっと容易

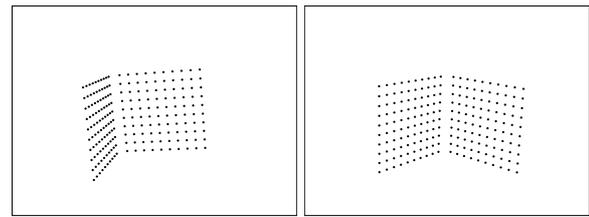


図 4: シミュレーションに用いた画像

である。

この性質によって、より条件の厳しい、画像の差異が大きい 2 枚の画像をうまく処理できる可能性が生まれる。カメラの移動が大きくなると、単純な相関だけでは、正しい対応を 1 つに限定して選ぶことがより難しくなってくる。このような場合に、複数の対応を使って投票を行えば効果を上げることが期待される。

さらに次のような可能性も生まれる。カメラを大きく動かすとき、単純な相関ではなく、回転に不変な相関や微分不変量（local jets）などを用いないと、類似度を正しく判定できなくなると思われる。だがこのような自由度の高い類似尺度を用いると、1 つの点に 1 つだけ決めるやり方では、画像の変化が少ない容易な場合でも、かえって誤対応を増やすことになる。複数の対応を許し、その中に正しい対応があればよいという風に考えれば、このような問題は解決される。

5 実験

5.1 合成画像による実験

図 4 のように、2 つの直交する平面上に格子状に分布させた 200 の特徴点に対し、カメラの位置と姿勢と変化させたときの画像を 2 枚合成した。各特徴点の画像座標には、同じ一様分布に従う誤差 $[-\sigma/2 : \sigma/2]$ を加え、対応をランダムに作成し、しかし常に誤対応の割合が 95% ($\epsilon = 0.95$) となるようにした。

このようなデータの 1 つに対する投票の結果を図 5 に示す。検出したピークから並進ベクトル t を作り、それが真の並進データと空間においてなす角度で精度を評価することにする。100 回、画像座標の誤差と特徴点の対応をランダムに作り変え、計算を行ってこの誤差の分布を求めた。これを図 6 に示す。

比較のために、RANSAC により並進パラメータを推定した。この場合の誤差分布を図 7 に示

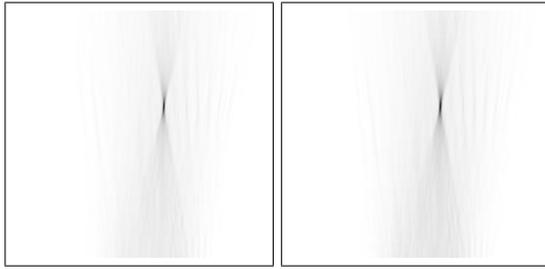


図 5: 図 4 に対する投票後の投票空間 (t_2t_3 平面) .
左 : $\sigma = 0.5$, 右 : $\sigma = 2.0$

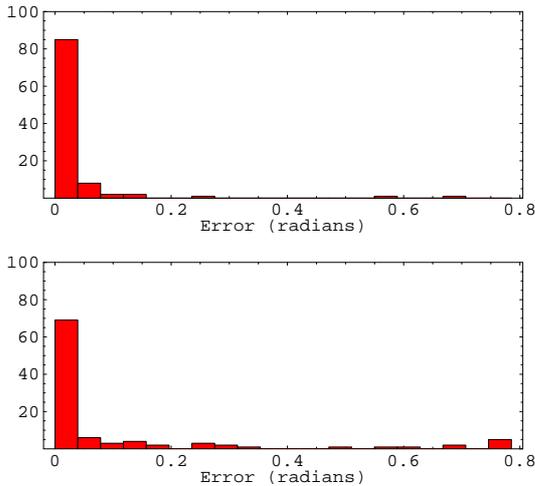


図 6: 投票による推定の誤差分布 . 上 : $\sigma = 0.5$,
下 : $\sigma = 2.0$

す . サンプル数は , 表 1 で見たように必要とされる 1200 を十分超える 5000 とした . また , 点がエピポーラ条件を満すかどうかの判定は , 誤差の存在を考慮して , エピポーラ線に対する距離が 1.0 未満であれば正しく , そうでなければ外れ値とすることで行った . 図 7 では , 推定値がまったく外れてしまっている場合が $\sigma = 0.5$ で 10 程度 , $\sigma = 2.0$ で 40 以上あるが , 図 6 ではそのようなことはなく , この結果から , 図 6 の投票の結果の方が確実に良いことが分かる .

前節で述べたように , 特徴点 1 つに複数の特徴点を関連付け , これを投票することができる . その例として , 上と同じデータで , 1 枚目の画像の特徴点 1 つに 2 枚目の画像の 20 個の特徴点を対応させて , これをすべて投票空間に投票した . 20 個のうち 1 個は常に正しい対応とした . これは , 各点に 1 つの点だけを対応付けるときに 95% の誤対応があるのに相当する . この場合の 100 回の試行による誤差の分布を図 8 に示す . 図 6 よりも良くなっていることが分かる .

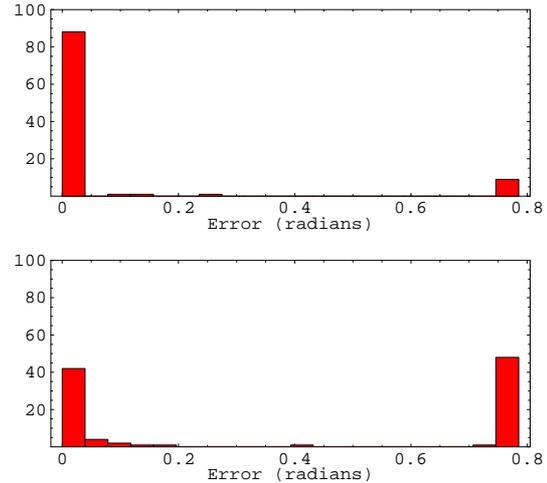


図 7: RANSAC による推定を行った場合の誤差分布 . 上 : $\sigma = 0.5$, 下 : $\sigma = 2.0$

5.2 実画像を用いた実験

実際に向きセンサ (データテック GU-3013) をカメラ (ニコン D1) に取付けて画像を撮影する実験を行った . 画像は元画像を 1/4 に縮小し , 500×328 画素とした .

センサをカメラに取付けただけでは , センサから得られる情報はカメラ座標に関する姿勢の情報ではない . カメラ座標系とセンサ自身の座標系との間の座標変換を知る必要がある . ビジュアルサーボなどで , カメラをコンピュータの先端に取付けたとき , カメラとハンドの座標系間の関係を定めるハンドアイキャリブレーションの研究があり , これを応用できる . ここでは , 基本的な手法 [10] を用いた . [11] を用いて , カメラの内部パラメータの推定と同時にセンサ-カメラ間のキャリブレーションを行うのに必要な外部パラメータの推定を行った .

実画像を用いて投票を行った結果を図 9 , 11 , 10 に示す . 対応の候補は , 相関を用いて 1 枚目の画像の特徴点 1 つにつき最大 20 個まで選び , 相関の値が 0.7 以上のものをすべて投票した . 投票空間 (t_2t_3 平面) の様子を図 9 に示す . 図 10 , 11 は抽出した特徴点を画像に重ね , さらに図 9 に示す投票空間から決めた並進 t をもとに , いくつかの代表的な点についてエピポーラ線を引いた様子である . 1 枚目の画像上に数字で示した点について , 2 枚目の画像上で対応するエピポーラ線が正しい点の位置を通過していることが分かる . 比較のために図 12 に , 未校正なカメラで撮影した回転と並進ともに未知の画像対を

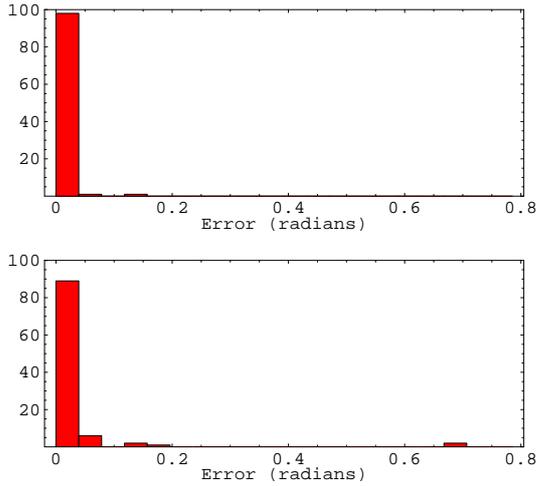


図 8: 1 対 20 の特徴点の対応を利用して投票した場合の誤差分布．上： $\sigma = 0.5$ ，下： $\sigma = 2.0$

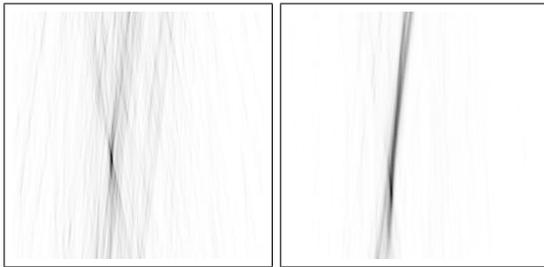


図 9: 図 10, 11 の画像に対する投票結果．右が図 10 で左が図 11 にのもの．

対象に対応点を求める [11] の手法を適用した結果を示す．図 10, 11 の 1 枚目の画像上に数字で示した点について，それに対応するエピソード線を表示してあるが，画像上通るべき点を通っていない．主に対象が平面上であること，図の特に背景部分が若干異なることが原因と思われる．

6 まとめと議論

本報告では，内部パラメータ既知のカメラに 3 次元向きセンサを取付け，撮影した 2 枚の画像において，未知の特徴点の対応を調べながら，投票によって並進を推定する方法を述べた．推定すべきが並進の 2 自由度しかないことにより，投票という枠組みを用いることができる．計算機シミュレーションによってランダムサンプリングと投票の比較を行い，投票による手法がよりロバストであることを確かめた．また，実画像を用いた実験により有効性を確かめた．



図 10: 実画像に対する結果．上：1 枚目の画像と抽出した特徴点．下：2 枚目の画像と特徴点，さらに投票によって求めた並進からエピソード線を求めて，それらを重ねて表示したもの．数字は対応する点とエピソード線を表す．

投票による長所には次のようなものがある．

1 つは，精度が向上する可能性である．ランダムサンプリングでは，少ないサンプルでパラメータを推定し，できるだけ全体を説明するようなものを選ぶという戦略をとっている．パラメータを推定するのに少ないサンプルしか利用しないことはこの種の手法の思想であるが，同時に問題を生じる．誤対応が非常に多く，特徴点の位置にも誤差が含まれる状況では正しいサンプルが生き残るとは限らない．投票によれば，正しいデータは正確なパラメータの位置近くに投票されるから，よりロバストにパラメータを推定できる可能性が高いと考える．

また，述べたように，1 つの点当り複数の点に対応させて投票に利用できることも利点である．より自由度が高い，回転に不変な相関などを用いることが可能となる．その他にも計算量の少なさも利点となる．ランダムサンプリングは，予め知りようのない誤対応の割合に応じて，必要なサンプル数が変動するやっかいな性質があるから，いきおい十分大きな数のサンプルを取らざるを得ず，計算量が多くなってしまふ．

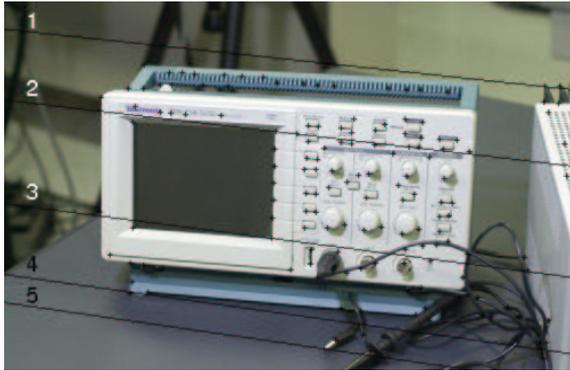
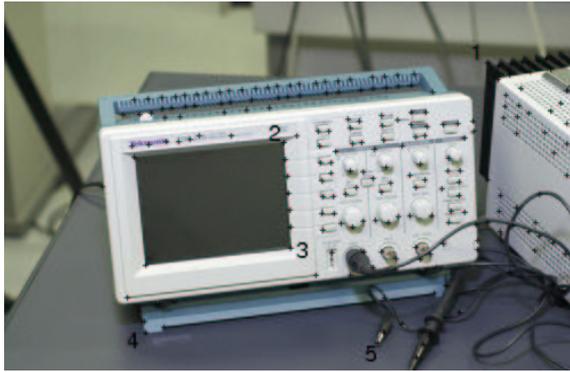


図 11: 実画像に対する別の結果．上：1 枚目の画像と抽出した特徴点．下：2 枚目の画像と特徴点と推定されたエピポーラ線を重ねて表示したもの．数字は対応する点とエピポーラ線を表す．

今後、特徴点の対応候補を定めるのに用いる類似尺度について、様々な可能性を調べることにしている．どの程度ロバスト性が向上させられるかについて、より詳細に調べて行きたい．

参考文献

- [1] P. Beardsley, P. Torr, and A. Zisserman. 3d model acquisition from extended image sequences. In *Proceedings of European Conference on Computer Vision*, volume 2, pages 683–695, 1996.
- [2] P. H. S. Torr, A. W. Fitzgibbon, and A. Zisserman. The problem of degeneracy in structure and motion recovery from uncalibrated image sequences. *International Journal of Computer Vision*, 32(1):27–44, 1999.
- [3] M. Pollefeys, R. Koch, and L. V. Gool. Self-calibration and metric reconstruction inspite of varying and unknown intrinsic camera parameters. *International Journal of Computer Vision*, 32(1):7–25, 1999.
- [4] Z. Zhang, R. Deriche, O. Faugeras, and Q. T. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artificial Intelligence*, 78:87–119, 1994.

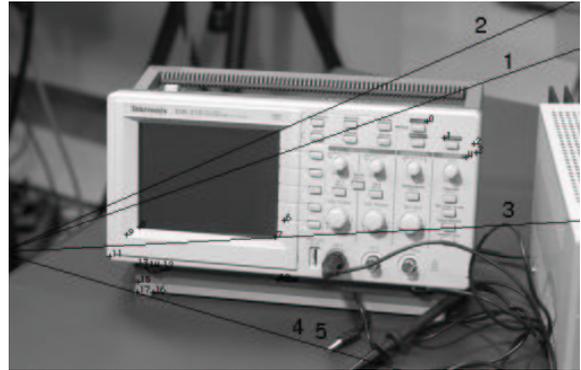


図 12: 図 10, 11 の画像に対して [4] によって F を推定し, 図 10, 11 と同じ点のエピポーラ線を重ねたもの (数字は点との対応を示す) .

- [5] P. H. S. Torr and D. W. Murray. The development and comparison of robust methods for estimating the fundamental matrix. *International Journal of Computer Vision*, 24(3):271–300, 1997.
- [6] P. H. S. Torr and C. Davidson. Impsac: Synthesis of importance sampling and random sample consensus. In *Proceedings of European Conference on Computer Vision*, 2000.
- [7] T. Mukai and N. Ohnishi. The recovery of object shape and camera motion using a sensing system with a video camera and a gyro sensor. In *Proceedings of IEEE International Conference on Computer Vision*, 1999.
- [8] C. Harris and M. Stephens. A combined corner and edge detector. In *Proceedings Alvey Conference*, pages 189–192, 1988.
- [9] C. Schmid and R. Mohr. Local grayvalue invariants for image retrieval. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 19(5):530–535, 1997.
- [10] R. Y. Tsai and R. K. Lenz. A new technique for fully autonomous and efficient 3d robotics hand/eye calibration. *IEEE Transactions on Robotics and Automation*, 5(3):345–358, 1989.
- [11] Z. Zhang. A flexible new technique for camera calibration. Technical Report MSR-TR-98-71, Microsoft Research, 1998.