

# 複数の環境カメラ画像情報を用いた モバイルカメラの位置姿勢推定

Pose Estimation of a Mobile Camera by using SIFT Features in  
Environmental Camera Images

濱田 修平 北原格 亀田能成 大田友一  
Shuhei HAMADA Itaru KITAHARA Yoshinari KAMEDA Yuichi OHTA

筑波大学 システム情報工学研究科  
(〒3058573 茨城県つくば市天王台 1-1-1, hamada@image.esys.tsukuba.ac.jp)

**Abstract:** This paper proposes a method for estimating the pose of a mobile camera by using SIFT extracted in multiple environmental images. At first, we extract SIFT key points in two environmental images, and search the corresponding points by matching SIFT keys. 3D information of the corresponding points are estimated to apply stereo vision algorithm. We generate a data-set to combine the estimated 3D information and SIFT keys in the environmental cameras. We also extract SIFT key points in the mobile image. By matching SIFT key between the data-set and mobile camera, camera parameters of the mobile cameras are estimated. Our approach has some advantages that it does not require prior knowledge of captured scene geometry or setting visual markers, and it is not affected by changing the lighting condition.

**Key Words:** MR(Mixed Reality), Calibration, SIFT, Mobile camera.

## 1. はじめに

現実空間の映像とコンピュータグラフィックスによって生成した仮想空間の映像を融合してユーザーに提示する、複合現実感に関する研究が、近年盛んに行われている[1]。複合現実感の研究課題として、現実空間と仮想空間の幾何的整合性の実現がある。これまでに、環境に特定のビジュアルマーカーを設置し、それをカメラで撮影する手法や、カメラに加えて3次元位置センサを併用する手法などが開発されているが、実利用シーンを考えると、特殊なセンサを用いずに画像中に写りこんだ自然特徴量を手がかりに位置合わせを行う手法が待ち望まれている。我々は、環境カメラによって撮影した映像の情報を併用することにより、3次元位置センサやビジュアルマーカーを用いることなく、撮影環境の変化にロバストなモバイルカメラの位置姿勢を推定する手法の研究に取り組んでいる。

## 2. 複数の環境カメラ画像を利用したモバイルカメラの位置・姿勢推定手法

本稿では、複数の環境カメラで撮影した映像の情報を用いて、モバイルカメラの位置・姿勢関係を求める手法を提案する。図1に示すように、提案する位置姿勢の推定処理に必要な入力情報は、モバイルカメラの画像と、その撮影

タイミングと同期して撮影した2台の環境カメラ画像のみであり、特殊な3次元位置センサなどが混在しない簡易な構成で実現可能である。また、撮影空間中に特殊な人工マーカーを配置したり、事前に撮影対象の3次元CGモデルを準備したりする必要はない。さらに、モバイルカメラ画像の撮影タイミングと同期した環境カメラ画像を用いて位置合わせを行うため、照明条件の変化に対してロバストであり、また、撮影空間に移動物体が存在するようなシーンへの適用が可能である。

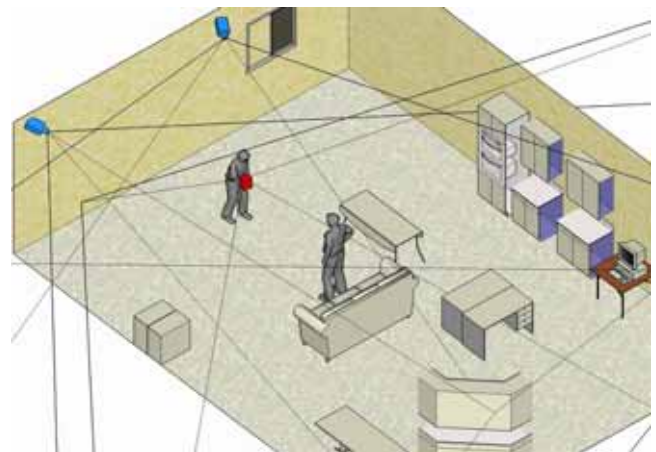


図1：環境カメラを用いたモバイルカメラの位置合わせ

提案手法の大きな処理の流れについて述べる。まず、カメラキャリブレーション済みの環境カメラの画像中から SIFT 特徴点を抽出する。この特徴点を用いて 2 台の環境カメラ画像間の対応点探索を行い、ステレオ法により特徴点の 3 次元座標を推定する。SIFT 情報に推定 3 次元情報を関連づけた“3 次元情報付特徴点群”を生成する。モバイルカメラ画像においても同様に SIFT 特徴点の抽出を行い、その SIFT 情報と 3 次元情報付特徴点群の SIFT 情報のマッチングを行うことにより、3 次元空間中の点とモバイルカメラ画像上での観測点の組み合わせを獲得し、位置姿勢の推定を行う。

提案手法により、監視カメラが設置された空間であれば、手軽にモバイルカメラの位置姿勢を推定することが可能となるため、複合現実感を利用したミュージアムガイドや撮影対象のプライバシー保護などへの応用が期待できる[5]。

### 3. 3 次元情報付特徴点群の生成

本研究では、複数の環境カメラで撮影した画像中の特徴点の対応付けを行い、特徴点の 3 次元情報を推定し、それを利用してモバイルカメラ位置姿勢推定を行う。3 次元情報付特徴点群の生成の手順は、以下の通りである。

1. 環境カメラのキャリブレーションを行い、カメラ射影行列を求める。
2. 環境カメラ画像において SIFT 特徴点の抽出を行う。
3. 特徴点の SIFT マッチングを行い、マッチ点の SIFT 鍵を求める。
4. 1 で求めたカメラ射影行列と 3 で求めたマッチ点から特徴点の 3 次元座標を求め SIFT 鍵に 3 次元位置情報を追加する。

以下、本章では、各節において上記手順の説明を行う。

#### 3.1 カメラキャリブレーション

環境カメラは、あらかじめキャリブレーションが行われ、画像座標系(u, v)と世界座標系(X<sub>w</sub>, Y<sub>w</sub>, Z<sub>w</sub>)の関係は既知であるとする。座標間の対応を表す変換行列は射影変換行列 P と呼ばれ、世界座標系とカメラ座標系の関係は式(1)で表される。

$$\lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = P \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (1)$$

#### 3.2 SIFT 特徴点の抽出

環境カメラで撮影した画像に対して、特徴点の抽出を行う。本研究での特徴点抽出には SIFT[3] を用いる。SIFT は局所不変特徴量を抽出する一手法であり、特性として画像

の拡大縮小、照明変動、視点の変化に強いことがあげられる。特徴の異なる複数のカメラで撮影した画像中での対応点探索では、SIFT の持つ不変性が有効に機能することが予想される。本節では、特徴点抽出アルゴリズムの概要を説明する。

Difference of Gaussian( DOG) を用いて特徴となる候補を検出する。対象となる画像に異なるスケールのガウシアンフィルタを適用し、隣接する画像の差分から DOG 画像を生成する。それらのスケールスペース内の極値に基づく特徴点抽出を行うことにより、特徴点の位置と特徴点が発見されたスケールが求まる。記述子としては、輝度勾配の方向ヒストグラム[3] を用い、ベストな状態で 128 次元の特徴量が得られる。各特徴点について、x 座標、y 座標、回転角、スケール、128 次元の特徴量をもつ SIFT 鍵が獲得される。

#### 3.3 特徴点マッチング

SIFT 特徴点同士のマッチングは、SIFT 鍵の値に最近傍探索を適用することによって行う。特徴量間の距離として、記述子間のユークリッド距離を用い、距離が最小となるものからマッチングの候補とする。高次元ベクトルでの鍵の探索はとても複雑であるが、k-d 木探索[4] を基にした、Best-Bin-First (BBF) アルゴリズムを用いることにより、効率的に高確率で最近傍の対応点を見つけることができる。この際、マッチングの信頼性を高めるために、2 番目に近い特徴量を用いたマッチング結果の評価を行う。任意に設定したしきい値よりも、最小距離の候補と 2 番目の候補との距離が大きいことを条件とし、そうでない場合を破棄することによって、曖昧な対応を除去し、マッチング精度を高めている。

#### 3.4 3 次元座標の推定

互いに平行ではない 2 台のカメラで撮影した 2 枚の画像において、共通の対応点が観測されている場合、その対応点の 3 次元座標  $M = [X, Y, Z]^T$  は、両カメラの射影変換行列から推定することができる。本手法では、3.1 節で紹介した射影変換行列と、3.3 節で説明した画像間の対応点情報を用いて、SIFT 特徴点の 3 次元座標を求める。2 枚の画像中の対応点  $m = [u, v]^T$  と  $m' = [u', v']^T$  が既知ならば、式(1)より  $s\overline{m} = P\overline{M}$ 、 $s'\overline{m}' = P'\overline{M}$  の関係が求まる。ここで  $m, m', M$  は  $m, m', M$  の同次ベクトルである。また、 $s, s'$  は奥行を表すスカラーであり、 $P, P'$  は各々のカメラの射影変換行列である。2 式より次式が導かれる。

$$BM = b \quad (2)$$

ここで、

$$B = \begin{bmatrix} up_{31} - p_{11} & up_{32} - p_{12} & up_{33} - p_{13} \\ up_{31} - p_{21} & up_{32} - p_{22} & up_{33} - p_{23} \\ u'p'_{31} - p'_{11} & u'p'_{32} - p'_{12} & u'p'_{33} - p'_{13} \\ u'p'_{31} - p'_{21} & u'p'_{32} - p'_{22} & u'p'_{33} - p'_{23} \end{bmatrix} \quad (3)$$

$$b = [p_{14} - up_{34}, p_{24} - vp_{34}, p'_{14} - u'p'_{34}, p'_{24} - v'p'_{34}]^T \quad (4)$$

であり,  $p_{ij}, p'_{ij}$  はそれぞれ  $P, P'$  の  $i$  行  $j$  列の要素である. この式は, 方程式の数が未知数より多いため, 方程式が全て線形独立ならば厳密な解が存在しない. そこで, 解の誤差を最小にする疑似逆行列  $B^+ = (B^T B)^{-1} B^T$  を用いて,  $M = B^+ b$  とすることにより, 対応点の 3 次元座標  $M(X, Y, Z)$  を推定する. 全ての SIFT 特徴点について 3 次元座標情報を推定し, 撮影空間中の 3 次元情報付特徴点群を生成する.

#### 4. モバイルカメラ位置姿勢推定

モバイルカメラの位置姿勢の推定は, 以下の手順でおこなう.

1. モバイルカメラ画像より SIFT 特徴量を抽出する.
2. 1 で抽出した SIFT 特徴点と 3.4 節で説明した 3 次元情報付特徴点群との SIFT 鍵を用いてマッチングをおこなう.
3. マッチした点の画像上の観測座標と 3 次元情報付特徴点群での 3 次元座標の関係からモバイルカメラの位置姿勢を推定する.

SIFT による特徴点の抽出およびマッチング処理は 3.2, 3.3 節で説明した手法をモバイルカメラ画像にも適用し, モバイルカメラ画像中の観測座標と撮影空間の 3 次元座標の対応点が求まる. ここで, 式(1)より座標  $(u, v)$  を求めると以下のような式になる.

$$u = \frac{p_{11}X_w + p_{12}Y_w + p_{13}Z_w + p_{14}}{p_{31}X_w + p_{32}Y_w + p_{33}Z_w + p_{34}} \quad (4)$$

$$v = \frac{p_{21}X_w + p_{22}Y_w + p_{23}Z_w + p_{24}}{p_{31}X_w + p_{32}Y_w + p_{33}Z_w + p_{34}} \quad (5)$$

式 (1)は任意の定数倍を許した方程式であるため, そこから求められる行列  $P$  には定数倍の不定性が残る. そこで,  $p_{34} = 1$  とおき, 残りの  $P$  の各要素について整理すると以下のような式になる. 画像上の位置  $(u, v)$  とそれに対応する空間中の位置  $(X_w, Y_w, Z_w)$  が得られれば, 上式のように,  $P$  の各要素に関する 2 つの方程式が得られる. 未知数の数は

全部で 11 個  $(p_{11}, p_{12}, \dots, p_{33})$  なので, 既知の位置座標の組が最低 6 個あれば, カメラ射影行列  $P$  を求めることができる. 以上の処理により, モバイルカメラの射影変換行列が求まる.

## 5. 実験

### 5.1 実験環境

撮影実験で用いたカメラは, モバイルカメラとして Sony 製デジタルカメラ DSC-T30 (2048 画素×1536 画素) を, 環境カメラとして Point Grey Research 製の Scorpion(IEEE-1394 デジタルカメラ: 1280 画素×960 画素) を 2 台使用した. 撮影は屋内で行い, 環境カメラは, 1.4[m] の高さの平行に約 1[m] の間隔を空けて同じ空間を重複して撮影するように設置した. 各環境カメラによって撮影した.

### 5.2 モバイルカメラのカメラ射影行列の獲得

環境カメラに, GML C++ Camera Calibration Toolbox [2] を適用しキャリブレーションを行う. 任意の大きさの白黒格子パターンを複数回撮影し, カメラの内部・外部パラメータを求め, これらの行列を計算して, 環境カメラのカメラ射影行列を求める.

次に, 環境カメラ画像間での SIFT 特徴点の抽出を行い, それらのマッチングを行う. 図 2 の左上が環境カメラ L で撮影した画像, 右上が環境カメラ R で撮影画像で, マッチングできた特徴点を矢印で重畳してある. 環境カメラ L の画像で 3453 点, 環境カメラ R の画像で 3062 点の特徴点を抽出し, マッチングにより 142 点の対応点を獲得した. 獲得した対応点とカメラパラメータより求めたカメラ射影行列より, 対応点の 3 次元座標を計算する.

モバイルカメラ画像に対して SIFT 特徴点の抽出を行い, 5250 点の特徴点を抽出した. さらに, 環境カメラの SIFT 特徴点間の対応点と, モバイルカメラの SIFT 特徴点のマッチングを行った結果を図 2 の環境カメラ L の画像 (左上) とのモバイルカメラ画像 (左下) に示す. 画像間のラインの上点と下点に対応点となる. 環境カメラ L の画像に対し 20 点の対応点を獲得することができる. 目視で確認した結果, そのうち 15 組の点は正しく対応が取れていることがわかった.

空間中の 3 次元座標と画像上の 2 次元座標の正しい対応点が最低 6 組以上あれば, カメラ射影行列を求めることが可能である. 本実験では, 獲得した対応点の組から最小 2 乗法を用いてモバイルカメラのカメラ射影変換行列を求めた.

獲得したモバイルカメラの射影変換行列を用いて, 実空間中に存在する机の上に仮想物体を CG で重畳させた結果を図 2 右下に示す. 特殊なセンサを用いずに画像中に写りこんだ自然特徴量だけを用いて, カメラの位置合せが実現できていることが確認できる.

## 6. おわりに

本研究では、環境カメラから得られる情報を用いてモバイルカメラの位置姿勢を推定する手法を提案した。SIFT特徴量を用いて撮影環境の3次元情報を持った特徴点群を生成することにより、特殊なセンサを必要とせず、画像中の自然特徴量を用いて、様々な環境下にロボタストに対応可能なモバイルカメラのカメラ射影行列を求め、複合現実感を実現した。今後の課題として、位置・姿勢獲得精度と処理速度の向上があげられる。

本研究の一部は、文部科学省科学研究補助金若手研究A「被写体のプライバシーを考慮したモバイルカメラによる高自由度映像監視技術に関する研究」によって実施したものである。

- [1]Y. Ohta, and H. Tamura, "Mixed Reality – Merging Real and Virtual Worlds-", Ohmsha,Ltd.1999.
- [2]<http://research.graphicon.ru/calibration/gml-c-camera-calibration-toolbox-5.html>
- [3]David G. Lowe, "Object recognition from local scale-invariant features," international Conference on Computer Vision, Corfu, Greece (September 1999), pp. 1150-1157.
- [4]J. S. Beis and D. G. Lowe. Shape indexing using approximate nearest-neighbour search in high-dimensional spaces. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1000.1006,1997
- [5]Itaru Kitahara, Kiyoshi Kogure, Norihiro Hagita, "Stealth Vision for Protecting Privacy", Proc. of 17th International Conference on Pattern Recognition (ICPR 2004), Vol.4, pp.404-407, (2004)

## 参考文献



図2 : (左上) 環境カメラ L 画像 (右上) 環境カメラ R 画像

(左下) モバイルカメラ画像 (右下) CG 重畳したモバイルカメラ画像