

自然特徴点ランドマークデータベースを用いた カメラ位置および姿勢推定に関する研究

奈良先端科学技術大学院大学 武富 貴史
奈良先端科学技術大学院大学 佐藤 智和
奈良先端科学技術大学院大学 横矢 直和

1. はじめに

CGなどの仮想環境をカメラで撮影した画像中へ位置合わせして重畳表示することで情報を付加するビデオスルー型拡張現実感技術は、ヒューマンナビゲーション、教育・作業支援、エンタテインメントなどのさまざまな分野への応用が期待される。ビデオスルー型拡張現実感では、仮想環境と現実環境との位置合わせに、重畳対象となる画像を撮影したカメラの位置・姿勢を推定する必要がある。カメラ位置・姿勢の推定技術は、ジャイロセンサや加速度センサなどのセンサ情報に基づく方法と撮影された画像情報に基づく方法に大別できる。初期の拡張現実感技術研究では、センサ情報に基づくカメラ位置・姿勢推定手法が主流であったが、近年では、計算機の処理能力の向上により画素単位での高精度な位置合わせが可能な画像情報に基づく手法が主流となっている。

画像情報に基づくカメラ位置・姿勢推定手法では、一般に、カメラの内部パラメータ（レンズ歪、画像中心、焦点距離）は既知で変化しないものとし、カメラの外部パラメータ（位置・姿勢）のみを推定する。このような画像情報に基づくカメラ位置・姿勢推定手法として、人工マーカを用いる手法¹⁾、3次元モデルを用いる手法²⁾、Visual-SLAMによる手法³⁾、自然特徴点ランドマークデータベースを用いる手法⁴⁾などさまざまな手法が提案されている。これらの手法の中でも、自然特徴点ランドマークデータベースを用いた手法⁴⁾は、利用環境内で撮影した全方位映像に対して3次元形状復元手法であるStructure-from-motion (SfM)法を適用することで、広域な屋外環境においても半自動でデータベースを構築することができる。また、この手法は、SfM法における最適化計算（バンドル調整）の際に、少数の基準点⁵⁾またはGPSより得られる位置情

報⁶⁾を絶対指標として用いることで、画像情報のみに基づく Visual-SLAM や SfM 法で問題となる蓄積誤差が生じないという利点を持つ。本稿では、このような特長を持つ自然特徴点ランドマークデータベースを用いたカメラ位置・姿勢推定手法について詳しく解説する。

2. 自然特徴点ランドマークデータベースの構築

自然特徴点ランドマークデータベースを用いた手法は、オフラインでのデータベース構築処理とオンラインでのカメラ位置・姿勢推定処理で構成される(図1参照)。ここでは、オフラインでのデータベース構築処理について述べる。ランドマークデータベースの構成要素を図2に示す。データベースにはランドマークごとに、(a) ランドマークの3次元位置、(b) 撮影地点に依存した情報、(c) ランドマークの優先度情報が保持される。以下では、これらのランドマーク情報の取得方法について述べる。

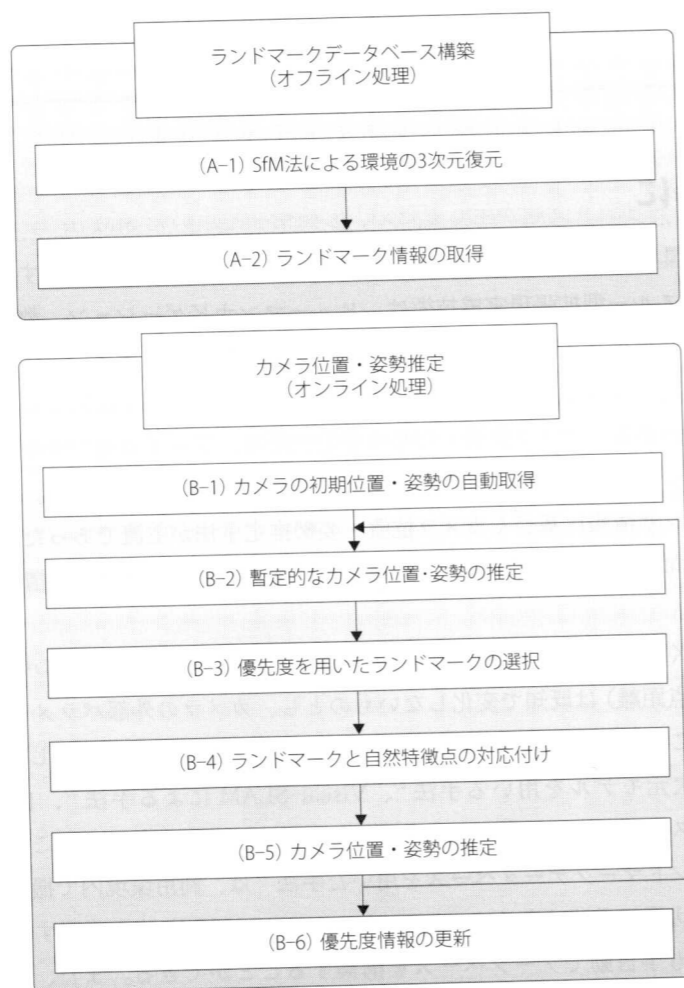


図1 ランドマークデータベースを用いたカメラ位置・姿勢推定処理の流れ

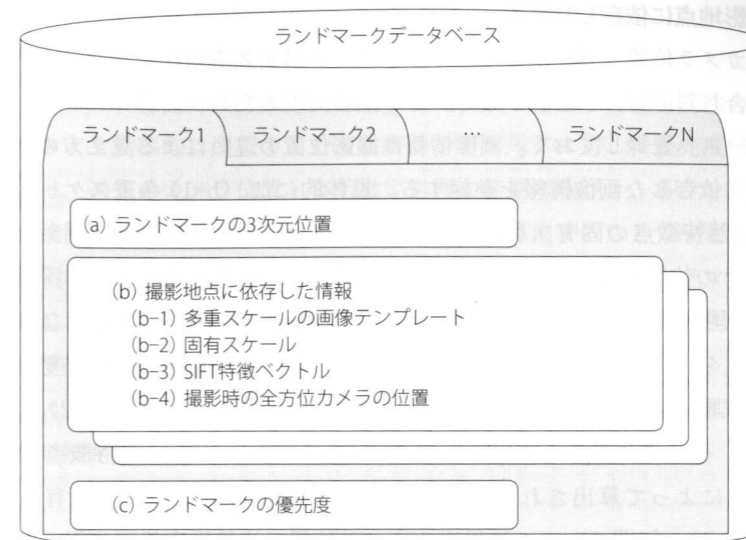


図2 ランドマークデータベースの構成要素

2.1 Structure-from-motion 法を用いた3次元情報の取得 (A-1)

ランドマークデータベースを構築するために、利用環境の3次元形状情報を動画からの3次元形状復元手法である SfM 法を用いて取得する。ただし、画像情報のみを用いた SfM 法では推定誤差が蓄積するという問題が生じるため、参照情報を用いた SfM 法^{5),6)}を用いることで、広域な環境においても蓄積誤差の生じない環境の3次元復元を行う。これらの手法では、まず、全方位カメラを用いて利用環境内の動画を撮影する。次に、動画中の自然特徴点を Harris オペレータ⁷⁾により検出しフレーム間で追跡する。この自然特徴点の追跡結果に基づき、自然特徴点の3次元位置および画像上の座標、全方位カメラの位置・姿勢を推定する。最後に、3次元位置が既知の少数の基準点または GPS から取得される位置情報を用いて、動画全体での最適化処理を行うことで、自然特徴点の3次元位置および推定されたカメラ位置・姿勢の累積的な推定誤差を最小化する。これにより、広域な環境においても蓄積誤差を含まない自然特徴点の3次元位置の推定を実現することができる。

2.2 ランドマーク情報の取得 (A-2)

[2.1]で述べた SfM 法により3次元位置が推定された自然特徴点をランドマークとし、各ランドマークに対してオンラインでのカメラ位置・姿勢の推定に必要なランドマーク情報を取得する。以下ではランドマークデータベースの各要素の情報取得方法について述べる。

2.2.1 (a) ランドマークの3次元位置

オンラインでのカメラ位置・姿勢推定処理では、ランドマークの3次元位置と画像上での観測座標を対応付けることでカメラ位置・姿勢の推定を行う。そのため、各ランドマークの3次元位置が必要となる。ランドマークの3次元位置は、[2.1]で述べた SfM 法により推定されたものであり、利用環境に固定された世界座標系で保持される。

2.2.2 (b) 撮影地点に依存した情報

オンラインのカメラ位置・姿勢推定処理では、データベース中のランドマークと入力画像中の自然特徴点を照合し対応付ける必要がある。この照合処理のために、各ランドマーク周辺の画像情報をデータベースへ登録しておく。画像情報は撮影位置の違いによる見え方の変化に対応するため、撮影地点に依存した画像情報を登録する。具体的には、(b-1) 多重スケールの画像テンプレート、(b-2) 自然特徴点の固有スケール、(b-3) SIFT 特徴ベクトル、(b-4) ランドマーク撮影時の全方位カメラの位置が保持される。多重スケールの画像テンプレートは、**図3**に示すように、カメラの投影中心とランドマークの3次元位置を結ぶ直線に対して垂直な平面に対して自然特徴点周辺のパターンを投影することで作成する。これにより、カメラの姿勢に依存しない画像パターンを生成することができる。また、自然特徴点の固有スケール (b-2)、SIFT 特徴ベクトルは (b-3) は、それぞれ Harris-Laplacian⁸⁾ によって算出される自然特徴点スケール情報、SIFT-descriptor⁹⁾ によって算出される SIFT 特徴ベクトルが登録される。固有スケール情報、SIFT 特徴ベクトルは、初期フレームに対するカメラ位置・姿勢推定処理に用いられる。また、オンラインのカメラ位置・姿勢推定処理において、対応付け候補となるランドマークを選択するためにランドマーク撮影時の全方位カメラの位置 (b-4) を登録する。この撮影位置情報を用いることで、オンラインでの対応付け候補選択処理において撮影可能なランドマークをデータベ

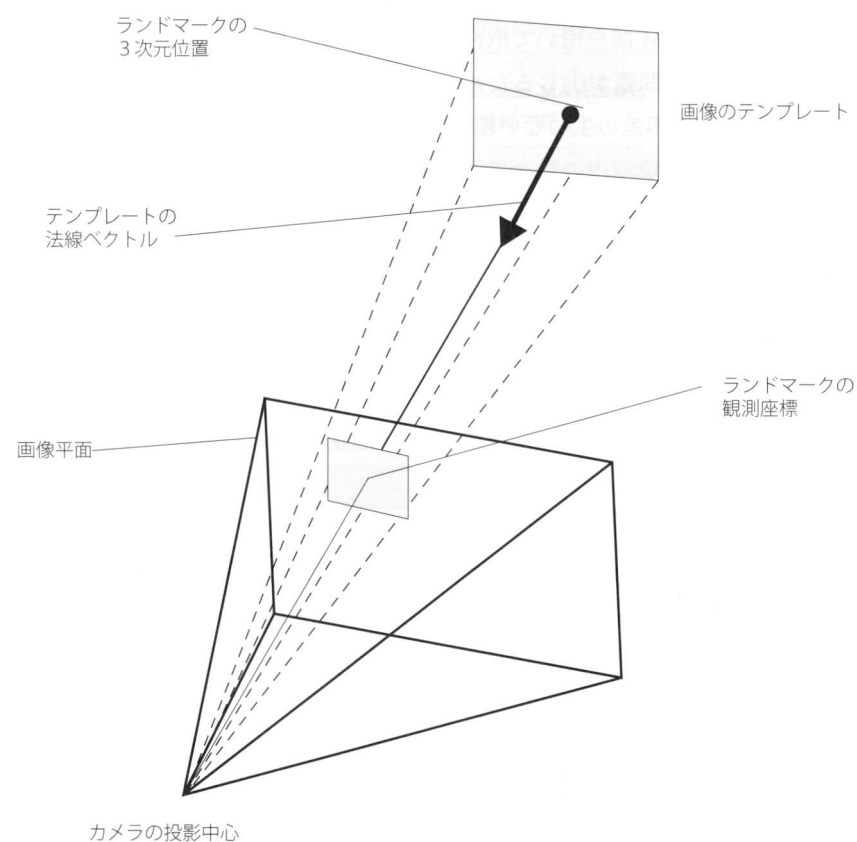


図3 画像テンプレートの生成

ス中より選択することが可能となる。

2.2.3 (c) ランドマークの優先度情報

オンラインでのカメラ位置・姿勢推定処理において、正しく対応付く可能性の高いランドマークをデータベースより効率的に選択するために各ランドマークに対して優先度情報を付加する。この優先度情報は、過去のカメラ位置・姿勢推定処理の結果に基づき、各ランドマークが正しく対応付きカメラ位置・姿勢の推定処理に利用される確率を算出することで決定される。ただし、データベース構築時は優先度を算出することができないため、全てのランドマークに対して同一の優先度を設定しておき、一般ユーザーによる利用の前に、データベース管理者が複数の学習用データを与え優先度情報を更新しておくことを想定する。優先度情報 P_i は、ランドマーク i がデータベースより選択された回数 D_i とオンラインでのカメラ位置・推定処理において誤対応として排除されずに推定処理に用いられた回数 E_i を用いて式 (1) で算出する。

$$P_i = E_i / D_i \quad (1)$$

この優先度情報については、各ユーザーの利用後に推定結果をフィードバックすることで更新する。この更新処理については [3.5] で述べる。

3. 自然特徴点ランドマークデータベースを用いたカメラ位置・姿勢推定

オンラインのカメラ位置・姿勢推定処理では、まず、初期フレームのカメラ位置・姿勢を推定 (B-1) した後、暫定的なカメラ位置・姿勢の推定 (B-2)、ランドマークの選択 (B-3)、ランドマークと自然特徴点の対応付け (B-4)、現フレームのカメラ位置・姿勢の推定 (B-5) を繰り返す。最後に、カメラ位置・姿勢推定結果をデータベースへフィードバックしランドマークの優先度情報を更新する (B-6)。以下では、オンラインでのカメラ位置・推定処理の各処理について述べる。

3.1 カメラの初期位置・姿勢の推定 (B-1)

初期フレームのカメラ位置・姿勢の推定処理は、ランドマークデータベースを用いた静止画像からのカメラ位置・姿勢推定手法¹⁰⁾により自動で取得する。この手法では、まず、入力画像中から自然特徴点を検出し、各自然特徴点に対してデータベース構築時と同様に固有スケール、SIFT 特徴ベクトルの算出を行う。次に、データベース中の全てのランドマークの SIFT 特徴ベクトルと入力画像中から抽出された自然特徴点の SIFT 特徴ベクトルを照合し対応付ける。これらの対応付けられたランドマークと自然特徴点の固有スケールの比に基づき、環境中での大まかな撮影位置 (2 自由度)・方位 (1 自由度) を絞り込む。最後に、推定された撮影位置・方位に存在するデータベース中のランドマークのみを用いて再度対応付け処理を行い、対応付け結果から誤対応を排除した後、PnP 問題を解くことで6自由度のカメラ位置・姿勢を推定する¹¹⁾。ただ

し、この初期フレームのカメラ位置・姿勢推定処理には、十数秒を必要とするため、初期位置・姿勢の推定処理が完了するまでの間はカメラを動かさないものとする。

3.2 暫定的なカメラ位置・姿勢の推定 (B-2)

フレーム間でランドマークを追跡することで、現フレームの暫定的なカメラ位置・姿勢を推定する。ランドマークデータベースを用いたカメラ位置・姿勢推定処理では、ランドマークの照合処理において、データベース構築時のカメラとオンラインの推定時のカメラの画角の違い、レンズ歪の影響を除去した画像パターンを生成し、比較を行うが、この画像パターン生成処理には多くの計算コストを必要とする。そこで、連続フレーム間でランドマークを追跡し、暫定的なカメラ位置・姿勢の推定を行っておくことで、データベース中のランドマークと照合すべき入力画像中の自然特徴点候補を限定し計算コストの削減を行う。

暫定的なカメラ位置・姿勢の推定処理では、まず、直前のフレームで自然特徴点と対応付けられカメラ位置・姿勢の推定処理に利用されたランドマークを選択し、現フレームにおける対応点を探査する。この探索処理では、連続フレームにおけるパターンの変形や明るさの変化は微小であることから、ランドマークの前フレームにおける画像上の位置を中心とする一定の範囲に存在する自然特徴点を対象として、SSD (Sum of Squared Differences) を評価尺度としたテンプレートマッチングにより現フレームにおける対応点の位置を決定する。次に、追跡されたランドマークから誤対応を排除した後、正しい対応関係を用いて PnP 問題を解くことで暫定的なカメラ位置・姿勢の推定を行う。

3.3 ランドマークの選択 (B-3)

データベース中より新たに対応付けるランドマークを選択する。ランドマークの選択処理では、まず、暫定的なカメラ位置・姿勢情報とランドマークの3次元位置を用いて幾何学的な制約に基づき撮影可能なランドマークを選択する。次に、撮影可能なランドマークより、正しく対応付く可能性の高いランドマークを優先度情報に基づき一定個数選択する。以下では、幾何学的な制約に基づくランドマーク選択と優先度情報に基づくランドマーク選択について述べる。

3.3.1 幾何学的な制約に基づくランドマークの選択

幾何学的な制約に基づくランドマークの選択処理では、暫定的なカメラ位置・姿勢、ランドマーク撮影時の全方位カメラの位置、ランドマークの3次元位置を用いて以下の条件を満たす画像テンプレートをデータベース中より選択する。

- (条件1) ランドマークが入力画像の画角内に存在する
- (条件2) 暫定的なカメラ位置とランドマークの3次元位置を結ぶ直線と、ランドマーク撮影時の全方位カメラの位置とランドマークの3次元位置を結ぶ直線の成す角が一定値以下
- (条件3) 暫定的なカメラ位置とランドマーク撮影時のカメラ位置の距離が一定値以下

上記の幾何学的な制約を用いることによって、暫定的なカメラ位置・姿勢から撮影可能なラン

ドマークの画像テンプレートをデータベース中より選択することができる。

3.3.2 優先度情報に基づくランドマークの選択

幾何学的な制約により選択されたランドマークの中から、正しく対応付く可能性の高いランドマークを優先度情報に基づき選択する。ここでは、まず、入力画像を格子状に分割し、各格子内で最も優先度の高いランドマークを選択する。次に、入力画像全体で優先度の高い順に一定個数選択する。このように、優先度情報を用いてランドマークを選択することで、正しく対応付く可能性の高い少数のランドマークをデータベースより選択することが可能となる。

3.4 ランドマークの対応付けとカメラ位置・姿勢推定 (B-4)、(B-5)

選択されたランドマークの画像テンプレートと入力画像中の自然特徴点周辺の画像パターンを照合し、対応点の探索を行う。この照合処理では、データベース構築時と同様に対応付け候補である入力画像中の自然特徴点に対して撮影姿勢に依存しない画像パターンの生成後、正規化相互相関に基づく照合を行う。ただし、各ランドマークの対応点候補となる自然特徴点は、ランドマークを暫定的なカメラ位置・姿勢を用いて画像上へ投影した座標を中心とする一定の範囲に存在するものに限定する。最後に、対応付け結果より、誤対応を排除した後、正しい対応付け結果を用いて PnP 問題を解くことで現フレームのカメラ位置・姿勢を推定する。

3.5 優先度情報の更新 (B-6)

カメラ位置・姿勢推定処理完了後、推定結果をデータベースへフィードバックすることでランドマークの優先度情報を更新する。更新処理では、オンラインでのカメラ位置・姿勢推定処理において、ランドマークの選択処理 (B-3) において各ランドマークが選択された回数 E_{new} 、カメラ位置・姿勢推定処理 (B-5) において誤対応として排除されずにカメラ位置・姿勢の推定処理に利用された回数 D_{new} を利用する。具体的には、式 (2) を用いて各ランドマークの優先度情報を更新する。

$$P_i = (E_{old} + E_{new}) / (D_{old} + D_{new}) \quad (2)$$

ここで、 E_{old} 、 D_{old} は、それぞれ過去に同じ環境内でカメラ位置・姿勢の推定を行った際の頻度情報でデータベースに保持されているものである。

4. カメラ位置・姿勢推定実験

実環境において、実際にランドマークデータベースを構築し、カメラ位置・姿勢の推定実験を行った。この実験では、全方位カメラ (Point Grey Research 社 (カナダ)、Ladybug) で約 75 m を移動撮影した動画像 (1260 フレーム) を入力としてデータベースを構築し、ビデオカメラ (SONY DSR-PD-150) を用いて撮影した 4 本の動画像に対してカメラ位置・姿勢推定を行った。ただし、オンラインでのカメラ位置・姿勢推定処理において、初期フレームのカメラ位置・姿勢

に対して、レーザレンジファインダで計測した密な形状情報を用いて画像パターンの見え方を補正する方法¹³⁾やSfM法によって復元された3次元形状に対して局所平面を当てはめることで視点位置に依存しない画像パターンを生成する手法¹⁴⁾が提案されている。

レーザレンジファインダを用いた手法では、より高精度な位置合わせが必要な場所において、全周レンジファインダを用いて密な形状情報を取得し、この形状情報を用いることで全方位カメラの撮影地点から離れた地点の画像テンプレートを仮想的に生成しておくことで、ランドマークの見え方の変化に対応している。また、SfM法によって復元された3次元形状に対して局所平面を当てはめる手法では、画像テンプレートの生成過程において、ランドマークと撮影したカメラ位置を結ぶ直線に垂直な平面へ画像を投影するのではなく、ランドマーク周辺の3次元点群に対して局所平面を当てはめ、この平面へ画像を投影することで視点位置に依存しない画像テンプレートを生成している。これにより、見え方が変化し易い地面などに存在するランドマークを正しく対応付けることが可能となり、データベース構築時のカメラの撮影経路から離れた場合においても、安定したカメラ位置・姿勢推定処理が実現できる。

6. まとめ

本稿では、自然特徴点ランドマークデータベースを用いたカメラ位置・姿勢推定手法について述べた。この手法では、事前に利用環境において構築したランドマークデータベースを用いてカメラ位置・姿勢の推定を実現している。データベースの構築処理では、SfM法を用いることにより半自動でランドマークデータベースを構築することができるが、この手法では、静的な環境での利用を想定しており、利用環境の3次元構造が変化した場合には再度データベースを構築する必要が生じる。今後の課題として、環境の変化が生じた場合におけるデータベース更新処理手法の開発などが挙げられる。

参考文献

- 1) H. Kato, H. Billinghurst: "Marker tracking and HMD calibration for a video-based augmented reality conferencing system", Proc. Int. Workshop on Augmented Reality, pp.85-94 (1999).
- 2) T. Drummond, R. Cipolla: "Real-time visual tracking of complex structure", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **27** (7), pp.932-946 (2002).
- 3) G. Klein, D. Murray: "Parallel tracking and mapping for small AR workspaces", Proc. Int. Symp.on Mixed and Augmented Reality, pp.225-234, (2007).
- 4) 武富貴史, 佐藤智和, 横矢直和: 拡張現実感のための優先度情報を付加した自然特徴点ランドマークデータベースを用いた実時間カメラ位置・姿勢推定. 電子情報通信学会論文誌, **J92-D** (8), pp.1440-1451 (2009).
- 5) 佐藤智和, 池田聖, 横矢直和: 複数動画からの全方位型マルチカメラシステムの位置・姿勢パラメータの推定. 電子情報通信学会論文誌 (D-II), **J88-D-II** (2), pp.347-357 (2005).
- 6) S. Ikeda, T. Sato, K. Yamaguchi, N. Yokoya: "Construction of feature landmark database using omnidirectional videos and GPS positions", Proc. Int. Conf. on 3-D Digital Imaging and Modeling, pp.249-256 (2007).
- 7) C. Harris, M. Stephens: "A combined corner and edge detector", Proc. Alvey Vision Conf., pp.147-151 (1988).

- 8) K. Mikolajczyk, C. Schmid: "Scale & affine harris laplace invariant interest point detectors", *Int. J. of Computer Vision*, **60** (1), pp.63-86 (2004).
- 9) D. G. Lowe: "Distinctive image features from scale-invariant keypoints", *Int. J. of Computer Vision*, **60** (2), pp.91-100 (2004).
- 10) 薄充孝, 中川知香, 佐藤智和, 横矢直和: ランドマークデータベースに基づく静止画像からのカメラ位置・姿勢推定. 日本バーチャルリアリティ学会論文誌, **13** (2), pp.161-170 (2008).
- 11) R. Klette, K. Schluns, A. Koschan: *Computer Vision-Three-dimensional Data from Image*, Springer (1998).
- 12) H. Tamura, H. Kato: "Proposal of international voluntary activities on establishing benchmark test schemes for AR/MR geometric registration and tracking methods", Proc. Int. Symp.on Mixed and Augmented Reality, pp.233-236 (2009).
- 13) T. Taketomi, T. Sato, N. Yokoya: "Real-time and Accurate Extrinsic Camera Parameter Estimation using Feature Landmark Database for Augmented Reality", *Int. J. of Computers and Graphics*, **35** (4), pp.768-777 (2011).
- 14) 川戸一希, 武富貴史, 佐藤智和, 横矢直和: 三次元点群への局所平面当てはめに基づく法線情報付きランドマークデータベースを用いたカメラ位置・姿勢推定. 画像の認識・理解シンポジウム (MIRU2011) 講演論文集, pp.1133-1140 (2011).