

修士 (工学)

仮想化現実モデルと画像に基づく携帯カメラの位置姿勢推定

A photo-shoot localization of mobile phone camera based on virtualized reality models and photos

西 田 純
NISHIDA Jun知能機能システム専攻
Master's Program in Intelligent Interaction Technologies指導教員: 葛岡 英明, 亀田 能成, 蔵田 武志 (知能機能工学域)
Supervised by: KUZUOKA Hideaki, KAMEDA Yoshinari and KURATA Takeshi
(Division of Intelligent Interaction Technologies)

This paper provides a study on an estimation of photo-shoot location and orientation about indoor environments for augmented reality applications. This study is based on a server with virtualized reality models and key frames. Key frames are photos with known photo-shoot locations and orientations. A real image taken by a mobile device is sent to a server, and the image is compared to model data. The model data include 3D coordinates of the feature points on the key frames and edges on the structure of the virtualized-reality model. A focus of the study is the performance of localization using mutual information, edge information, SURF, ORB and FREAK features. In this work, some experiments are performed with real images taken by a mobile device and the model data from virtualized reality environment models.
keywords: Camera localization, Augmented reality, Template matching, Image matching, Feature point correspondence.

1 序論

拡張現実 (Augmented Reality : AR) アプリなどを用いた位置や姿勢に基づく情報サービスが国内外で注目を浴びている¹⁾²⁾³⁾。実環境で撮影した画像中に CG などの仮想物体を描画するためには、一般にアプリのユーザもしくはカメラの位置・姿勢を推定することが必要であり、現在まで様々な研究が進められてきた。中でも、画像処理を用いてカメラの位置・姿勢を推定する手法は、高精度な推定を実現することが可能である。本研究では特に、携帯カメラで撮影した画像を入力とし、3次元モデルの情報をデータベースとして用いたモデルベースのカメラ位置・姿勢推定手法に着目する。

データベースを用いた手法として例えば、文献⁴⁾では撮影位置と姿勢が既知のパノラマ画像群と入力画像とをアフィン変換や射影変換を用いて対応付けすることで位置・姿勢推定を実現している。しかしながら、奥行き差の大きなシーンでは、撮影位置からのずれに伴う視差によって見えが局所的に変化し、対応付けが困難となるため、少しずつ視点を変えてパノラマ画

像を撮影していかなければならない。例えば、典型的な屋内環境であれば、1~2m ほどのパノラマ画像撮影による基盤データベース整備が求められる。そのため、効率的に任意な視点での位置・姿勢推定を実現するためには、撮影画像に含まれる画像特徴と実環境の3次元モデルの構造とを対応付けるような3次元モデルベースの手法が適していると言える。

本研究のための予備調査として、3次元モデルの構造と携帯端末内蔵のカメラ (以降、携帯カメラと呼ぶ) で実環境を撮影した画像 (以降、実画像と呼ぶ) とを対応付ける実験を行った⁵⁾⁶⁾。本実験では、実画像と仮想化現実モデルを用いて生成した画像 (以降、生成画像と呼ぶ) との比較による絶対位置・姿勢推定手法を想定し、色相・彩度のヒストグラム, SSD, 正規化相互相関 (ZNCC), 及び SIFT の各特徴量がどのように振る舞うかについて調査した。その結果、実画像と生成画像を比較する場合には画像全体の見えの類似度尺度として ZNCC が最も有効であり、大局的な位置・姿勢推定ができる可能性が示された。一方で、ZNCC による類似度が最も

高い生成画像を位置・姿勢の推定結果として AR システムに適用するには精度が不十分であった。

そこで本研究では、3次元モデルの情報を事前に付加した実画像（以降、キーフレームと呼ぶ）を用いてデータベースを作成し、そのキーフレームと実画像を特徴点、エッジ、面の情報それぞれの特徴量を用いて対応付けを行い、AR システムへの適用に妥当なカメラの位置・姿勢の推定を行う。

以下では、まず、本研究の位置付けと方針について述べる。次に、本論文の構成を述べる。

1.1 研究の位置付けと方針

AR アプリケーションにおいて、CG などの仮想物体を描画するためには、カメラの絶対位置・姿勢を推定することが必要である。一般に、カメラの絶対位置・姿勢を推定する手法は、マッチングのみで推定する手法、マッチングとトラッキングを併用する手法がある。本研究では、マッチングとトラッキングの併用する方式の推定を想定してマッチング方法を調査した。

実画像と生成画像の対応付けにおいて、色相・彩度のヒストグラム、SSD、正規化相互相関、及び SIFT の各特徴量がどのように振る舞うかについて調査した。実画像と実画像に近い視点の生成画像を比較した際に、正規化相互相関を類似度尺度として用いた場合に最も良い結果を示した。SIFT 特徴量を用いた特徴点検出では、実画像と比較して生成画像では十分な数の特徴点を得ることができなかった。また、特徴点の対応付けにおいても誤対応が多くカメラの位置・姿勢を推定するためには性能が不十分であった。

そこで、本研究では3Dモデルの作成に用いた画像を用いることで、その画像の撮影位置・姿勢と各画素の3次元座標の既知な画像を作成してキーフレームとし、実画像と比較することでカメラの位置・姿勢を推定する。カメラの位置・姿勢を推定する手法として、特徴点を用いた手法、エッジを用いた手法、面の情報を用いた手法の3種類の手法について実験した。特徴点の検出手法として、SURF (Speeded Up Robust Features)⁷⁾、ORB (Oriented FAST and Rotated BRIEF)⁸⁾、FREAK (Fast Retina Keypoint)⁹⁾を用い、エッジ検出では実画像からキャニーエッジ検出器

¹⁰⁾を用いて検出したエッジと3Dモデルから出力した構造上のエッジを用いた。また、面の情報として相互情報量 (Mutual Information: MI)¹²⁾を用いた。

1.2 本論文の構成

以下、2節では位置・姿勢推定の従来研究について概観し、3章では構築したARシステムについて述べる。次に、4節ではカメラの位置・姿勢推定に利用した特徴量について述べ、5章では実験内容とその結果、応用例を述べる。最後に、6章でまとめと今後の展望について述べる。

2 位置・姿勢推定の従来研究

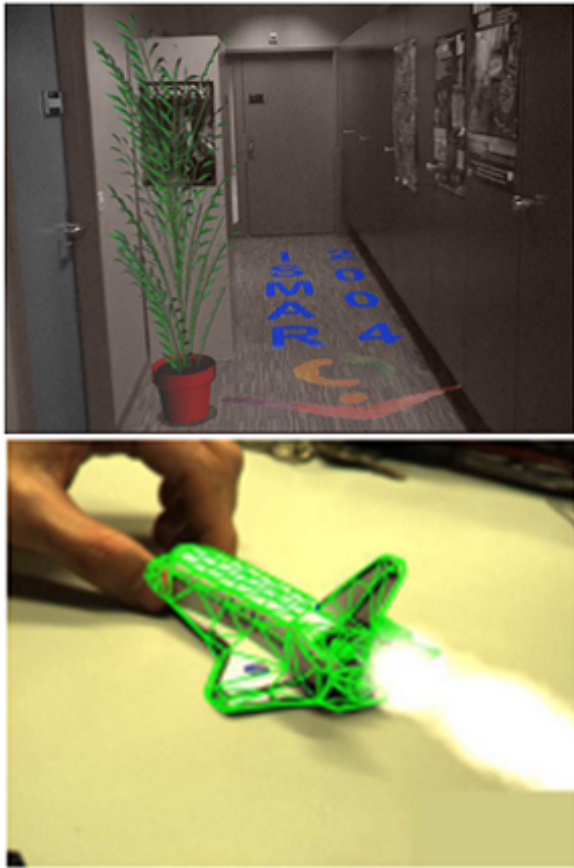
現実環境に仮想環境を重畳表示する拡張現実感技術は、Figure 1に示すように、様々な研究・開発が盛んに行われている¹¹⁾。これらの拡張現実感技術を用いたアプリケーションを開発するためには、仮想物体を違和感なく合成するため、実時間で動作し、かつ高精度なカメラの位置・姿勢推定技術が必要である。

カメラの位置・姿勢を推定する手法には、センサを用いた手法やユーザが持つカメラで取得した画像を基に画像処理を行う手法が主流である。本研究では特に、画像処理を用いた方法に着目する。画像処理を用いた手法は、事前準備の不必要な手法と事前準備の必要な手法に分類できる。以下では、はじめに、センサを用いた手法について述べ、次に、画像処理を用いた手法について述べる。

2.1 センサを用いた位置・姿勢推定手法

屋外環境では、GPSをはじめとするGNSS (Global Navigation Satellite Systems)を用いた研究開発事例¹³⁾や実サービスが既に数多く存在する。ただし、GNSSは基本的に屋外でのみ利用可能であるため、屋内環境での位置姿勢計測やそれに基づく情報サービスをどのように実現するかについては、未だに様々な技術課題が残されている。

屋内測位の手法としては、まず、Wi-Fi やRFID、IrDA (光無線データ通信)に基づくものがあげられる¹⁴⁾¹⁵⁾。これらの手法の中には実用化されたものも多いが、物理的なインフラ (タグや基地局等)、基盤データベース (インフ

Figure 1: Examples of AR applications ¹¹⁾

ラの配置や電波強度等), もしくはその両方を必要とするため, 初期コストやメンテナンスコストをいかに下げるかという問題がある. また, GNSS にも言えるが, 一般には AR 等の詳細な位置情報サービスを実現するには精度が不十分であり, 高精度な位置・姿勢の情報は別の手段で取得しなければならない.

2.2 画像処理を用いたカメラの位置・姿勢推定手法

画像を用いた手法は事前準備の不必要な手法と事前準備の必要な手法に分類できる. 事前準備の不必要な手法の例として, 3次元空間への特徴点のマッピングとカメラの位置・姿勢推定を同時に行う PTAM が挙げられる ¹⁶⁾. このような手法では, カメラの位置・姿勢はフレーム間の相対的な位置・姿勢を扱っているため, 本手法単独では, 絶対的な位置・姿勢が得られない. そのため, ナビゲーションサービスのように絶

対的な位置・姿勢が必要なサービスへの適用は難しい.

一方で, 事前準備の必要な手法として AR の分野での適用例が多いマーカベースの手法 ¹⁷⁾ では, マーカがカメラに写っている際には, 比較的安定して位置姿勢を推定することができる. 本手法では, 実環境中にマーカを設置する必要があるため, デスクトップ環境などの比較的狭い範囲での AR に適している一方で, 広域屋内環境をカバーするためのコストが高いという課題がある.

マーカを用いない画像ベースの位置姿勢推定手法も提案されている. より任意な視点での位置姿勢推定のためには, 撮影画像に含まれる画像特徴と実環境の 3次元モデルの構造とを対応付けるような 3次元モデルベースの手法が有効である ¹⁸⁾¹⁹⁾. この場合も, 広域屋内環境への適用のためには, 基盤データベース (3次元モデル) 構築が必須となるが, 近年, 効率的に実環境の 3次元形状を復元する手法が多く提案されている. 例えば, 広範囲を正確に復元する方法として測距センサを取り付けたロボットを用いた手法は, 普及までの課題はあるが今後の発展が期待されている ²⁰⁾²¹⁾. また, カメラさえあれば適用可能な画像ベースの 3次元モデル構築手法 ²²⁾²³⁾ は, 上述の基盤データベースの共創的な整備を実現するための有望な手法の 1つである. Figure 2 はインタラクティブ屋内 3D モデラ ²³⁾ を用いて生成した仮想化現実モデルと仮想化現実モデルから生成した画像の一例である.

本節で述べた位置姿勢計測または推定技術にはそれぞれ問題点や課題があり, 1つの手法や単純な方法論ですべての状況や要求に対応するのは現状では非現実的である. そのため, 位置や姿勢に基づく情報サービスを, 広域屋内環境において妥当なコストと精度で実現するための現実的な手段の 1つとして, 慣性航法に代表される相対測位と絶対位置・姿勢計測との連携 ²⁴⁾ が注目されている. そこで本研究では, 歩行者推測航法 (PDR) ²⁵⁾ と, 携帯端末内蔵のカメラで撮影される実画像と生成画像との比較に基づく絶対位置姿勢推定との連携についての検討を想定し, インタラクティブ屋内 3D モデラ ²³⁾ で作成した 3次元モデルと実画像とを対応付けるこ



Figure 2: Upper: An overview of the created model. Center: A photo used to create a model. Lower: An image generated from 3D model.

とでカメラの位置・姿勢を推定する。

3 システムの概要

歩行者推測航法 (PDR)²⁵⁾ と、携帯端末内蔵のカメラで撮影される実画像と生成画像との比較に基づく絶対位置姿勢推定の二つの計測手法により構成される位置・姿勢推定システムについ

て述べる。本システムでは、常に PDR により相対位置と姿勢を推定しているため、計算コストの低減や物理的なインフラや基盤データベース整備コストの低減と、位置姿勢情報の常時更新とを両立させることができる一方で、相対測位手段には誤差を累積させてしまうという問題点がある。また、初期値として絶対位置情報を与えておく必要もある。そのため、絶対測位が必要とされる。PDR システムとカメラ位置・姿勢推定システムとの連携の概要を Figure 3 に示す。PDR による相対位置・姿勢測位を常時動かしつつ、携帯カメラからの入力画像から絶対位置・姿勢を推定した結果で PDR の位置・姿勢を補正することで、相対測位による誤差の蓄積を低減する。実画像と 3 次元モデルとの比較に基づく絶対位置方位推定は、携帯カメラからの入力画像をサーバに無線送信し、サーバ側で仮想化現実モデルからキーフレームの特徴点の 3 次元座標を得て、位置・姿勢を推定するといったクライアント・サーバ方式を採用している。

4 仮想化現実モデルを用いたカメラの位置姿勢推定

本研究では 3 次元モデルの作成に用いた画像を用いることで、画像の撮影位置・姿勢及び各画素に対応するオブジェクトの 3 次元座標が既知なキーフレームを作成し、このキーフレームと実画像と比較することでカメラの位置・姿勢を推定する。カメラの位置・姿勢を推定する手法として、特徴点を用いた手法、エッジを用いた手法、面の情報を用いた手法の 3 種類の手法を研究対象として取りあげ実験を行った。以下では、それぞれの手法について詳しく述べる。

4.1 特徴点を用いた手法

特徴点を用いたカメラの位置・姿勢の推定手法の処理の流れを Figure 4 に示す。はじめに、事前準備として画像の撮影位置・姿勢、及び各画素の 3 次元座標の既知な画像をキーフレームとして用意する。次に、入力画像とキーフレーム画像それぞれから特徴点を抽出し、対応付けを行い、入力画像の特徴点の 2 次元座標とその点に対応付けられたキーフレームの特徴点の 3 次元座標の組が得られる。最後に、複数の 2 次元

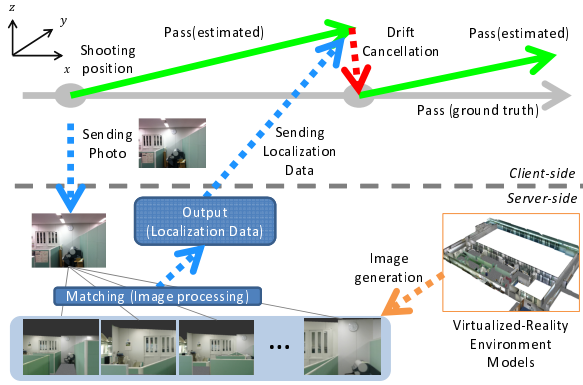


Figure 3: Cooperation among model based tracking system and PDR system

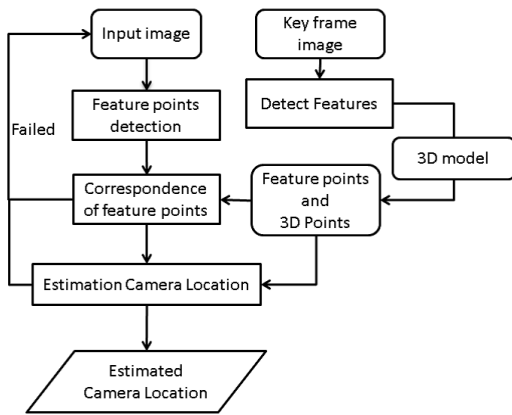


Figure 4: Flow chart of the estimation using feature points

座標と3次元座標の組からRANSAC²⁷⁾によるロバスト推定を用いてカメラの位置・姿勢を推定する。

4.2 エッジを用いた手法

エッジを用いたカメラの位置・姿勢の推定手法の処理の流れをFigure 5に示す。本研究で用いる仮想化現実モデルは、テクスチャ付きのポリゴンによって構成され、各ポリゴンのエッジは構造上のエッジかどうかの属性を持つ。構造上のエッジの例をFigure 6に示す。本手法では、はじめに、入力画像からエッジ検出をし、モデルから生成される構造上のエッジを取得した。次に、エッジの投影に用いるカメラパラメータを複数生成し、モデルから取得した構造上のエッ

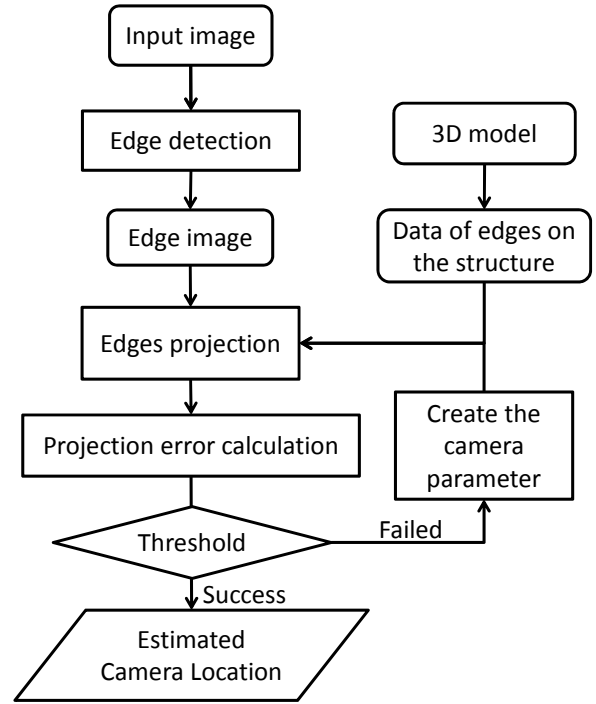


Figure 5: Flow chart of the estimation using edge information

ジを入力画像上へ投影した。最後に、検出したエッジと投影したエッジの投影誤差をそれぞれのカメラパラメータで評価し、閾値以上で投影誤差の最も小さいものをカメラの位置・姿勢の推定結果とした。以下に投影誤差の計算手順を述べる。

エッジ投影時のカメラパラメータ P_t の時を考える。

1. カメラパラメータ P_t で投影した構造上のエッジ E_i を一定長さごとに分割し、サンプル点 $e_{i,j}$ を得る (Figure 7)。
2. サンプル点 $e_{i,j}$ から構造上のエッジ E_i に垂直な方向の入力画像の輝度勾配を探索する。
3. 輝度勾配が閾値以上の極値をエッジ候補 $e_{i,j}$ としてすべて検出する。
4. 検出したエッジ候補とサンプル点の距離の2乗を $\Delta_t(E_i, e_{i,j,k})$ とする。

$\Delta_t(E_i, e_{i,j,k})$ を各構造上のエッジについて計算し、複数の誤差の値を統合してエッジの投影



Figure 6: Upper: An example of the image. Lower: Examples of the edges on the structure.

誤差とする．さらに，カメラパラメータ P_t におけるエッジの評価値を v_t として算出する．ただし， c はデータの標準偏差である．

$$v_t = \frac{1}{N} \sum_i \sum_j \rho_{Tuk}^*(\Delta_t(E_i, e_{i,j,1}), \dots, \Delta_t(E_i, e_{i,j,K})) \quad (1)$$

$$\rho_{Tuk}^*(x_1, \dots, x_n) = \min_i \rho_{Tuk}(x_i) \quad (2)$$

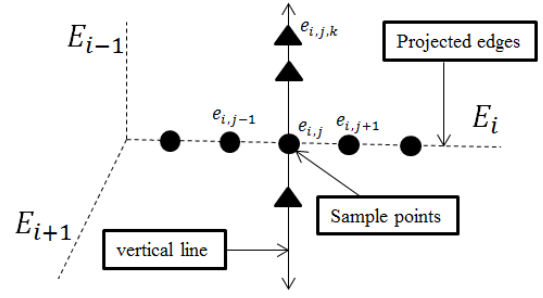


Figure 7: Sample points on the edges

$$\rho_{Tuk}(x) = \begin{cases} \frac{c^2}{6} (1 - (1 - (\frac{x}{c})^2)^3) & \text{if } |x| \leq c \\ \frac{c^2}{6} & \text{if } |x| > c \end{cases} \quad (3)$$

複数の評価値の統合

本手法では，検出したエッジと投影したエッジの投影誤差を評価する際に，複数のエッジを投影し，それぞれのエッジで投影誤差が計算される．そのため，各群のデータ数の違う複数のデータを評価する必要があり，そのようなデータ群を対象に評価する際に有効な手法としてテューキー法がある¹¹⁾．テューキー法はデータ群とデータ群を多重比較する手法で，式 (3) で定義される．テューキー法による評価値をエッジの投影誤差とし，評価値が最も小さくなるカメラパラメータを推定結果とした．

4.3 面を用いた手法

面の情報として相互情報量と呼ばれる特徴量が提案されている¹²⁾．これは位置合わせのための特徴量でテンプレートマッチングやトラッキングに用いられる．相互情報量を用いたカメラの位置・姿勢の推定手法の処理の流れを Figure 8 に示す．はじめに，計算で用いるエントロピーと結合エントロピーについて説明する．

エントロピー

エントロピーは画像の情報量を示す指標で，ある画像 I において輝度値が $0 \sim (k-1)$ の k 値画像について考える．ある輝度値 r の出現確率を P_r としてエントロピー $h(I)$ は以下の式 (4) で表

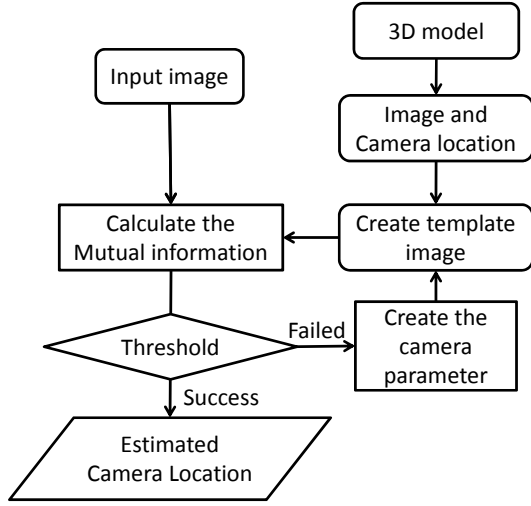


Figure 8: Flow chart of the estimation using mutual information

される.

$$h(I) = -\sum P_r \log_2 P_r \quad (4)$$

結合エントロピー

同じ解像度の画像 $I1$ と画像 $I2$ を考える. それぞれの画像の同一座標の画素の輝度値を比べて画像 $I1$ が輝度値 r , 画像 $I2$ が輝度値 t の組み合わせを取る確率 $P_{I1,I2}(r,t)$ は以下の式 (5) で表される. 結合エントロピー $h(I1, I2)$ は確率 $P_{I1,I2}(r,t)$ を用いて式 (6) で表される.

$$P_{I1,I2}(r,t) = \frac{1}{N} \sum_X \phi(r - I1(x)) \phi(t - I2(x)) \quad (5)$$

$$h(I1, I2) = -\sum_r \sum_t P_{I1,I2}(r,t) \log_2(P_{I1,I2}(r,t)) \quad (6)$$

ただし, N は画像の全画素数, $\phi(x)$ は以下の式 (7) で表される.

$$\phi(x) \begin{cases} 1(x=0) \\ 0(x \neq 0) \end{cases} \quad (7)$$

$I1(x)$, $I2(x)$ は画像 $I1$, $I2$ においてある座標 x における画素の輝度値を示す.

相互情報量

相互情報量 $MI(I1, I2)$ は前述のエントロピーと結合エントロピーを用いて,

$$MI(I1, I2) = h(I1) + h(I2) - h(I1, I2) \quad (8)$$

画像 $I1$ の輝度値が r を取る確率を $P_{I1}(r)$, 画像 $I2$ の輝度値が t を取る確率を $P_{I2}(t)$ として,

$$MI(I1, I2) = \sum_r \sum_t P_{I1,I2}(r,t) \log \left(\frac{P_{I1,I2}(r,t)}{P_{I1}(r)P_{I2}(t)} \right) \quad (9)$$

$$P_{I1}(r) = \frac{1}{N} \sum_X \phi(r - I1(X))$$

$$P_{I2}(t) = \frac{1}{N} \sum_X \phi(t - I2(X))$$

と表される. 本手法では, 式 (9) で表される値 $MI(I1, I2)$ を用いた値を相互情報量の評価値とする.

本手法では, カメラパラメータを複数生成し, それぞれのカメラパラメータで投影したテンプレート画像と入力画像を用いた相互情報量の評価値を計算し, 評価値が最大値となるカメラパラメータを位置・姿勢の推定結果として用いた.

5 実験

本研究では, はじめに, 生成画像を用いて大域的な絞り込みが可能であるか, また, 生成画像と実画像間で特徴点の対応付けが可能であるかを調査するための予備実験として, 実画像と生成画像を対応付ける際に, 色相・彩度 (HS) を用いたヒストグラムの相関, SSD (Sum of Squared Difference), 及び ZNCC (Zero-mean Normalized Cross-Correlation) の計 3 種類の特徴量を類似度尺度として用いた場合の有効性を比較評価した. また, 実画像と生成画像それぞれで SIFT 特徴に基づいて特徴点の対応付けを行い, 実画像と生成画像の相対的な位置・姿勢を推定した. 生成画像のカメラパラメータは既知であるので, これによりモデル座標系での撮影位置・姿勢が推定される.

続いて, 本実験では, 実画像とキーフレームを用いたカメラの位置・姿勢推定精度が AR システムに適用可能か, 並びに複数の特徴量を組み合わせることにより精度の向上が見込めるかを調べ

るため、特徴量として、SURF、ORB、FREAK、エッジ、相互情報量を用いてカメラの位置・姿勢推定を行った。特徴点を用いた手法において、予備実験で扱った SIFT 特徴量は処理時間が長く、実時間処理が困難であると判断したため、SIFT 特徴量よりも高速な特徴量である SURF、ORB、FREAK について調査した。

5.1 生成画像と実画像によるカメラの位置・姿勢推定

5.1.1 入力画像と生成画像

本評価では、入力画像として、Apple 社の iPhone4 で撮影した画像 (2592×1936 pixels) を 4:3 になるようトリミングし、 640×480 に縮小したものを用いる。生成画像 (640×480 pixels) は、モデル座標系でのカメラパラメータ (仮想カメラ) を設定し、描画することで生成する。ここでは、実験を単純化するため、仮想カメラの高さを 160cm に固定し、実画像もほぼ同じ高さ (身長 170cm 前後のユーザが眼前にカメラを構えていることを想定) で撮影した。また、仮想カメラの姿勢変化については、ロールとピッチは固定し、水平方向の回転のみ、つまり、方位 (ヨー) のみとした。その上で、位置は水平方向の各軸 100cm 刻み (範囲: 各軸 ± 300 cm), 方位は、類似度の評価では 5 度刻み (範囲: 360 度), SIFT の評価では 5 度刻み (範囲: 正解方位を中心に ± 60 度) で仮想カメラの位置・姿勢を設定し、実験用の生成画像を得た。SIFT の範囲を ± 60 としたのは、方位が大きく違う 2 つの画像上に同じ物が写ることはほとんどなく、評価で扱う SIFT による対応付けで正解対応の数が 0 となることが明白であるためである。Figure 9 に本実験で用いた筆者の居室の仮想化現実モデルを示す。また、Figure 10 に、入力画像、並びにその撮影位置・方位に近い仮想カメラからの生成画像、及び画像処理の対象領域を制限するためのマスク画像を示す。本研究で用いる仮想化現実モデルは、テクスチャ付きのポリゴンで構成され、各ポリゴンは壁、床などの属性を持つ。本実験では、モデルのない領域の類似度計算への悪影響を回避するため、壁のみを対象にした画像処理を行うようにマスク画像を設定した。

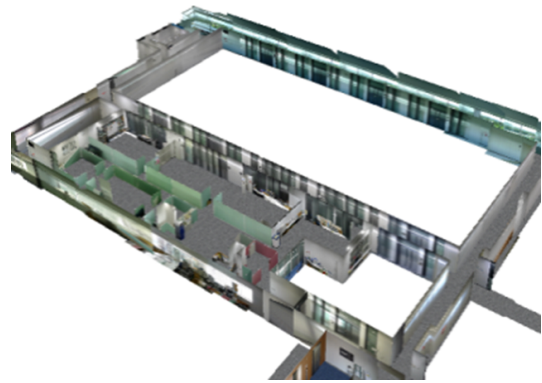


Figure 9: An example of the virtualized reality model

5.1.2 類似度に基づく推定結果の評価方法

ここでは、画像の類似度に基づくカメラの位置・姿勢の推定、及びその結果の評価方法について述べる。

はじめに、入力画像と生成画像から各類似度尺度を用いて類似度を計算する。次に、類似度の高い生成画像を閾値処理により選択する。選択された生成画像群の中にカメラパラメータが入力画像に近い画像が含まれる必要があるため、本実験ではこの閾値と絞り込まれた生成画像の分布との関係性を評価した。

5.1.3 SIFT に基づく推定手法の評価方法

ここでは、SIFT 特徴に基づくカメラの位置・姿勢の推定、及びその結果の評価方法について述べる。

はじめに、実画像と生成画像それぞれで SIFT 特徴に基づいて特徴点を抽出する。次に、文献²⁶⁾の方法を用いて、実画像、生成画像のそれぞれから抽出された特徴点の対応付けを行う。対応付けの結果に基づいて生成画像を選択し、入力画像と生成画像の相対的な位置・姿勢を推定する。生成画像のカメラパラメータは既知であるので、これにより実画像での撮影位置・姿勢が推定される。

ここでは、SIFT 特徴を実画像、生成画像のそれぞれに適用し対応付けを行う際の振る舞いについて確認する。本実験では、対応付けの精度を、対応付けられた入力画像中の特徴点と生成画像中の特徴点それぞれのモデル座標系での三

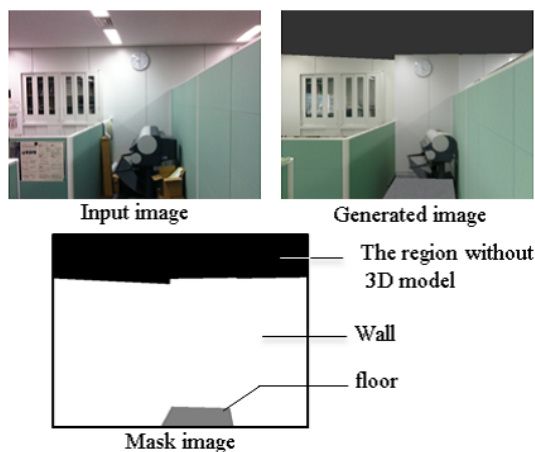


Figure 10: Examples of the input image, generated image and mask image

次元座標のユークリッド距離を用いて評価した。

入力画像中の特徴点の三次元座標を得るために、以下の手順で三次元座標を取得した。(1) 手で仮想カメラを変化させながら、入力画像と見えがなるべく一致するように生成画像(評価用生成画像)とそのデプスマップ (Figure 11) を生成し、(2) そのデプスマップを用いて入力画像の各画素の三次元座標を取得した。ただし、仮想化現実モデルの精度や評価用画像の調整が手動であることなどを考慮し、各画素の周囲に Figure 12 に示す探索領域と探索点を設定し、探索点ごとに生成画像中の特徴点とのモデル座標系での 3 次元座標のユークリッド距離を求め、その最小値を用いて、閾値処理による正誤判定を行った。

5.1.4 実験結果と考察

類似度の評価

Figure 13 は、HS ヒストグラム、SSD、ZNCC の 3 種類の特徴量を用いて入力画像と生成画像との類似度を計算した結果である。入力画像は、A,B,C の計 3 地点においてそれぞれ図中に示された方位で画像を 1 枚撮影することで得た。Figure 13 はそれらと生成画像との類似度の分布マップとなっており、図中の各円は、その中心位置で仮想カメラの方位を 360 度変化させた場合の類似度を示している。円と類似度の提示方法の詳細を Figure 14 に示す。

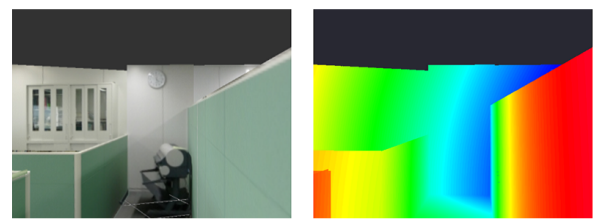


Figure 11: Left: Generated image, Right: Depth map image

Figure 15 に、類似度を用いた閾値処理により選択された生成画像数の割合(絞り込みの効果)を示す。横軸は類似度を示し、縦軸は生成画像数の割合を示す。グラフ中の数値は実画像の位置・姿勢と同じ設定の仮想カメラで生成した生成画像(以後、正解生成画像)の類似度と、その類似度で閾値処理した際に残る生成画像数の割合である。

正解生成画像を含むように類似度による閾値処理を行うには閾値をこの正解生成画像の類似度よりも低くしなければならない。そこで、各特徴における閾値の最小値を調べた結果と、HS ヒストグラムの場合は、A 地点: 43%, B 地点: 69%, C 地点: 40%, SSD の場合は、A 地点: 24%, B 地点: 67%, C 地点: 97%, ZNCC の場合は、A 地点: 10%, B 地点: 7%, C 地点: 96% となった。

この評価実験では、ZNCC を類似度とした場合に比較的良好な結果が得られたため、ZNCC について D, E, F, G, H, I 地点 6 カ所での追加実験を実施した。撮影位置は D,E,F 地点は A 地点付近、G,H 地点は B 地点付近、I 地点は C 地点付近である。方位は、それぞれ、A,B,C 地点の付近で方位を変更して撮影した。Figure 16 に Figure 15 と同様の形式のグラフを示す。ZNCC を類似度とした結果で、D 地点: 15%, E 地点: 9.5%, F 地点: 37%, G 地点: 1%, H 地点: 10%, I 地点: 19% であった(9 地点の平均: 22.7%, 最小値: 1%, 中央値: 10%, 最大値: 96%)。

実験の結果、ZNCC を類似度とした場合、ある程度効果的に推定範囲の絞り込みができることがわかった。しかし、C 地点では閾値の最小値が 96% となり、正解生成画像の類似度で閾値処理すると絞り込むことが困難である結果となっ

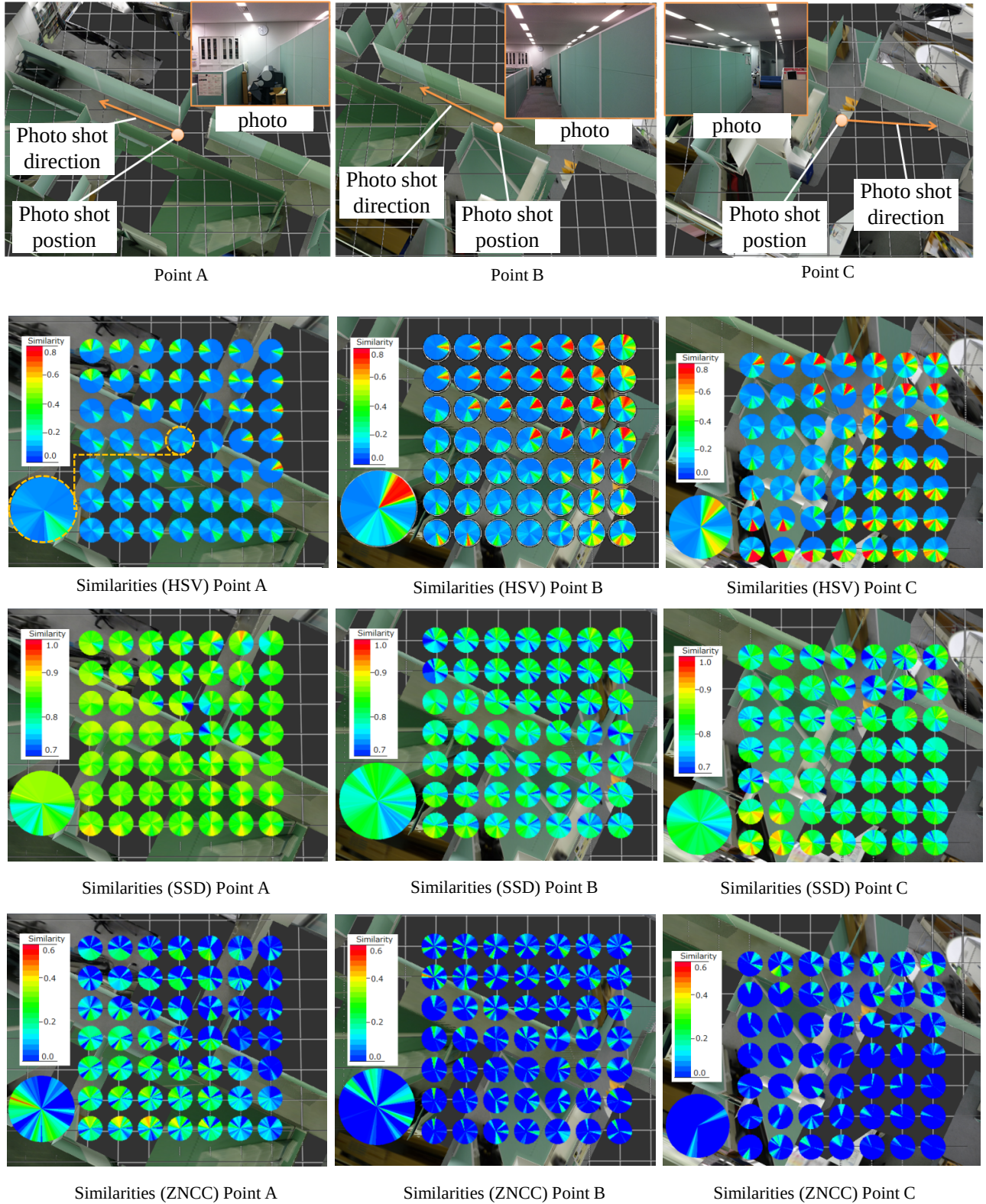


Figure 13: An experimental results of three types of similarities (Upper: Hue-saturation histogram-based, Center: SSD-based, Lower: ZNCC-based)

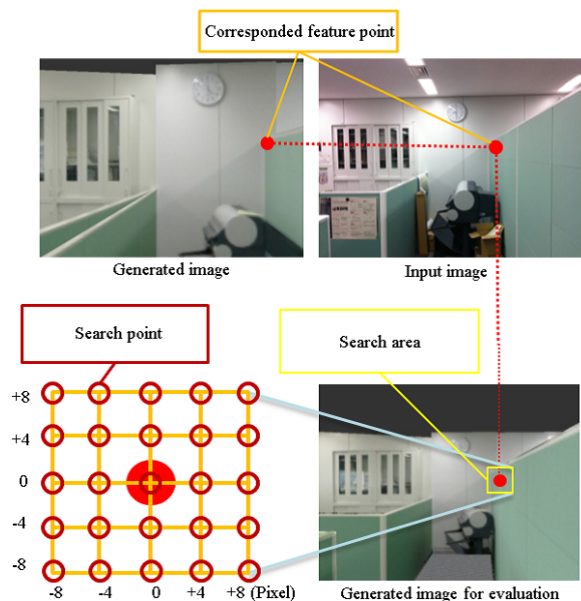


Figure 12: Search points and a search area for SIFT evaluation

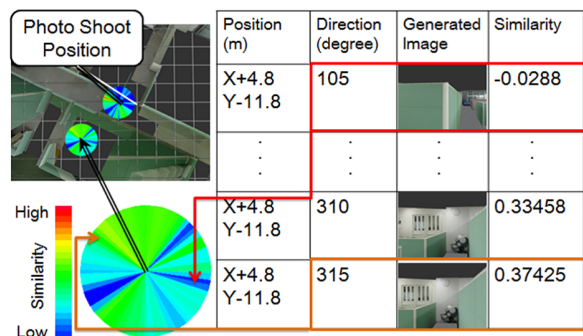


Figure 14: A circle show the position and orientation of a camera used to generate an image from models. Center of the circle shows the camera position. Each sector indicates an orientation of a camera orientation and a similarity.

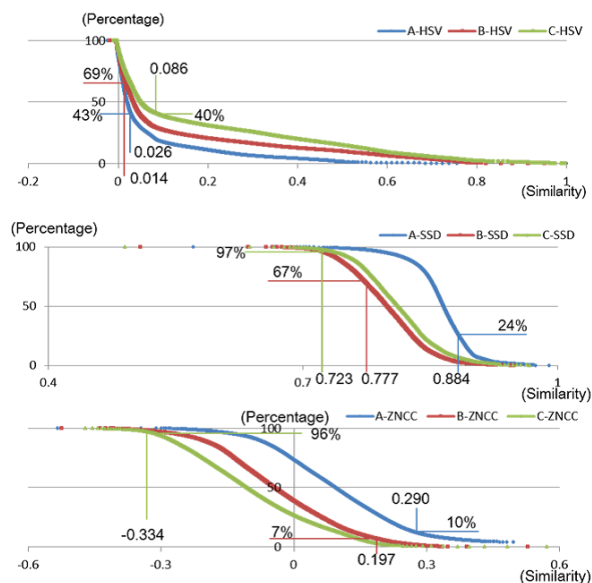


Figure 15: Relation between the threshold of the similarity and the percentage of the number of the estimation.

た。この原因として、C 地点ではガラスなどの反射や透過をする部分がテクスチャとして使われていたことが考えられる。このような部分は視点によって見え方が大きく異なってくる。そのため、今後の展望として、例えば、モデルの各ポリゴンの属性として、壁、床等の他に、ガラス、窓、鏡といった透過や鏡面反射にする属性を追加し、適応的にそのような領域を扱うことで絞り込みの性能を向上させることができると考えられる。

SIFT の評価

Figure 17 は、実画像と生成画像のそれぞれで抽出された特徴点の対応付けを行い、対応付けが正しいと判定された特徴点对の数を、A、B、C の計 3 地点で調べた結果である（正誤判定の閾値：30cm）。ここでは、入力画像と生成画像の相対的な位置・姿勢を実際に推定は行っていないが、C 地点を除き全ての地点において推定に必要な数の特徴点对が得られた。一方、C 地点の正解生成画像では位置・姿勢の推定に十分な数の特徴点对が得られず、異なる視点の生成画像の方が特徴点对の数が多かった。他の 2 地点では正解生成画像でもある程度は特徴点对が得られたが、

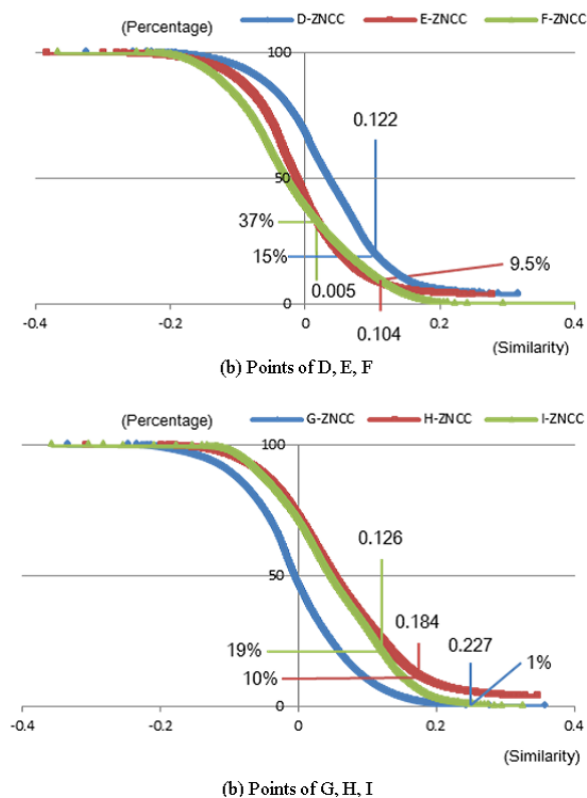


Figure 16: Relation between the threshold of the similarity (ZNCC) and the percentage of the number of the estimation.

やはり正解生成画像以外でより多くの特徴点対が得られた。なお、正しい特徴点対数は最大で 14 であった。本手法では、RANSAC(RANdom SAmple Consensus)²⁷⁾ 等、ロバスト推定によって位置・姿勢を得ることを想定しているが、誤対応が多く、正しい対応が少ないと正しい推定が困難となる。この評価実験では、SIFT に基づいた特徴点の抽出を行い、その対応付けの評価をした。その結果、誤対応が多く、正しい推定が困難であると考えられる。そのため、今後の展望として、生成画像中の特徴点（局所特徴）を扱う方法が実画像と同じでよいのかどうかの検討や、より積極的な仮想化現実モデルの持つ情報の活用等が必要となる。

5.2 キーフレームによるカメラの位置姿勢推定

本実験では、特徴量として、特徴点、エッジ、面の情報を用いて作成したキーフレームと入力画像を比較することでカメラの位置・姿勢を推

定した。推定した位置・姿勢の評価を行うためには、入力画像を実際に撮影した正しい位置・姿勢という真値が必要となる。しかしながら、物理的に距離や姿勢を正確に測るのは難しいため、本実験では真値に準ずる値として参照値を作成した。以下では、まず、本実験で用いた入力画像について述べる。

5.2.1 入力画像

本実験では、入力画像として 3 つのシーンを用意し、それぞれのシーンで入力画像とキーフレームを用いてカメラの位置・姿勢を推定した。各シーンに含まれる画像の例を Figure 18 に示す。以下、各シーンの入力画像とキーフレームについて述べる。

シーン 1 は、入力画像として施設のモデルから生成した生成画像（解像度 640*480 ピクセル）44 枚を用いた。キーフレームを作成するための画像も同様の生成画像を用いて、インタラクティブ 3D モデラ²³⁾を用いてモデリングを行った。キーフレームは 10 フレーム目の画像を用いて作成した。

シーン 2, 3 は、入力画像として Apple 社の iPad2 で撮影した画像（解像度 768*1024 ピクセル）を 180*240 に縮小したものをを用いた。シーン 2 はプラスチック製の箱を積み上げたものを周囲から iPad2 を用いて撮影した画像 34 枚、シーン 3 は異なる角度からプラスチック製の箱を撮影した画像 20 枚を用いた。シーン 2 のキーフレームは、シーン 2 の 10 フレーム目の画像のカメラパラメータ付近の画像を用いて作成し、シーン 3 のキーフレームは、シーン 3 の 7 フレーム目の画像のカメラパラメータ付近の画像を用いて作成した。シーン 2, 3 では、AR アプリケーションへの実利用を想定し、iPad2 からネットワークを介してサーバへと画像を送信する際に送信レートを高く保つため画像の縮小を行った。キーフレームを作成するための画像も iPad2 で撮影し同様に縮小したものをを用い、インタラクティブ 3D モデラ²³⁾を用いてモデリングを行った。実スケールに合わせたモデルを作成することで、モデリングに使用した画像の撮影位置と画像中の各画素の 3 次元座標を実環境と対応付ける事ができる。

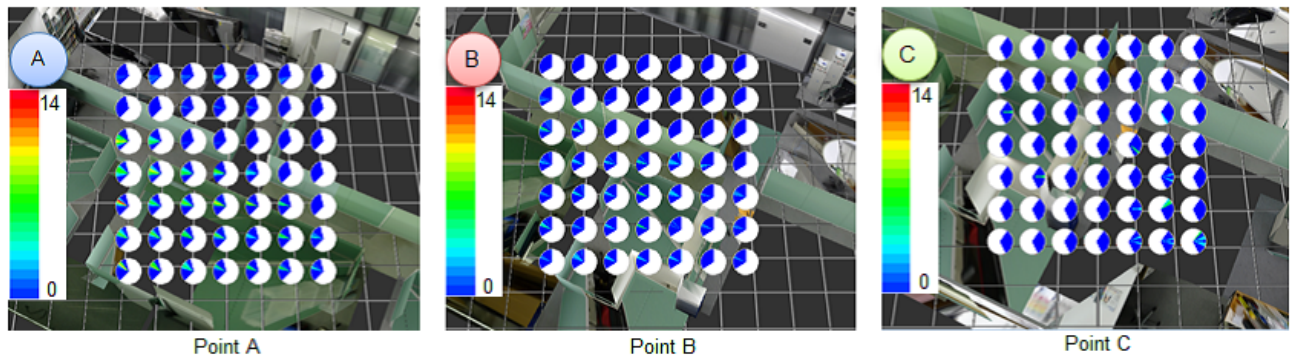


Figure 17: The number of corresponded points with SIFT features (A circle show the position and orientation of a camera used to create a generated image from models. a center of circle shows the camera position. Color of each section shows the number of the corresponded points.).

5.2.2 データベースの作成とその利用

本実験では、特徴点、エッジ、面の情報を用いてカメラの位置・姿勢推定を行った。使用する特徴量に応じてモデルから情報を得てデータベースを作成した。以下、データベースの作成手順と各特徴量を用いたカメラの位置・姿勢推定の方法についてそれぞれ述べる。

特徴点を用いた推定手法

特徴点の検出では SURF⁷⁾ と ORB⁸⁾ を用いた。また、ORB で特徴点を検出して FREAK⁹⁾ アルゴリズムで記述した際の推定結果についても実験を行った。以下に、データベースの作成手順を示す。

1. キーフレームとなる画像を iPad2 で撮影し、縮小した画像を用いてモデルを作成
2. 撮影した画像の撮影位置・姿勢をモデルから取得
3. キーフレームとなる画像から特徴点を検出し、その 2 次元座標を取得
4. 2 次元座標に対応した実環境における 3 次元座標をモデルから取得

以上で得た、画像、撮影位置・姿勢、特徴点の 2 次元座標と 3 次元座標、及び特徴量のセットをキーフレームとしてデータベースを作成した。

次に、入力画像から特徴点抽出を行い、キーフレームとして登録された特徴点データと対応

付けをし、RANSAC²⁷⁾ を用いてカメラのパラメータを推定した。

エッジを用いた推定手法

以下に、エッジを特徴量として用いる際のデータベースの作成手順を示す。

1. 任意のカメラで画像を撮影し、モデルを作成
2. 撮影した画像の撮影位置・姿勢をモデルから取得
3. モデルから構造上のエッジを自動抽出し、エッジの両端の 3 次元座標を取得
4. 以上で得た、構造上のエッジをデータベースに登録した。また、ここで得た画像の撮影位置・姿勢は構造上のエッジを投影する際のカメラパラメータの基準とした。

入力画像のエッジ検出には、Canny エッジ検出器¹⁰⁾ を用いた。検出したエッジとモデルから取得した構造上のエッジをキーフレームの撮影位置・姿勢を基準に、モデル座標系で位置を X 軸、Y 軸、Z 軸の各方向に $\pm 100\text{mm}$ の範囲を刻み幅 50mm 、姿勢を X 軸、Y 軸周りに ± 2 度の範囲を刻み幅 1 度ずつ変更したカメラパラメータを生成し、そのカメラパラメータによって投影されたエッジと入力画像で検出したエッジの投影誤差を計算したものを評価値とし、評価値が最も小さくなるカメラパラメータを推定結果とした。

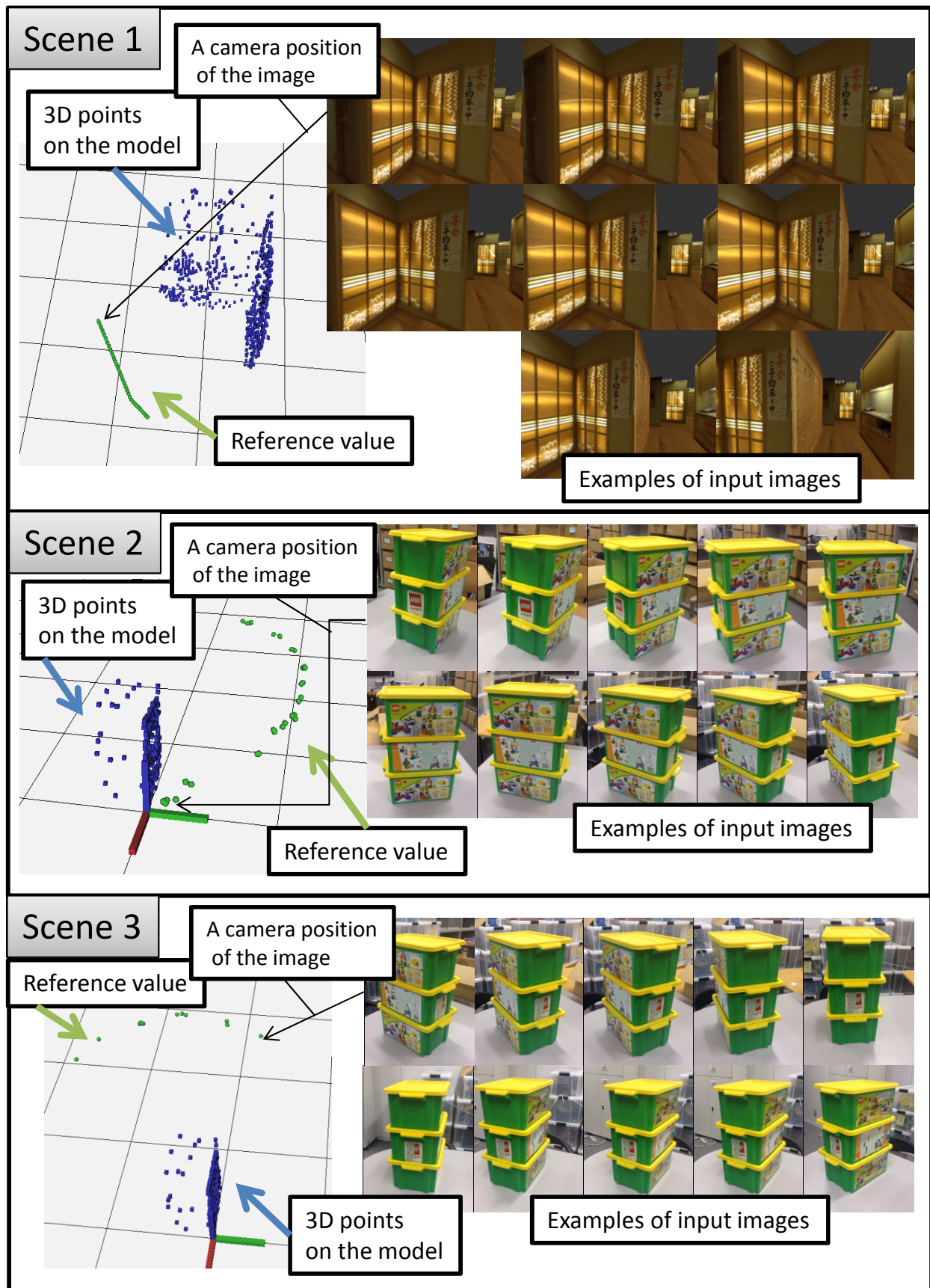


Figure 18: Examples of input images of each scene and reference values . (Upper: Scene1, Center: Scene2, Lower: Scene3.)

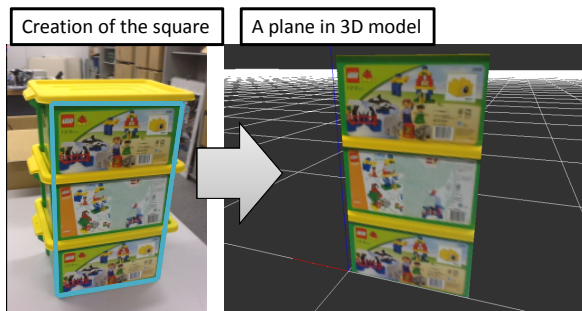


Figure 19: An overview of making a template

面を用いた推定手法

以下に、相互情報量を特徴量として用いる際のデータベースの作成手順を示す。

1. テンプレートとなる画像を iPad2 で撮影し、縮小した画像を用いて 3D モデルを作成
2. 撮影した画像の撮影位置・姿勢をモデルから取得
3. 画像中のモデリングされた四角形の平面領域を選択 (Figure 19)
4. 3. で選択した四角形の 4 隅の 3 次元座標を 3D モデルから取得
5. 以上で得た、画像、撮影位置・姿勢、テンプレートの 4 隅の 3 次元座標のセットを一つのキーフレームとしてデータベースを作成

次に、面の情報を用いた推定の際のテンプレートの作成、及びカメラパラメータの推定方法を述べる。本実験では、エッジの場合と同様にカメラパラメータを生成した。さらに、そのカメラパラメータからキーフレーム画像を射影変換してテンプレートを作成した。最後に、入力画像とテンプレートから相互情報量を計算して評価値とし、評価値が最も大きくなるカメラパラメータを推定結果とした。

5.2.3 参照値の作成

本実験では、一つの入力画像に対して特徴点、エッジ、面の情報を用いて作成したデータベースを利用してそれぞれの手法について一つのカメラの位置・姿勢を推定した。そこで、推定したカメラ位置・姿勢の精度評価をするために、各入力画像に対してカメラ位置・姿勢の参照値を

作成し、位置の誤差、姿勢の誤差、仮想物体の投影誤差についての評価を行った。

参照値の作成には 3D モデルを利用する。入力画像中で同一平面上の 2 次元座標に対応したモデル座標系の 3 次元座標を手動で 4 組取得する。4 つの 2 次元座標と 3 次元座標の組みから P4P 問題を解くことでカメラパラメータを求める²⁸⁾。ただし、4 組の 3 次元座標が同一平面上にあり、かつその 3 点以上が同一直線上に存在しない点を用いる。

5.2.4 実験結果と考察

本実験では、推定されたカメラ位置・姿勢と参照値を比較して、位置の誤差、姿勢の誤差、仮想物体の投影誤差について評価した。位置の誤差は、推定位置と参照値の位置のモデル座標系におけるユークリッド距離で算出した。姿勢の誤差は、推定された姿勢と参照値の姿勢からそれぞれのモデル座標系におけるカメラの視線方向のベクトルを計算し、2 つのベクトルのなす角度に比例する値²⁹⁾とした。仮想物体の投影誤差については、はじめに、参照値の画像を 3*3 の 9 つの領域に分割し、各領域の中心に投影される仮想点を設定した上で、入力画像上に仮想点を投影をした。次に、両画像上における仮想点の二次元座標を求め、2 点間の距離 (ピクセル) を仮想物体の投影誤差の値とした。Figure 20 に、仮想点、参照値の画像、入力画像、それぞれの関係を示す。以下では、各シーンにおける位置、姿勢、仮想物体の投影誤差の評価結果について述べる。ただし、グラフ中、線が存在しない部分は推定が失敗した事を示し、入力画像全フレームに対して各特徴量を用いてカメラの位置・姿勢推定を行った。

シーン 1 における位置、姿勢、仮想物体の投影誤差の評価結果を Figure 21-(B) に示す。特徴点を用いた手法では多くのフレームでカメラの位置・姿勢が推定でき、位置の誤差や仮想物体の投影誤差も小さい。推定の成功したフレーム数は、SURF が 30 フレーム、ORB が 32 フレーム、FREAK が 43 フレームと FREAK を用いた手法が最も多かった。しかしながら、姿勢誤差と仮想物体の投影誤差を見ると特徴点を用いた手法では、FREAK のみ 40 フレーム以降で誤差

が大きくなっている事がわかる。

シーン 2 における位置，姿勢，仮想物体の投影誤差の評価結果を Figure 22-(B) に示す．特徴点を用いた手法において，推定に成功したフレーム数を見ると，SURF で 5 フレーム，ORB で 1 フレーム，FREAK で 0 フレームであった．ORB の位置の誤差を見ると，推定位置が参照値から大きく離れていることが分かる．しかしながら，投影誤差は SURF とほとんど変わらない．これは，ORB でのマッチングが誤対応を多く含んでいたためであると考えられる．エッジ及び相互情報量を用いた場合の位置の誤差と姿勢の誤差が 10～15 フレーム付近で低くなり，離れる程高くなっているのは，テンプレートやエッジの投影を行う際のカメラパラメータの基準位置が 10 フレーム付近のカメラパラメータに近いためと考えられる．相互情報量の評価値とエッジの評価値の投影誤差との関係については後述する．

Figure 23-(B) に，シーン 3 における位置，姿勢，仮想物体の投影誤差の評価結果を示す．特徴点を用いた手法では，20 フレームすべてにおいて推定が失敗した．キーフレームにより 3 次元座標が得られるのはモデリングをした物に限られる．そのため，モデルを作成した部分の特徴点が十分な数検出できないと画像全体での特徴点のマッチングができて 3 次元座標の情報を持った特徴点の数が少ないためにカメラの位置・姿勢推定を失敗していると考えられる．背景までモデリングをした場合にはもう少し推定可能範囲が広がる可能性がある．

相対測位との連携を考えた場合，画像処理によって推定されたカメラパラメータを用いて位置・姿勢を補正するため，誤差が大きい推定値は出来る限り出力されない方が望ましい．その観点から考えて，本実験で扱った特徴点を用いた手法の中では，SURF が最も推定に有効である．

エッジを用いた手法はいずれのシーンにおいても誤差が大きいものも推定結果として出力された．これは，入力画像から検出されるエッジに構造上のエッジ以外のエッジが多く含まれており，誤対応が多くなっているためと考えられる．エッジを用いた手法については，構造上のエッジを検出しやすいような手法に変更するな

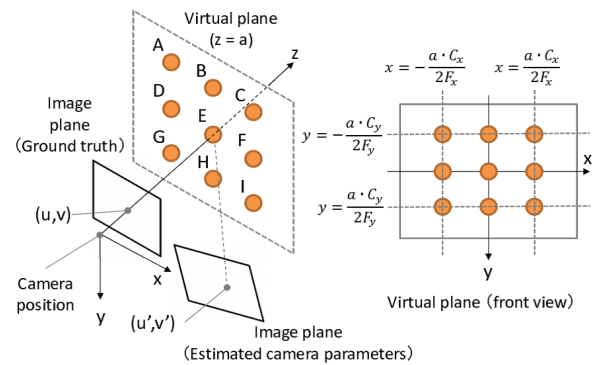


Figure 20: Description of the projection error 30)

ど，改良の余地がある．

相互情報量を用いた手法はシーン 2，シーン 3 において誤差の大きい出力もなく，比較的誤差が小さい．Figure 24 に，シーン 2 及びシーン 3 における仮想物体の投影誤差と相互情報量による評価値の関係を示す．仮想物体の投影誤差が大きい場合に評価値が小さい値を示し，仮想物体の投影誤差が小さい場合に評価値が大きな値を示すことが分かる．そのため，相対測位との連携において，推定したカメラパラメータを用いて位置・姿勢を補正する手法として適していると考えられる．

また，シーン 3 の結果 (Figure 23-(B)) から，特徴点を用いた手法ではカメラのパラメータを推定出来ないシーンにおいて，相互情報量を用いた手法を用いることでカメラパラメータを推定できる可能性があることがわかった．そのため，特徴点を用いた手法と相互情報量を用いた手法を組み合わせることで，特徴点を用いた手法単独でカメラの位置・姿勢を推定するよりも，広い範囲でのカメラの位置・姿勢が可能であると考えられる．

5.3 計算コストについて

Table 1 に各特徴量を用いた推定の計算時間を示す．なお，推定には以下のスペックの計算機 (Windows7, Intel(R) Core(TM)i5 CPU M520 2.4GHz) を用いた．相互情報量及びエッジは，カメラパラメータを生成して一つの評価値が計算されるまでの時間である．そのため，実際にカメラの位置・姿勢推定に必要な時間は，この値

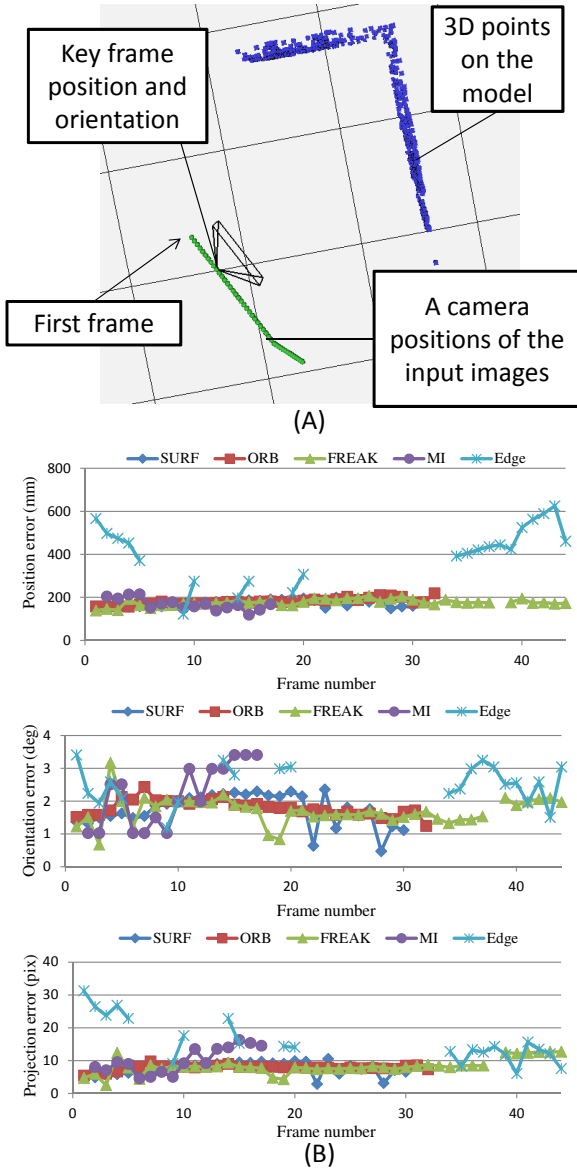


Figure 21: In scene 1. (A) A camera position and orientation of the key frame image. (B) Experimental results of the estimated camera location error using generated image. (Upper: Position error. Center: Direction error. Lower: Projection error.)

とカメラパラメータの生成数の積となる。探索範囲を広げたり、探索密度を上げるとカメラパラメータの生成数が増加するため、計算時間が増加する。SURF などの特徴点を用いた手法でかかる時間は約 300ms であるため、この時間と同程度の時間を許容すると仮定すると相互情報量を用いて一つのカメラパラメータを評価する

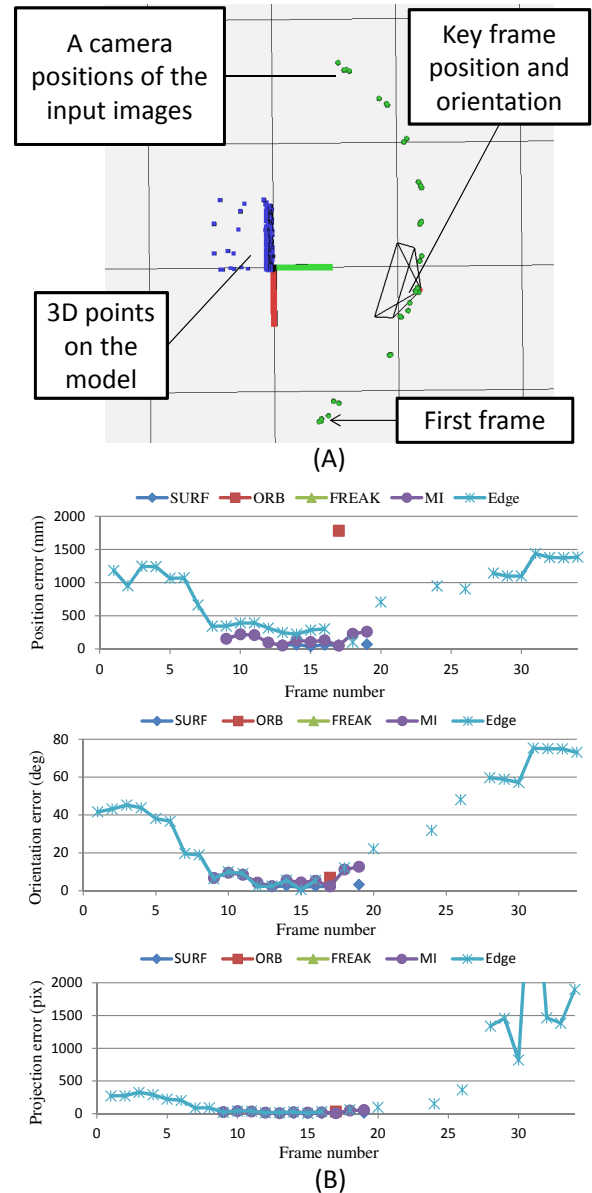


Figure 22: In scene 2. (A) A camera position and orientation of the key frame image. (B) Experimental results of the estimated camera location error using real image. (Upper: Position error. Center: Direction error. Lower: Projection error.)

時間が 27.7ms であるため、単純に計算すると $300/27.7 \simeq 10.83$ となり、約 11 個のカメラパラメータしか評価することができない。ただし、相互情報量を用いて 1 つのカメラパラメータを評価する計算は並列処理可能であり、時間あたりの処理数を増やすことが可能である。並列処理の可能なスレッド数を N とすると、 N 倍の数の

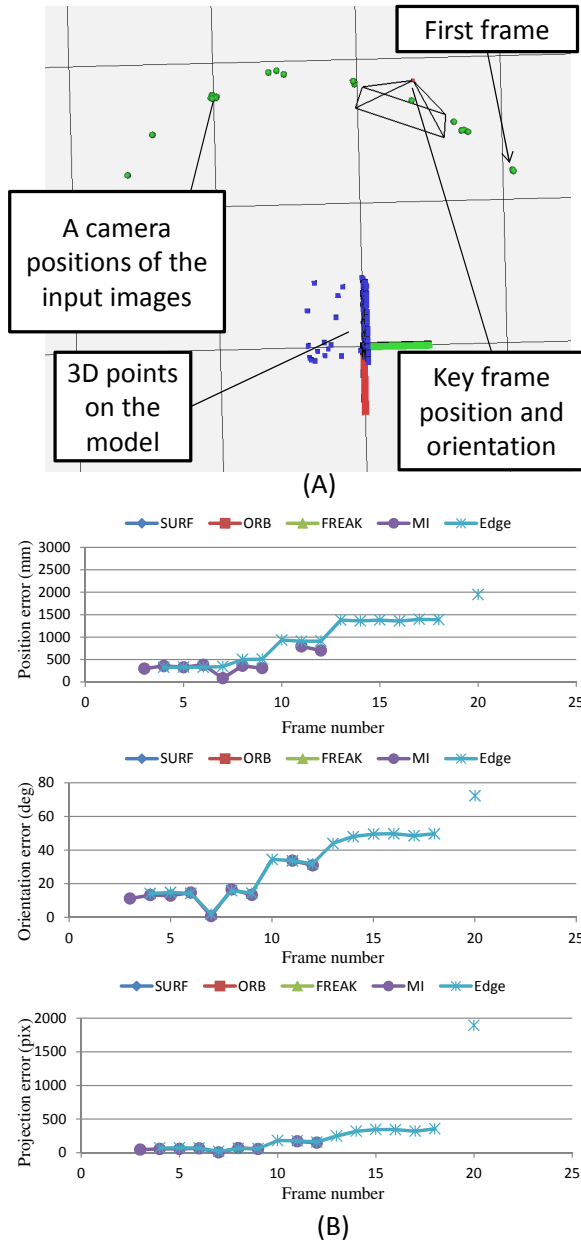


Figure 23: In scene 3. (A) A camera position and orientation of the key frame image. (B) Experimental results of the estimated camera location error using the less feature points scene. (Upper: Position error. Center: Direction error. Lower: Projection error.)

カメラパラメータを評価することが可能になる。そのため、実時間で推定することを考えると、本実験において高い推定精度を得られた相互情報量を用いた手法をそのまま AR アプリケーションなどに適用することは難しいが、並列処理の

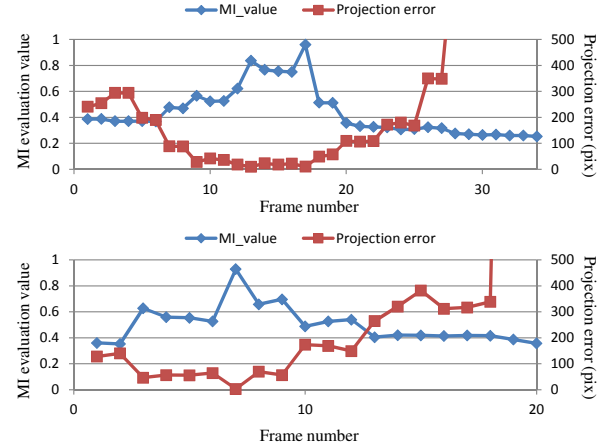


Figure 24: Experimental results of MI evaluation values and projection errors.(Upper: Scene 2. Lower: Scene3.)

ような実装上の改善や計算するカメラパラメータの数を減らすために、位置・姿勢の探索アルゴリズムを改善することで AR アプリケーションへ適用出来る可能性があると考えられる。

5.4 AR システムへの応用

本節では、位置・姿勢に依存した情報の提示を必要とする AR アプリケーションへの応用例について述べる。本システムの概要を Figure 26-(A) に示す。歩行者推測航法 (PDR)²⁵⁾ と、携帯端末内蔵のカメラで撮影される実画像と生成画像との比較に基づく絶対位置姿勢推定の二つの計測手法により構成され、画像を撮影する携帯端末となるクライアントは iPad2, 画像からカメラの位置・姿勢を推定するサーバは以下のスペックの計算機 (Windows7, Intel(R) Core(TM)i5 CPU M520 2.4GHz) を用いた。PDR システムは、加速度センサ, ジャイロセンサ, 地磁気センサ, 気圧センサが内蔵されたセンサモジュール (Samsung RZ Galaxy) で構成される。

本システムを用いて、ディスプレイに関する

Table 1: Computation time (ms)

SURF	ORB	FREAK	Edge	MI
300.2	339	305.3	79.8	27.7

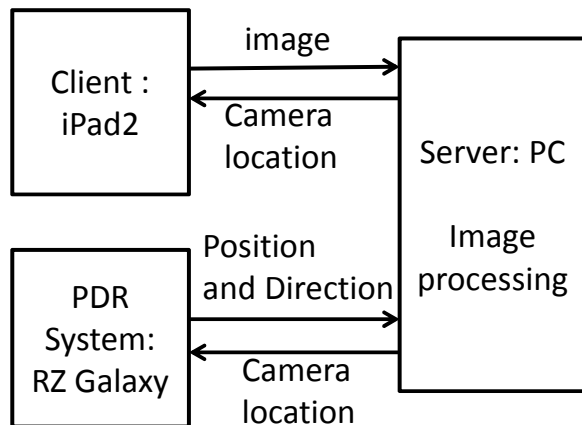


Figure 25: Structures of the AR system

国際ワークショップ: IDW/AD'12 (The 19th International Display Workshops) にてデモンストレーションを行った (Figure 26-(C)). デモブースの概観を Figure 26 に示す. カメラの位置・姿勢推定には, SURF 特徴を用いた特徴点によるカメラ位置・姿勢推定手法を利用した. データベースの作成のため, 周辺環境のモデリングを行い, 次に, 設置したポスターを iPad2 で撮影して, キーフレームの作成を行った. Figure 26-(B) にポスターの画像を示す. 本システムは, iPad2 により撮影された画像とキーフレームを用いて, カメラの位置姿勢を推定し, 常時相対測位を行っている PDR システムの位置・姿勢の補正を行った. 画像処理により推定された位置・姿勢により, ポスター上にラベルを重畳表示した画面の例を Figure 26-(D) に示す. 本デモでは, 仮想物体としてポスターに記載された項目に対応したラベルを表示した. さらに, このラベルを iPad2 のスクリーン上でタッチすると, 内容に合致した動画や pdf ファイルなどの, 詳細なコンテンツが閲覧できる. Figure 26-(E) に, 詳細なコンテンツとして動画が再生された際の画像を示す.

6 結び

本研究では, 実画像と生成画像との比較による携帯カメラの絶対位置・姿勢推定手法を想定し, HS ヒストグラム, SSD, ZNCC, 及び SIFT の各特徴量がどのように振る舞うかを予備実験により評価した. また, 入力画像として 3 つのシーンを用意し, それぞれのシーンで入力画像

とキーフレームを用いて, 特徴点, エッジ, 面の情報を用いてカメラの位置・姿勢を推定し, 位置, 姿勢, 仮想物体の投影誤差を評価した.

予備実験の結果から, ZNCC を用いた手法で大局的な絞り込みができること, 及び生成画像と実画像を対応付けるには SIFT 特徴を用いた方法の性能は不十分なことがわかった. また, 本実験の結果から, 相対測位との連携を考えると, SURF が最も推定に適していた. 一方で, 相互情報量を用いた手法は精度を保つために複数のパラメータで計算をする際の計算時間に課題は残るが, 高い精度でカメラパラメータを推定可能であることがわかった.

本実験の結果から, 特徴点を用いた手法において, 本実験で扱った特徴点を用いた手法の中では, SURF が最も推定に有効であることを示し, 特徴点を用いた手法に, 相互情報量を用いた手法を組み合わせることで, カメラの位置・姿勢の推定可能範囲を拡大する可能性について示した. また, 相互情報量を用いた手法の処理を並列処理することで, 実時間でのカメラ位置・姿勢推定への適用が期待できることがわかった.

他にも, 取り組むべき課題は多く, 本研究で取り扱った特徴量, 及び各評価結果を活用しつつ, 画像やデータベースの検索で用いられる高速化の手法を組み込むことや, 屋内ナビ等や他の AR アプリのようなユーザとシステムのインタラクションが想定される状況での位置・姿勢推定を想定すると, 常に自動で推定値を出すだけではなく, ユーザと協力して位置や姿勢の絞り込みを行っていくといった発展も考えられる. 今後, そのようなインタラクティブな測位やインタラクション技法についての検討が期待される.

謝辞

本研究を遂行するにあたり, 指導教官として終始丁寧な御指導, 御助言を頂いた, 筑波大学大学院システム情報工学研究科知能機能システム専攻 葛岡英明教授に心より感謝の意を表します.

本研究の全過程において, 丁寧で的確な御指導, 御鞭撻を賜った産業技術総合研究所サービス工学センター 蔵田武志行動観測・提示技術チーム 研究チーム長に心より感謝申し上げます.

学位論文審査において, 貴重な御指導と御助

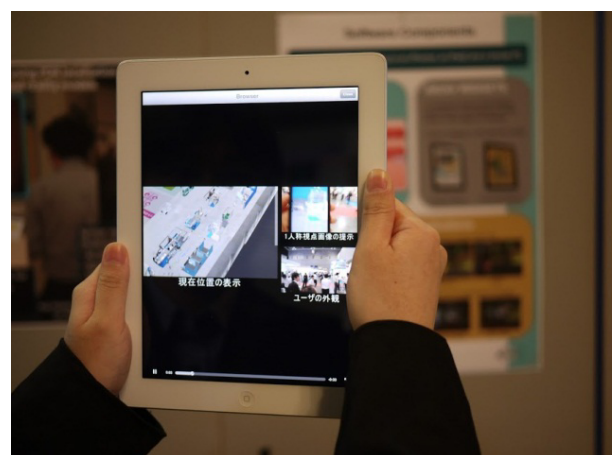
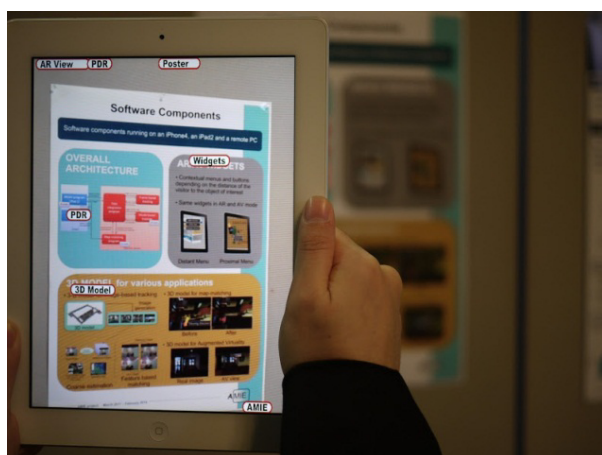
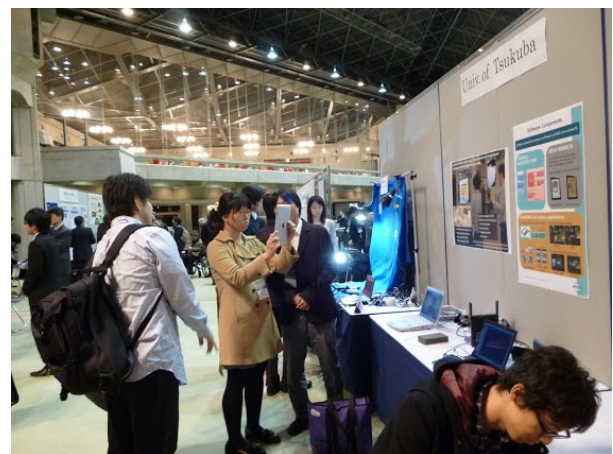
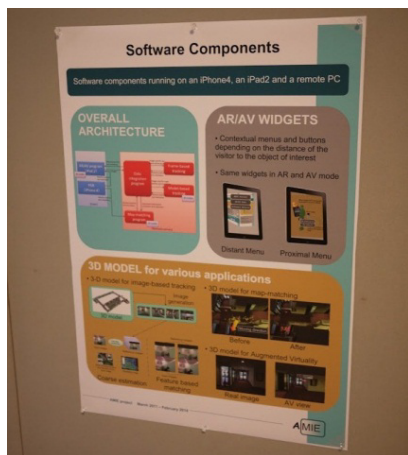
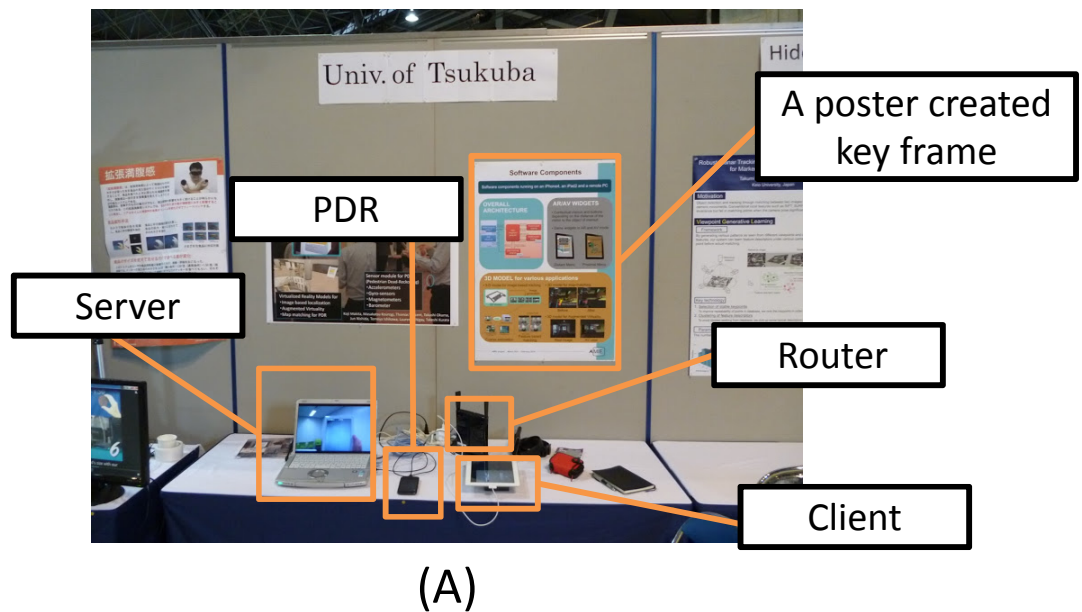


Figure 26: Images of the demonstration in IDW/AD'12. (A) System configuration. (B) An example of the poster for making key frames. (C) An overview of the demonstration. (D) An example of additional contents.

言を頂いた筑波大学大学院システム情報工学研究科知能機能システム専攻 亀田能成准教授に謹んで謝意を表します。

本研究において、日頃から研究の進め方、考え方、まとめ方など研究の基礎から懇切丁寧に指導して頂きました産業技術総合研究所サービス工学センター 興梠正克研究員、大隈隆史研究員、牧田孝嗣特別研究員、石川智也特別研究員に心より感謝致します。

本研究のための実験を行うにあたり、AR アプリケーションの開発、および実験環境の提供に協力頂いた、ジョセフ・フーリエ大学の Laurence Nigay 教授、Thomas Vincent 氏、Jeremie Francone 氏に心より感謝致します。

最後になりますが、3 年を共にしたグループウェア研究室のメンバーの皆様、大学および大学院生活を支えてもらった家族に深く感謝の意を表す。

本研究の一部は、JST 戦略的国際科学技術協力推進事業 (研究交流型)「日本-フランス (ANR) 研究交流」の支援を受けて実施されました。

参考文献

- 1) N. Kawai, M. Yamasaki, T. Sato, and N. Yokoya: “AR marker hiding based on image inpainting and reflection of illumination changes,” Proc. IEEE International Symposium on Mixed and Augmented Reality (ISMAR2012), pp. 293-294, Nov. 2012.
- 2) F. Okura: “Spacetime freeview generation using image-based rendering, relighting, and augmented telepresence,” Proc. ACM Multimedia (MM2012) Doctoral Symposium, pp. 1437-1440, Oct. 2012.
- 3) M. Swoboda, T. Nguyen, U. Eck, G. Reitmayr, S. Hauswiesner, R. Ranftl, C. Sandor: “BurnAR: Feel the Heat,” Proc. IEEE International Symposium on Mixed and Augmented Reality (ISMAR2011), Basel, Switzerland, October, 2011.
- 4) M. Kourogi, T. Kurata and K. Sakaue: “A Panorama-based Method of Personal Positioning and Orientation and Its Real-time Applications for Wearable Computers,” Proc. Fifth International Symposium on Wearable Computers (ISWC’01), pp. 107-114, 2001.
- 5) K. Makita, J. Nishida, T. Ishikawa, T. Okuma, J. Yamasita, H. Kuzuoka and T. Kurata: “Feasibility Study on the Estimation of Photo Shoot Position and Direction Based on Virtualized Reality Environment Models,” Proc. ICAT 2011, pp. 137, 2011.
- 6) 西田 純, 石川 智也, 牧田 孝嗣, 興梠 正克, 山下淳, 葛岡 英明, 蔵田 武志: “仮想化現実環境モデルに基づく携帯カメラの位置方位推定に向けて,” Proc. Vision Engineering Workshop 2011 (ViEW2011) 講演論文集, 2011, pp.316-323.
- 7) H. Bay, T. Tuytelaars, and L. V. Gool: “SURF: Speeded Up Robust Features,” Proc. 9th European Conference on Computer Vision (ECCV ’2006), Springer LNCS volume 3951, part 1, pp. 404-417, 2006.
- 8) E. Rublee, V. Rabaud, K. Konolige, and G. Bradski: “ORB: an Efficient Alternative to SIFT or SURF,” In Proc. of ICCV’11.
- 9) A. Alahi, R. Ortiz and P. Vandergheynst: “FREAK: Fast Retina Keypoint,” Proc. Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference, pp. 510-517, 2012.
- 10) J. Canny: “A computational approach to edge detection,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-8, no. 6, pp. 679-698, Nov 1986.
- 11) L. Vacchetti, V. Lepetit, P. Fua, “Combining Edge and Texture Information for

- Real-Time Accurate 3D Camera Tracking,” Proc. IEEE International Symposium on Mixed and Augmented Reality (ISMAR2004), pp.48-57, 2004.
- 12) A. Dame, E. Marchand. “Accurate real-time tracking using mutual information,” Proc. IEEE International Symposium on Mixed and Augmented Reality (ISMAR2010), pp. 47-56, 2010.
- 13) G. Reitmayr and T. Drummond: “Initialization for Visual Tracking in Urban Environments,” Proc. IEEE International Symposium on Mixed and Augmented Reality (ISMAR2007), pp. 161-172, 2007.
- 14) H. Yoshida, S. Ito and N. Kawaguchi: “Evaluation of Pre-Acquisition Methods for Position Estimation System using Wireless LAN,” The Third International Conference on Mobile Computing and Ubiquitous Networking (ICMU 2006), pp. 148-155, 2006.
- 15) R. Tenmoku, M. Kanbara, and N. Yokoya: “A wearable augmented reality system using positioning infrastructures and a pedometer,” Proc. IEEE International Symposium on Wearable Computers, pp. 110-117, 2003.
- 16) G. Klein, D. Murray: “Parallel Tracking and Mapping for Small AR Workspaces,” Proc. IEEE International Symposium on Mixed and Augmented Reality (ISMAR2007), 2007.8.
- 17) S. Saito, A. Hiyama, T. Tanikawa, M. Hirose: “Indoor Marker-based Localization Using Coded Seamless Pattern for Interior Decoration,” Proc. IEEE Virtual Reality 2007, pp. 67-74, 2007.
- 18) G. Bleser, H. Wuest, and D. Stricker: “Online camera pose estimation in partially known and dynamic scenes,” Proc. IEEE International Symposium on Mixed and Augmented Reality (ISMAR2006), pp. 56-65, 2006.
- 19) G. Simon and M. Berger: “Pose estimation from planar structures,” IEEE Computer Graphics and Applications, vol. 22, No. 6, pp. 46-53, 2002.
- 20) D. Hahnel, W. Burgard and S. Thrun: “Learning Compact 3D Models of Indoor and Outdoor Environments with a Mobile Robot,” Robotics and Autonomous Systems, vol.44, no.1, pp.15-27, 2003.
- 21) B. Jensen, J. Weingarten, S. Kolski and R. Siegwart: “Laser Range Imaging using Mobile Robots: From Pose Estimation to 3d-Models,” Proc. of 1st Range Imaging Research Day, pp.129-144, 2005.
- 22) M. Langer and S. BenHimane: “An Interactive Vision-based 3D Reconstruction Workflow for Industrial AR Applications,” Proc. IEEE International Symposium on Mixed and Augmented Reality (ISMAR2010), Workshop on Augmented Reality Super Models, 2010.
- 23) T. Ishikawa, K. Thangamani, M. Kourogi, A. P. Gee, W. Mayol, K. Jung, and T. Kurata: “In-Situ 3D Indoor Modeler with a Camera and Self-Contained Sensors,” In Proc. HCI2009, LNCS 5622, pp. 454-464, 2009.
- 24) T. Ishikawa, M. Kourogi, and T. Kurata, “Economic and Synergistic Pedestrian Tracking System with Service Cooperation for Indoor Environments,” International Journal of Organizational and Collective Intelligence, Vol.2, No.1, pp.1-20, 2011.
- 25) M. Kourogi, and T. Kurata, “Personal Positioning Based on Walking Locomotion Analysis with Self-Contained Sensors and

- a Wearable Camera,” Proc. IEEE International Symposium on Mixed and Augmented Reality (ISMAR2003), pp. 103-112, 2003.
- 26) B. Jeff and D. G. Lowe: “Shape indexing using approximate nearest-neighbor search in high-dimensional spaces,” Conference on Computer Vision and Pattern Recognition, pp. 1000-1006, 1997.
- 27) M. A. Fischler, and R. C. Bolles: “A paradigm for model fitting with applications to image analysis and automated cartography,” Comm. of the ACM, Vol. 24, pp. 381-395, 1981.
- 28) M. a. Abidi and T. Chandra. “A new efficient and direct solution for pose estimation using quadrangular targets: Algorithm and evaluation,” Proc. IEEE PAMI, 17(5), 534-538, 1995.
- 29) A. Petit, G. Caron, H. Uchiyama, E. Marchand: “Evaluation of Model based Tracking with TrakMark Dataset,” In Proc. of TrakMark2011, 2011.
- 30) K. Makita, T. Okuma and T. Ishikawa: “A benchmark suite with virtualized reality models for supporting evaluation and data set generation,” In Proc. of TrakMark2012, 2012.

西田 純



筑波大学大学院システム情報工学研究科
知能機能システム専攻所属