

# 文字認識と SFM を用いた携帯端末ユーザの位置推定

山森 康生<sup>†</sup> 廣瀬 詢<sup>†</sup> 藤里 和樹<sup>‡</sup> 頼光 拓真<sup>‡</sup> 檀 寛成<sup>†</sup> 安室 喜弘<sup>†</sup>

関西大学 環境都市工学部<sup>†</sup> 関西大学大学院 理工学研究科<sup>‡</sup>

## 1 はじめに

近年、スマートフォンの普及により地図情報を利用したサービスが非常に身近なものとなっている。コストのかからないフリーソフトやインストールを必要としないウェブアプリケーションでさえ、地図情報を提示する機能だけではなく、カーナビゲーション、店舗案内など動的で幅広い位置情報サービスを提供している。災害時においても、現在位置を基準に移動経路を検索できるということは、避難指示の観点からも非常に有効である。しかし、屋内や地下など、GPS 信号が届かない場所にユーザが入ってしまうと、これらのサービスを受けることが技術的に難しく、GPS 電波の有無に関わらず、シームレスに自分の位置情報を得る方法が望まれている。

## 2 先行研究

地下街における位置推定手法には、ビーコンを設置する方法や Wi-Fi 電波強度や地磁気等の電磁波を利用する方法のほか、カメラ付き携帯端末により撮影した画像の自然特徴点を利用した推定手法がある[1,2]。ビーコンなどのランドマークを設置する場合には、コストや設備面で難点がある。電磁波による方法は、時間帯や場所によって不安定な場合がある。画像情報による自己位置推定は、自然特徴点が多いシーンでは比較的頑健な方法であるが、特徴点が多いほど、検出処理や事前情報との照合処理に計算コストがかかる。本研究では、画像情報に基づいた手法を基本とし、都市部で発達している地下街や交通機関などの環境において、文字情報が豊富であることに注目し、経路案内に堪える程度の速さと正確さを併せ持つ位置推定手法を提案する。

## 3 提案手法

### 3.1 基本方針

本研究では、まず、サービス対象とする地下空間において視点と方向を変えながら写真を撮影し、画像を収集する。収集した画像に対して SFM (Structure From Motion) を適用して対象空間の 3 次元モデルを作成する。次に、収集した各画像において、文字認識処理を行い、識別された文字の画像上の位置と、SFM で再構成された 3 次元空間での座標との対応づけを行い、この対応関係を参照できるようにデータベース化しておく。新たにユーザ端末で撮影された画像に対しても、文字認識処理により文字情報を抽出し、事前に登録された文字情報を参照して、対応する 3 次元座標を取得する。最終的に、1 つの画像内で複数の文字について、画像上の 2 次元位置と、3 次元モデルの 3 次元座標との幾何学的な関係が得られるので、整合する撮影位置としてユーザの位置を求めることができる。

### 3.2 位置推定方法

一般に、画像上の 2 次元座標と空間中の 3 次元座標との対応関係からカメラの外部パラメータを求める問題は PnP (Perspective-n-Point) 問題として定式化され、その解法が知られる。ここでは、厳密なカメラの位置や姿勢ではなく、ユーザの立ち位置を求めるという目的から、次のように問題を単純化し、速やかに算出する。

まず、ユーザ端末で撮影された画像から文字認識処理によって識別された複数の文字について、3 次元モデルにおける座標と対応づけしておく。図 1 に示すように、画像上で識別された複数の文字の 2 次元座標とカメラの画角との関係から、カメラの焦点  $O_c$  から各文字の方向を指す 3 次元ベクトルを求める。これら複数の文字の方向ベクトルについて互いの内積を計算しておく。一方、事前に作成したデータベースから、識別された文字の 3 次元空間座標を参照し、ユーザ位置の候補となる座標  $O_w$  から各文字を指す方向ベクトルを求めることができ、同様に互いの内積を計算できる。画像から計算される内積の値との二乗誤差を評価値として、これを最小とするユーザの立ち位置を解とする。

Position estimation of mobile terminal users using character recognition and SFM

Kouki Yamamori<sup>†</sup>, Makoto Hirose<sup>†</sup>, Kazuki Fujisato<sup>‡</sup>, Takuma Yorimitsu<sup>‡</sup>, Hiroshige Dan<sup>†</sup>, Yoshihiro Yasumuro<sup>†</sup>

<sup>†</sup>Faculty of Environmental and Urban Engineering, Kansai University

<sup>‡</sup>Graduate School of Science and Engineering, Kansai University

3-3-35 Yamate, Suita, OSAKA, 564-8680, JAPAN

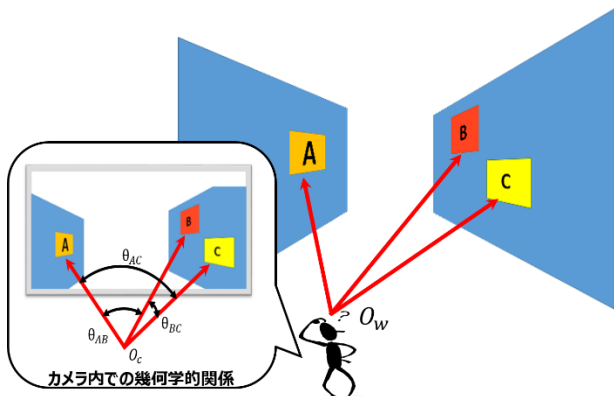


図1：提案手法概念図

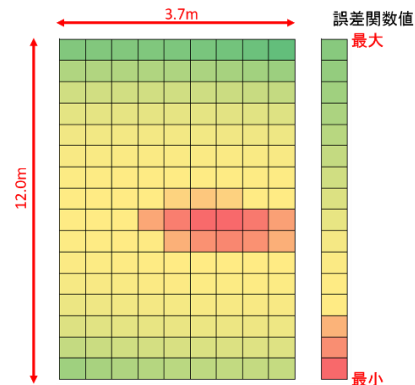


図3：位置推定結果

#### 4 実験

本研究では、人通りが多く、多種多様な店を擁する場所として、梅田地下街（大阪駅前第3ビルB2F）の一角を対象として位置推定の実験を行った。ウェアラブルカメラ（Panasonic製HX-A500: 3840x2160画素, 水平画角136°）により直線状の通路を12m程度の距離を通行しながら撮影した画像34枚をVisual SFM[3]により3次元点群を取得し、さらにPMVS（Patch-based Multi-View Stereo）によりメッシュ化して3次元モデルを生成した。図2に生成したモデルを示す。文字認識処理としてCloudVision APIを用いて取得画像から文字情報を抽出し、文字列の中心に相当する画像座標を記録した。文字（列）座標とメッシュモデルでの3次元座標との対応付けを行った上で、カメラの焦点距離と画角、画素数の関係から、カメラのローカル座標系（3次元）における文字の方向ベクトルを求めた。今回は図2に示すように、文字列の中から「和牛」「GAME」「エレベータ」「地下鉄」の4つを選択して用い、通路上の3.7m×12.0mの領域を対象としてユーザの位置を推定するものとした。SFM座標における対象領域を横9縦16の144グリッドに切り分け、グリッドの各点で内積値の二乗誤差値を算出する。なお、歩行しながらの撮影であること

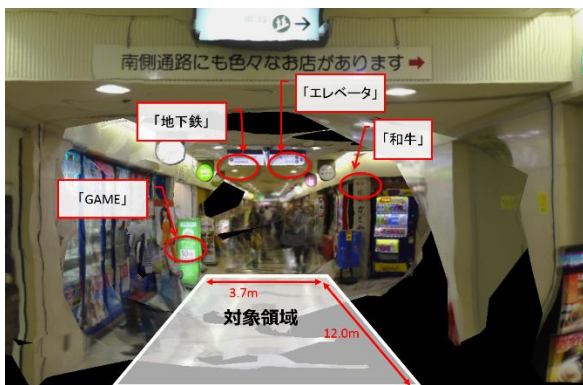


図2：対象とした空間

を考慮して、床面からの高さは1.7mに相当する高さで固定した。算出された誤差の分布を図3に示す。1枚の写真入力から、進行方向については0.75~1.50m（1~2グリッド）程度、横方向については、1.2~1.8m（4~6グリッド）程度の位置推定分解能が得られた。誤差分散に方向性が見られるのは、写真を撮影した視線方向と、検出された文字列の方向に依るものと考えられるが、通路での位置として、どのブロック、どの店舗の付近にいるかなど、行動の記録や、経路の誘導などに有効な位置情報であると考えられる。今回の実装では、事前の文字情報と3次元座標との対応付けや、ユーザ端末からの入力画像から抽出される文字情報の選択などを手動で行っているため、これらの自動化が今後の課題である。

#### 5 おわりに

本研究では、SFMと文字情報を用いたユーザ携帯端末位置推定手法を提案し実験により有効性を確認した。今後は、実験室などの環境における真値との比較による精度検証や、観測される文字情報の分布や指向性について分析を進める予定である。

謝辞：本研究は科学研究費補助金、関西大学先端科学技術推進機構研究グループの助成を受けた。

#### 参考文献：

- [1] 山崎俊太郎, 持丸 正明, 金出 武雄, “一人称ビジョンシステムのための自己位置推定法”, 電子情報通信学会技術研究報告, pp.73-78, 2010.
- [2] 佐藤 智和, 池田 聖, “景観画像データベースを用いたユビキタス端末の自己位置推定に関する研究”, 戦略的情報通信研究開発推進制度(SCOPE) 第6回成果発表会, pp.20-21, 2010.
- [3] C. Wu, Toward Linear-time Incremental Structure from Motion, 3DV 2013, pp.127-134, 2013.