

カメラ画像に写る物体の位置を利用した AR 注釈対象特定法

三 田 裕 策^{†1} 志 築 文 太 郎^{†2} 田 中 二 郎^{†2}

モバイル型の Augmented Reality (AR) の一表現手法に、カメラから取得した現実環境の画像（以下カメラ画像）に注釈を重畳表示する手法がある。これまでの AR システムでは注釈対象特定法に、カメラ画像のみを用いる方法や、GPS や電子コンパス等の機器の情報をを用いる方法がある。この手法は、注釈対象と注釈を結びつけて表示できる利点を持つ。しかし、この手法の問題として、本来の注釈対象以外の物体に注釈が表示される場合がある。この問題が発生した場合、ユーザは注釈がその別の物体に対する情報であると誤った認識を行う。

本研究ではこの問題を解決するために、カメラ画像に写る注釈対象を特定し、注釈対象に注釈を重畳表示する手法を示した。更に、本手法による注釈表示方法を行うプロトタイプの実装を行った。本手法が我々の意図した注釈表示を行うかを調べる実験を行い、その結果について考察を行った。また、考察を基に本システムの改良を行った。

A Method to Identify AR Annotated Objects Utilizing Locations of Real World Objects in Camera Images

YUUSAKU MITA,^{†1} BUNTAROU SHIZUKI^{†2}
and JIRO TANAKA^{†2}

Some Augmented Reality (AR) systems for mobile devices overlay annotations on images taken by camera (hereafter, we call this camera images). To identify an object which is to be annotated, traditional AR systems use camera images information or devices such as GPS and electronic compass. AR systems can connect annotated objects with annotations by using these inputs. However, these systems have a problem that annotations are overlaid on wrong objects. If the problem occurs, users would recognize annotations as an information of wrong objects.

In this research, we show a novel technique to identify an object from camera images which is to be annotated. We implemented a prototype of an AR system which uses our technique to overlay annotations. We investigated whether our system overlay annotations as we intend, and had discussion about results. We also improved our system from observation we got from the discussion.

1. はじめに

モバイル型の Augmented Reality (AR) の一表現手法に、カメラから取得した現実環境の画像（以下カメラ画像）に注釈を重畳表示する手法（以下 AR 注釈表示）がある。この AR 注釈表示の実現には、画面内に写っている物体が注釈を表示すべき物体（以下注釈対象）か否かを特定する必要がある。これまでの AR システムではその方法として、カメラ画像に写る物体の頂点を目印とする方法¹⁾、GPS や電子コンパス等の機器から得たユーザの位置及び方角とデータベースから取得した注釈対象の位置情報を対応付ける方法（例：セカイカメラ^{*1}や Laya^{*2}等のアプリケーションや、廣瀬らの手法²⁾）等が用いられてきた。特に廣瀬らの手法¹⁾はカメラ画像に写る物体を特定するので、注釈を正確にその物体に関連づけて表示可能な方法として有望である。

しかし、我々はこれまでの注釈対象特定法には注釈が注釈対象に重畳表示されない問題があると考えた。カメラ画像のみを用いる方法では、カメラ画像の注釈対象の認識が条件であり、注釈対象の一部が隠れてしまうと注釈を表示できない場合がある。GPS や電子コンパスの情報をを用いる方法では、その情報を精度良く取得できない場合、図 1 のように注釈対象とずれた位置に注釈が表示される。また、注釈対象の手前に物体が存在し、注釈対象を遮っている場合、注釈を表示した位置に別の物体が写っているため、図 2 のように注釈が別の物体に対して重畳表示される。これらの問題が発生した場合、注釈対象以外の物体に対して注釈が表示されるため、ユーザは表示された注釈がその別の物体に対する情報であると誤った認識を行う。

本研究では、カメラ画像に写る物体の位置を利用した AR 注釈対象特定法を示す。本手法は、GPS 及び電子コンパスから取得したカメラの位置と方角に加えて、カメラ画像に写る物体までの距離及び角度を算出し、これらをデータベースに格納されている注釈データと比較して注釈対象の特定を行う。なお、本手法は、GPS 及び電子コンパス、単一カメラ、注釈対象の位置情報のみを利用するため、既存の AR 注釈表示が可能であるスマートフォン等の

^{†1} 筑波大学大学院システム情報工学研究科コンピュータサイエンス専攻

Department of Computer Science, Graduate School of System and Information Engineering, University of Tsukuba

^{†2} 筑波大学システム情報系

Faculty of Engineering, Information and System, University of Tsukuba

^{*1} Sekai Camera. <http://sekaicamera.com/>

^{*2} Laya. <http://layar.jp/>



図 1 表示位置がずれた注釈表示の例
注釈が正しい注釈対象（赤色）とは異なる建物に対して表示されている
Examples of annotation which position is shifted.



図 2 注釈対象が隠れている注釈表示の例
注釈が正しい注釈対象（緑色）を遮っている建物に対して表示されている
Examples of annotation to hidden target.

端末全てにおいて使用できる利点がある。

以下、本論文の構成を示す。第 2 節では関連研究を述べる。第 3 節では本手法に用いる注釈対象特定法を述べ、第 4 節では本手法による注釈表示を行うプロトタイプについて述べる。第 5 節では実装した本システムを用いて実際に注釈表示を行った実験とその考察を述べる。第 6 節ではまとめを述べる。

2. 関連研究

Honkamaa らは、屋外においてカメラ付きの携帯情報端末を特定の位置に向けてかざすと、その位置に対応した注釈が画像に表示されるシステム³⁾を開発した。この研究では、GPS とコンパス、Google Earth から取得した建物の 3 次元形状の情報をを用いて、カメラ画像に注釈を重畳表示する。本研究は、カメラ画像に写る注釈対象に対して注釈の表示を行う点が同じであるが、Google Earth のような地理情報システムの情報が無い場面においても、注釈対象の特定を行うことができる。

石川らは屋外 AR システムによる現実環境と仮想環境の幾何学的ずれの解消を目的とした高精度定位法⁴⁾を提案した。この研究では、カメラ画像から建物と空の境界線を取得し、建物の 3 次元形状から得た境界線とマッチングを行うことにより現実環境と仮想環境のずれを解消する。この手法の使用条件として、建物及び空がカメラ画像に写ること、また、建物の 3 次元形状を予め入力する必要がある。山本らは複数の特性の異なるカメラを用いる

ことにより複合現実型視覚情報提示システム⁵⁾を開発した。この研究では複数の定点カメラから取得した画像を利用して、現実空間と仮想空間を精度良くマッチングさせる。本研究では現実空間と仮想空間の対応方法として、GPS 及び電子コンパス、単一カメラ、注釈対象の位置情報のみを利用して注釈対象の特定を行うため、既存の AR 注釈表示が可能であるスマートフォン等の端末において、追加の機器や情報源を用いることなく使用できる。

Klein らは 3 次元空間の特徴点マッピングを行い、その空間の追跡を行う手法である Parallel Tracking and Mapping (PTAM)⁶⁾を開発した。PTAM は、初めて写す場所でも実行可能であり、単一カメラにより 3 次元空間のマッピング及び追跡を行うことができる特徴を持つ。これらの特徴は、使用場所を限定せずに単一カメラにより利用できる利点を持つ。本研究では、PTAM のマッピング及び追跡方法を利用してシステムの作成を行う。

Lee らは手のひらと指先を注釈対象として特定する Handy AR⁷⁾を提案した。この研究では 5 本の指の先の位置を利用して手のひらの 1 点を原点とし、手のひらの向きを軸の方向とする座標系を作成する。小谷らは、注釈対象の形状を考慮して注釈の表示位置を決定する手法⁸⁾を提案した。この手法では、ユーザが注釈対象の位置及び領域を入力し、これを利用して注釈対象の形状を取得する。廣瀬らは、画像内の頂点追跡に基づいたマーカレス位置合わせを行う研究¹⁾を行った。この手法では、物体の 3 次元モデルを用いた物体の頂点追跡により正確な位置合わせを行なう。使用条件として、カメラ画像中に 3 次元形状が既知の物体を配置する必要がある。本研究ではこれらの研究と同じく特定のマーカを使用せずに情報の重畳表示を行うが、注釈対象を特定する際に注釈対象の形状に条件を持たず、事前に注釈対象の形状を入力する必要がある。

3. カメラ画像に写る物体の位置を利用した AR 注釈対象特定法

本手法は、カメラ画像に写る物体までの距離及び角度を利用し、その物体が注釈対象であるかを特定する。この距離を求めるために、複数画像を用いた視差を利用する。距離を求める他の方法として、3 次元カメラを用いる方法⁹⁾や、大きさが既知である物体をカメラ画像に写す方法¹⁰⁾があるが、複数画像を用いた視差を利用する方法は、既存の携帯情報端末において他の機器や情報源を追加無しに動作するメリットを有する。

3.1 特徴点の利用

本手法ではカメラ画像から取得した特徴点を用いる。この理由は、次の通りである。

- 特徴点は何もない部分や無地の壁等の周囲の画素との輝度差が小さい部分からは検出されないが、画像中の注釈対象となる部分はオブジェクトであることが多く、その部分が

らは特徴点が取得できる。

- 全画素ではなく、特徴点のみに対して距離と角度の算出を行うことにより、注釈対象の特定を高速に行う。

本システムではカメラ画像から特徴点を取得し、その距離を取得する方法として、Parallel Tracking and Mapping (PTAM)⁶⁾ を利用する。PTAM では、カメラを水平に動かして得た動画から複数画像を抜き出し、その複数画像の視差を利用して特徴点の3次元位置を取得する。この特徴点の3次元位置は画像に写る物体の位置として扱われる。以下、本論文ではこのPTAMにより作成された3次元空間を特徴点空間と呼ぶ。本手法では特徴点空間における情報の内、特徴点の位置、カメラの位置、カメラの向きを取得する。

3.2 注釈対象特定手順

本手法では、注釈対象の特定を3つのフェイズに分けて行う。第1フェイズでは、カメラと注釈対象の絶対位置を用いて、カメラから注釈対象までの距離及び方角を算出する。第2フェイズでは、特徴点空間内の特徴点の3次元座標を利用して、カメラから各特徴点までの距離及び方角を算出する。第3フェイズでは、算出した距離及び方角それぞれの比較により、注釈対象から取得した特徴点を特定する。

第1フェイズ

カメラの位置の緯度 N_t 及び経度 E_t と、注釈対象の緯度 N_n 及び経度 E_n を利用して、カメラから注釈対象までの距離 d_1 及び方角 θ_1 を算出する。これらの関係を図3に示す。

第2フェイズ

特徴点空間から取得した特徴点の3次元位置座標 (X_m, Y_m, Z_m) 及びカメラの3次元位置座標 (X_c, Y_c, Z_c) と、カメラの向きを利用して、カメラから特徴点までの距離 d_2 及び方角 θ_2 を算出する。これらの関係を図4に示す。

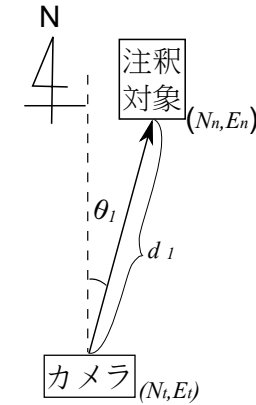


図3 カメラから注釈対象までの距離 d_1 と方角 θ_1
 d_1 :distance from camera to annotated, θ_1 :direction of annotated from camera.

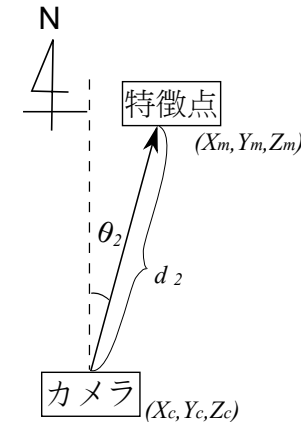


図4 カメラから特徴点までの距離 d_2 と角度 θ_2
 d_1 :distance from camera to feature point, θ_1 :direction of feature point from camera.

第3 フェイズ

第1 フェイズ及び第2 フェイズの算出結果から注釈対象を特定する。まず方角を比較し、その角度差が閾値 T_θ 以下だった特徴点の内、距離の差が最も小さい特徴点を取得する。その距離の差が閾値 T_d 以下の場合、その特徴点を注釈対象を表す特徴点として扱う。

4. プロトタイプ

本手法を用いた注釈表示を行うプロトタイプを作成した。本プロトタイプでは、電子コンパスと GPS モジュールから情報を取得するために Arduino を用いた。また、カメラとしてオートフォーカス機能を搭載した Web カメラを用いた。オートフォーカスのカメラを用いることで、ピント合わせを自動で行うことができ、ピントの合った鮮明な画像からより多くの特徴点を取得できる。表1に本プロトタイプに用いたデバイスを記し、図5にデバイスの全体像を示す。これらのデバイスを1つに纏めることにより、カメラの位置と向きに応じた値を GPS モジュールと電子コンパスからそれぞれ取得できる。また実装には Visual Studio 2008 環境の C++言語を用いた。本プロトタイプはデスクトップ計算機上で動作する。

表1 使用デバイス
Used devices.

カメラ	Logicool HD Pro Webcam C910
Arduino	Arduino Pro Mini 328 3.3V 8MHz
GPS モジュール	RS232C レベルコンバータ内蔵 GPS モジュール GT-723F
電子コンパス	HMC6352 搭載デジタルコンパスモジュール

本プロトタイプでは、 T_θ 及び T_d を、それぞれ 8 度、15m とした。これらの閾値はシステムを使用した経験により設定した値である。なお、カメラから注釈対象までの距離 d_1 は現実空間上の距離（単位：m）であるのに対して、カメラから特徴点までの距離 d_2 は特徴点空間の大きさを基準とした距離であるため、2つの距離の縮尺を変換する必要がある。本プロトタイプではこの変換のために用いる定数をユーザに入力させる。ユーザは注釈表示を行う際に、表示された注釈が妥当になるように確認しながらその値を調節する。

本プロトタイプの注釈を表示するまでの流れを示したシステムフローチャートを図6に示す。

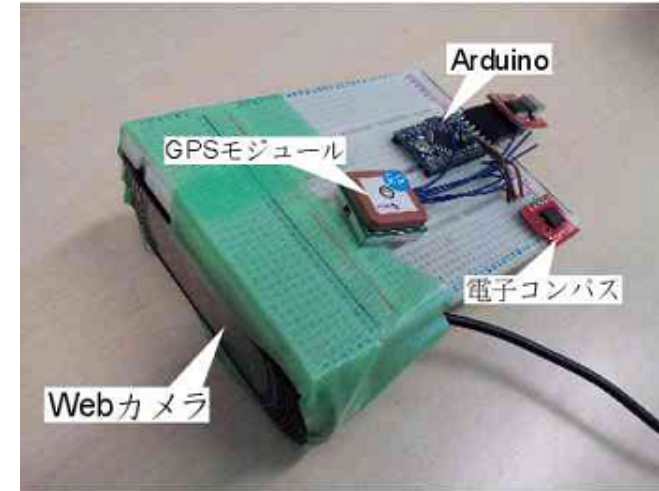


図5 プロトタイプの全体像
The prototype.

5. 実験

実装したプロトタイプが我々の意図した注釈表示を行うかを調べるために実験を行った。

5.1 実験方法

カメラの位置と注釈対象の位置を同一に設定した条件内において、本手法と GPS と電子コンパスのみを用いた従来手法を用いた注釈表示を行う。その後、それぞれの方法による注釈の表示位置を比較する。なお、従来手法には、本システムから画像処理の部分を除いたものを用いた。本実験で用いるデータベースには予め注釈対象の緯度及び経度が入力されている。実験当日の天気は晴れであり、午前10時から午後3時の間に行った。実験では、筑波大学構内の屋外である10箇所において以下の手順を踏んだ。

- (1) 指定した箇所において、本手法により注釈表示を行う
- (2) (1)と同様に、同じ箇所において従来の手法により注釈表示を行う
- (3) (1), (2)のそれぞれの方法による注釈表示位置の比較を行う

また、表示する注釈には下向きの三角形を用いた。三角形の下側の頂点が画像の注釈対象

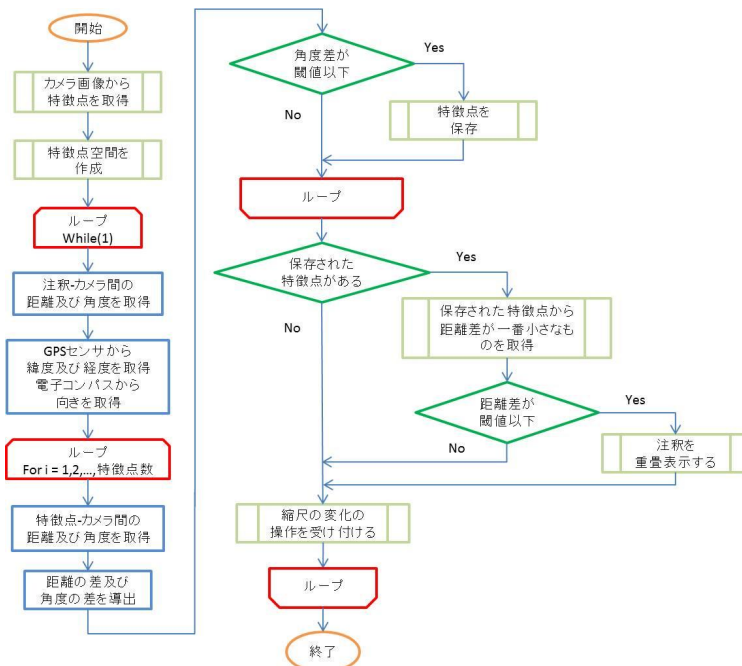


図 6 システムフローチャート
System flowchart.

を示す位置である。本手法においては赤色の三角形を、従来手法においては青色の三角形を表示した。注釈が表示されている画像中の色のついた点は、カメラ画像から取得した特徴点である。

5.2 結 果

実験結果を纏めたものを表 2 に示す。表には、注釈対象の特定の成否、注釈対象までの距離、実験箇所の備考を示す。

表 2 実験結果
Result of experiment.

実験場面	特定の成否	距離 (m)	備考
場面 1	成功	43	
場面 2	成功	85	逆光であった
場面 3	成功	43	
場面 4	成功	53	逆光であり、注釈対象の手前にネットが存在した
場面 5	成功	142	注釈対象の全部が隠れていた
場面 6	失敗	233	逆光であった
場面 7	失敗	44	注釈対象付近が明るかった
場面 8	失敗	33	注釈対象付近が明るかった
場面 9	失敗	56	注釈対象の手前にネットが存在した
場面 10	失敗	178	

10 箇所の内、5 箇所において目標とした注釈対象の特定を行うことができた。この内、従来手法と本手法の表示位置の差が明確な場面 1 と場面 2 について、注釈対象を示す画像、本手法による注釈表示の画像、従来手法による注釈表示の画像を提示する。

場面 1

場面 1 では、図 7 の 2 つの建物の内、黄色に塗りつぶされている左奥の建物が注釈対象である。従来手法では、図 8 に示すように誤って手前の物体に対して注釈が表示された。しかし、本手法では、図 9 に示すように注釈対象でない手前の物体を避けて正しく注釈が表示された。



図 7 場面 1 における注釈対象の位置
Position of the target at scene1.



図 10 場面 2 における注釈対象の位置
Position of the target at scene2.

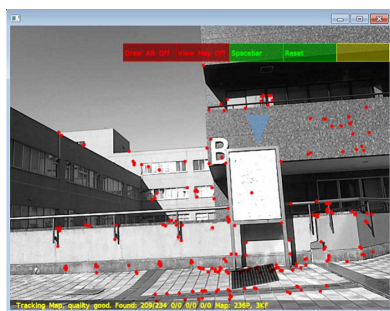


図 8 場面 1 における従来の手法による注釈表示
Annotation by the conventional method at scene1.



図 9 場面 1 における本手法による注釈表示
Annotation by our method at scene1.

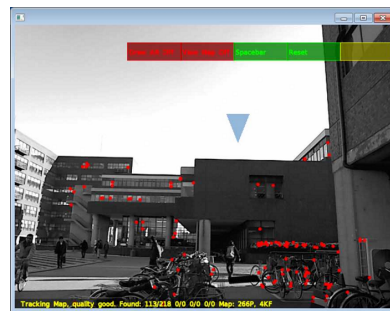


図 11 場面 2 における従来の手法による注釈表示
Annotation by the conventional method at scene2.

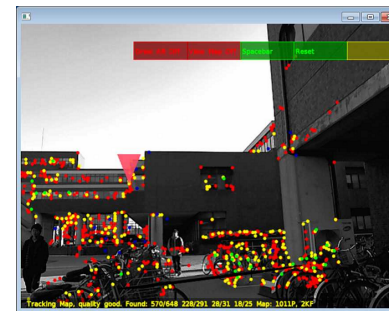


図 12 場面 2 における本手法による注釈表示
Annotation by our method at scene2.

場面 2

場面 2 では、図 10 の建物の内、中央付近の黄色に塗りつぶされている建物が注釈対象である。従来手法では、図 11 に示すように空中に注釈が表示された。しかし、本手法では、図 12 に示すように注釈対象に対して正しく注釈が表示された。

5.3 考察と改良案

本節では、半数の箇所において目標通りに動作しなかった原因を考察する。また、この考察から得た問題に対処するためのシステムの改良案を述べる。

5.3.1 画像の明暗

場面 7 や場面 8 の様に、カメラ画像の注釈対象付近が明るい場合、注釈対象から特徴点が取得されなかった。本手法では特徴点の取得にモノクロ画像を用いるが、実験環境が晴れ

の日の昼間の屋外という条件であったためにカメラ画像の輝度が高くなり、画像に一樣に白色になる部分が発生した。特徴点は周囲の部分との輝度差が大きな点であるので、一樣な白色の部分からは特徴点が検出されない。しかし、逆光下の場面 3 や場面 4 の様に、カメラ画像の輝度が高い場合においても注釈対象付近の輝度が低い場合は特徴点を取得できた。

このため、特徴点を取得する際に注釈対象が写る部分の輝度を下げていることを考えている。物体識別後に、カメラ画像の注釈対象の部分に基づいた輝度の調整を行う。

5.3.2 注釈対象までの距離

カメラから注釈対象までの距離が遠い箇所において、注釈対象の特定が出来なかった。注釈対象までの距離が遠い場合、特徴点の 3 次元位置を取得するためにカメラの位置を動かした時のカメラ画像間の視差が微少になり、特徴点の距離が上手く取得できないことが原因であると考えられる。

例として、5.3.1 節の画像の明暗の問題が発生しない箇所において、注釈対象の特定ができた最も遠い注釈対象までの距離は場面 3 の 85m であり、注釈対象の特定ができなかった最も近い注釈対象までの距離は場面 10 の 178m であった。

この問題を解決するため、注釈対象までの距離が遠い場面では、従来手法による注釈表示位置に対して、本手法による物体までの距離測定を行い、以下の条件に従った注釈表示方法を考えている。

注釈を表示しない 注釈対象より手前に物体を認識した場合

注釈を表示する 注釈対象より手前に物体が無い場合

5.3.3 閾値

本プロトタイプでは、注釈対象の特徴点を識別するときの閾値 T_θ 及び T_d をそれぞれ 8 度、15m とした。そのため、閾値の範囲内に注釈対象の特徴点が存在せず、注釈対象以外の特徴点が存在する場合、注釈対象以外の特徴点に注釈が表示された。

この問題を解決するため、両閾値をカメラから注釈対象までの距離に応じて変化させることを考えている。具体的には、カメラから注釈対象までの距離が短ければ T_θ を大きくかつ T_d を短く設定する。

現在、角度と距離の閾値を変化させる方法を実装している。カメラから注釈対象までの距離 (m) が d の時、角度の閾値 (rad) を $\tan^{-1}(\frac{20}{d})$ 、距離の閾値 (m) を $\frac{d}{5}$ としている。ただし、この閾値の係数の設定は経験則により行ったため、実験により調整を行っている。

6. ま と め

本研究では、カメラ画像を利用した、画像に写る物体までの距離及び方角の算出により、AR システムの注釈対象を特定する手法を示した。更に、本手法による注釈表示を行うプロトタイプを実装して実験を行い、半数の箇所において目標とした注釈対象特定を行うことができた。また、目標となる注釈対象特定が行われなかった箇所の考察を行い、システムの改良を行った。

今後は課題として挙げた発展案の実装を行った上で、携帯情報端末において本手法の実装を行い、使用場面を問わず利用できるシステムとなるよう改善を行う予定である。

参 考 文 献

- 1) 廣瀬亮, 斎藤英雄: 頂点追跡に基づく AR のためのマーカレス位置合わせ手法, 情報処理学会研究報告, CVIM, pp. 103–110 (2005).
- 2) 唐津豊, 中澤仁, 高汐一紀, 徳田英幸: ORGA:加速度センサを利用したマーカレス AR によるセンサ情報可視化システム, 電子情報通信学会技術研究報告. USN, ユビキタス・センサネットワーク, pp. 79–84 (2010).
- 3) Honkamaa, P., Siltanen, S., Jappinen, J., Woodward, C. and Korkalo, O.: Interactive outdoor mobile augmentation using markerless tracking and GPS, *Proc. the Virtual Reality International Conference VRIC Laval France*, pp. 285–288 (2007).
- 4) 石川高志, 全へい東: 画像処理を用いた屋外 AR システムのための高精度定位, 情報処理学会研究報告. CVIM, pp. 181–188 (2005).
- 5) 山本治由, 津田崇博, 塚田正人, 北原格, 亀田能成, 太田友一: 屋外における複合現実型情報提示の広域化: 特性の異なるカメラ群の利用, 電子情報通信学会技術研究報告. MVE, マルチメディア・仮想環境基礎, pp. 23–28 (2006).
- 6) Klein, G. and Murray, D.: Parallel Tracking and Mapping for Small AR Workspaces, in *Proc. Sixth IEEE and ACM International Symposium on Mixed and Augmented Reality*, pp. 1–10 (2007).
- 7) Lee, T. and Hollerer, T.: Handy AR: Markerless Inspection of Augmented Reality Objects Using Fingertip Tracking, *Proc. IEEE International Symposium on Wearable Computers*, pp. 83–90 (2007).
- 8) 小谷享広, 牧田孝嗣, 神原誠之, 横矢直和: 注釈対象の形状を考慮した AR オーサリングシステム, 電子情報通信学会技術研究報告. PRMU, パターン認識・メディア理解, pp. 1–6 (2007).
- 9) Vacchetti, L., Lepetit, V. and Fua, P.: Combining edge and texture information for real-time accurate 3D camera tracking, *In Proceedings International Symposium on Mixed and Augmented Reality*, pp. 48–56 (2004).

- 10) Kato, H. and Billinghurst, M.: Marker Tracking and HMD Calibration for a Video-based Augmented Reality Conferencing System, *Proc. the 2nd IEEE and ACM International Workshop on Augmented Reality*, pp. 85–94 (1999).