

定点カメラ映像を援用した屋外歩行者のための複合現実型情報提示

武政 泰輔[†] 亀田 能成[†] 大田 友一[†]

[†] 筑波大学 〒305-8573 茨城県つくば市天王台 1-1-1

E-mail: [†] (takemasa, kameda, ohta)@image.esys.tsukuba.ac.jp

あらまし 屋外歩行者が手にしているカメラ付 PDA のディスプレイ上での複合現実型視覚支援法について述べる。歩行者から見て死角に相当する領域について、死角部分の CG モデルとテクスチャがリアルタイムに PDA カメラ映像に正確に重畳され、かつそのテクスチャはオンラインで更新される。その結果、死角領域に存在する動物体を PDA ディスプレイ上でリアルタイムに確認することができる。このような情報提示を可能にするために、屋外に多数設置した定点カメラの映像を活用する。特に本稿では、CG モデルを正確に重畳するための、物理センサと画像センサを併用した PDA カメラ位置姿勢推定法を提案する。本手法の特徴は、画像センサに定点カメラを用いその映像を利用することでランドマーク検出を安定化させ、正確な PDA 位置姿勢推定を実現していることである。実際に定点カメラを屋外環境中に 4 台設置したシステムを実装し、実験を通して提案手法の有効性を確認した。

キーワード 屋外型複合現実感, PDA, 定点カメラ, ランドマーク検出, 姿勢推定

Outdoor Mixed-Reality PDA Utilizing Surveillance Cameras

Taisuke TAKEMASA[†] Yoshinari KAMEDA[†] and Yuichi OHTA[†]

[†] University of Tsukuba 1-1-1 Tennodai, Tsukuba, Ibaraki, 305-0005, Japan

E-mail: [†] (takemasa, kameda, ohta)@image.esys.tsukuba.ac.jp

Abstract This paper describes a mixed reality system on a PDA to support pedestrians. Our goal is to eliminate blind spots by presenting users camera images overlapped with 3D models. The textures of the 3D models are updated in real time by using videos captured at many outdoor surveillance cameras. The videos are also used to renew landmark templates to ensure accurate landmark detection and to achieve robust estimation of the pose of the PDA. We set up four outdoor surveillance cameras in an experimental outdoor environment. We succeeded in estimating PDA's pose by the proposed landmark detection method so that objects moving at blind spots were visualized on the PDA display in real time.

Keyword Outdoor MR, PDA, Surveillance Camera, Landmark Detection, Pose Estimation

1. はじめに

複合現実感 (Mixed Reality:MR) とは、我々が存在する現実世界とコンピュータで生成した仮想世界を繋ぎ目なく融合した感覚を提示する技術の総称である。近年、この複合現実感を屋外環境中で実現することで、歩行者ナビゲーションや観光案内といった様々な応用が期待されている[1][2]。

我々は、カメラ付ハンディ型 PDA (以下、PDA) のディスプレイ中で複合現実感を実現し、屋外歩行者を想定したユーザが手軽に利用できる視覚情報提示システムを提案している[4]。本システムの大きな特徴は、屋外に監視目的などのために多数設置した定点カメラ映像を利用し、建物の陰などによって死角となりユーザ視点からは直接見ることでできない現実世界の状況を提示映像からリアルタイムに確認できることである。ユーザが違和感なく死角を認知するために、死角に存在する建物の CG モデルは幾何的整合性を保ち位置ず

れなく現実世界の映像上に重畳されなければならない。そのため、現実世界を撮影するカメラ (以下、PDA カメラ) の世界座標系における位置と向き (以下、位置姿勢) を正確に計測することが必要となる。この PDA カメラの位置姿勢計測には、物理センサのみを用いる方法、物理センサと画像センサを組み合わせる方法などが挙げられる。

物理センサのみを用いる方法では、GPS や地磁気方位センサなどを用いて屋外環境における PDA カメラの位置姿勢を初期的に推定し、慣性センサや歩測センサを用いてその変化を追従することが考えられる[1][2]。しかし、本システムが求める位置ずれのない提示を行うためには、最終的に画像上で重畳位置のずれを評価し解消することが必要不可欠である。

そこで本研究では、ランドマークを屋外環境中に多数配置し PDA カメラ映像中で数点検出することにより、物理センサで求めた位置姿勢の初期値を補正する

手法を提案する．重量位置のずれを，PDA カメラ映像上でランドマークの投影位置と検出位置の二次元の距離として測定するため，この位置ずれをなくす位置姿勢の補正値を容易に求めることが可能となる．しかし，すべてのユーザ行動範囲をカバーするだけのランドマークを環境中に配置することは困難である．また，探索候補となるランドマークを増やすと，計算コストが膨大になる．さらに，ランドマークと光源・PDA カメラの相対的な位置関係の変化により，ランドマークの見え方が変化する問題もある．

見え方が変化する問題に対し，我々はユーザに視覚情報を提供している定点カメラの映像をランドマークのテンプレート更新にも利用することで，ランドマーク検出を安定化させることを提案してきている[4]．これに加え本論文では，物理センサから計測値を利用することで探索するランドマーク候補を減らし，更にランドマークを階層化しそれぞれに異なった適切な探索方法を用いることで，より短い時間で正確な PDA カメラの位置姿勢推定を実現する．

以降，2 章では定点カメラ映像を用いた死角部分の提示について説明する．3 章では物理センサと画像センサを組み合わせた PDA カメラ位置姿勢推定法について述べ，4 章ではランドマーク観測による位置姿勢推定について詳しく説明する．5 章では定点カメラ 4 台を用いたシステムの実装について説明し，PDA カメラの位置姿勢補正，及び死角部分のリアルタイムを行った実験について述べる．最後に 6 章でまとめと今後の課題について述べる．

2. 定点カメラ映像を用いた死角部分の提示

本システムでの複合現実感技術を用いた視覚情報提示例を図 1 に示す．図 1 において，上段左に示すのが現実環境をモデル化した CG モデルである．この CG モデルは，建物施工時の立面図及び配置図から作成している．図 1 上段右に示すのが，PDA カメラで撮影された現実環境である．このカメラ撮像上に CG モデルを半透明化し位置ずれなく重畳することで，図 1 下段に示すように，死角部分の様子が透けて見えるような視覚情報提示が可能となる．しかし，オフラインで構築した CG モデルを提示するだけでは死角部分における人や自動車の動きといった現実世界の変化に対応できない．そのため，CG モデルを現実世界の変化に合わせて随時更新する必要がある．

本研究では，CG モデルの形状については変化しないと仮定し，テクスチャを現在のものにリアルタイムに更新する提示方法を提案する．このテクスチャ更新に，屋外で監視等の目的で設置されている定点カメラ映像を利用する．ユーザは人や自動車の存在を建物な

どの CG モデルテクスチャに映り込んでいる様子から認識できることになる．

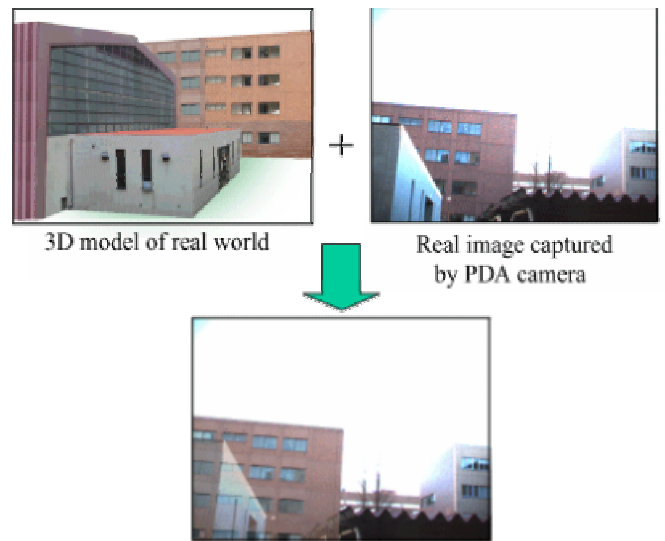


図 1：死角部分の視覚情報提示例

2.1. テクスチャの表現

CG モデルのテクスチャは方形で表現する．テクスチャ各端点について世界座標系での三次元座標，そのテクスチャに対応する定点カメラ画像上での二次元座標 c_k ($k=0,1,2,3$)，及び CG モデル中でテクスチャが乗る平面（以下，テクスチャ平面）上での二次元座標 g_k ($k=0,1,2,3$)をあらかじめ与えておく．

2.2. テクスチャの更新

CG モデル中のテクスチャそれぞれについて，世界座標系から PDA カメラ画像中への投影変換行列 P を用いて各テクスチャ端点を投影し，1 点以上 PDA カメラ画像内に投影されたテクスチャを更新対象として選択する．

更新対象として選択されたテクスチャについて，対応

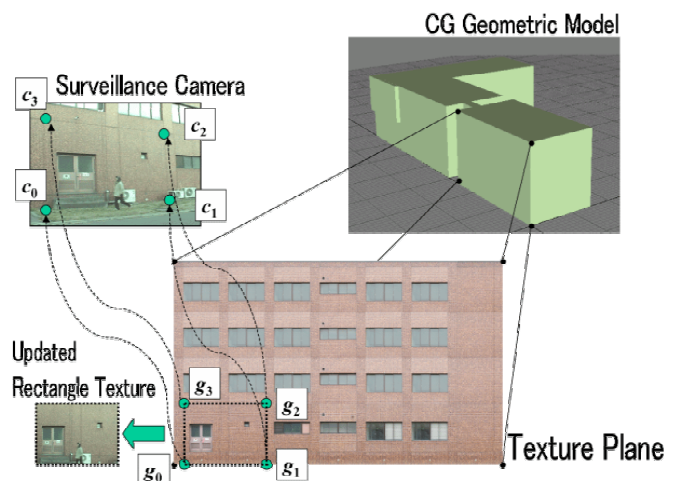


図 2：定点カメラ映像からのテクスチャ更新

付けられた定点カメラ映像から該当する部分画像を切り出し、最新のテクスチャに随時更新する。このとき、定点カメラ映像をテクスチャ平面に合わせるため、図2に示す $\mathbf{c}$ と $\mathbf{g}$ の関係から求めた二次元アフィン変換行列 $\mathbf{Q}$ を用いる (式 2-1)。

$$\lambda \begin{bmatrix} c_{xi} \\ c_{yi} \\ 1 \end{bmatrix} = \mathbf{Q} \begin{bmatrix} g_{xi} \\ g_{yi} \\ 1 \end{bmatrix} \quad (2-1)$$

3. PDA カメラの位置姿勢推定手順

前章で述べた CG モデルをユーザ視点の変化に応じて正確に PDA カメラ映像上に重畳するためには、PDA カメラの位置姿勢を正確に求める必要がある。PDA カメラの位置姿勢とは、世界座標系から PDA カメラ画像中への投影変換行列 $\mathbf{P}$ の外部パラメータである。

本研究では、以下の3つの手順からこの外部カメラパラメータを正確に求める。

- (1) 物理センサによる位置姿勢の初期値（初期外部カメラパラメータ）の推定
- (2) 慣性ジャイロセンサによる姿勢変化の追従
- (3) 画像センサによる初期外部カメラパラメータ補正

位置については、GPS や屋外環境中に埋め込んだビーコン、RFID タグ[1][3]などを用いて推定できるとされている。一方で姿勢は、地磁気方位センサなどの計測値から算出することが可能である。本研究で実装したプロトタイプシステムでは、位置は固定であるとし、姿勢の推定には INTERSENSE 社製の慣性ジャイロセンサ (InertiaCube²) を利用する。この慣性ジャイロセンサでは、北を X 軸、東を Y 軸、地面方向を Z 軸とするセンサ座標系からの roll, pitch, yaw の回転で姿勢の初期値を獲得することが可能である。

このようにして求めた初期外部カメラパラメータを 4×4 の変換行列 $\tilde{\mathbf{M}}$ で表現すると、世界座標系での三次元位置 $\mathbf{X}_w$ から PDA カメラ座標系での三次元位置 $\tilde{\mathbf{X}}_c$ への変換を式 3-1 のように表すことができる。

$$\tilde{\mathbf{X}}_c = \tilde{\mathbf{M}} \mathbf{X}_w \quad (3-1)$$

しかし、式 3-1 で求めた $\tilde{\mathbf{X}}_c$ は真の値 $\mathbf{X}_c$ に対してずれている可能性がある。これは現在位置として与える座標値が正確でないこと、電磁波などによって地面方向を回転軸とする Yaw の回転に誤差が生じ易いことに起因する。そこで PDA カメラ映像中から現実世界の中のランドマークを数点検出し、ICP アルゴリズムによって 4×4 の補正行列 $\mathbf{K}$ を求める。この $\mathbf{K}$ を式 3-2 のように掛け合わせ、初期カメラパラメータを補正する。

$$\tilde{\mathbf{X}}'_c = \mathbf{K} \tilde{\mathbf{M}} \mathbf{X}_w \quad (3-2)$$

初期外部カメラパラメータ補正後、再び慣性ジャイロセンサによって姿勢変化に追従して $\tilde{\mathbf{M}}$ を更新する。

4. ランドマーク観測による位置姿勢補正

ランドマークの検出方法として、松崎らは、定点カメラ映像を利用してランドマークのテンプレートを作成し天候や日照の変化に左右されないランドマーク検出手法を提案している[5]。この手法では、定点カメラで撮影されたランドマーク画像をテンプレートとして用意し、マッチングによって PDA カメラ映像中からランドマークの検出を行う。定点カメラで撮影しているランドマークがユーザ視点の PDA カメラ映像中に映りこんでいる可能性が高ければ検出することが可能であるため、様々な環境において利用できる可能性が高いと考えられる。

本研究では、建物の角や窓枠など特徴的な形状をランドマークとして利用し二次元画像（ランドマークテクスチャ）で表現する。ランドマークテクスチャの獲得には、ランドマーク観測専用の定点カメラを配置するのではなく、歩行者に視覚情報を提供する複数台の定点カメラの映像を援用する。定点カメラ映像を基にランドマークテクスチャをオンラインで随時更新することで、安定したランドマーク検出を実現する。以下、その検出方法と、複数の検出点から ICP アルゴリズムによって位置姿勢の補正を行うまでの手順を述べる。

4.1. ランドマークテンプレートの生成

画像によるマッチングを行う際、PDA カメラの位置姿勢変化から生じる PDA カメラ映像中と定点カメラ映像中でのランドマークの見え方の違いが問題となる。そこで図3に示すように、まず投影行列 $\mathbf{P}$ を用いてランドマークテクスチャを PDA カメラ映像中に投影し、ランドマークテクスチャがどのように見えるか調べる。次に、ランドマークテクスチャと定点カメラ映像との二次元アフィン変換を求め、変形後の対応する定点カメラ画像上の画素を内挿してランドマークテンプレートを生成する。

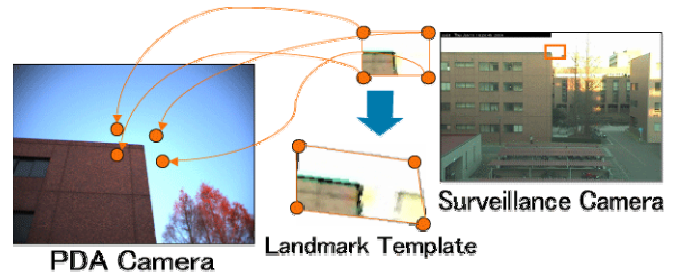


図3：ランドマークテンプレートの生成

4.2. ランドマークの分類と探索手順

ランドマークの探索には、比較的検出が容易であると考えられるランドマークから順に探索を行い、その結果に基づいて他のランドマークの探索を行う。1つのランドマークが正しくマッチングされた時点で、他のランドマークの探索領域を絞り込むことができ、マッチングに要する時間を削減することが可能になる。

ここで、ランドマーク探索を開始する際の PDA カメラ画像中には建物と空との境界線が映り込んでいるという前提条件を設け、この条件に基づいてランドマークを表 1 に示す 2 種類に分類する。各ランドマークの例を図 4 に示す。探索手順は、最初にフォーカスランドマークの探索を行い、その結果を基に周辺ランドマークの探索を行う。詳細を 4.2.1～4.2.4 節に示す。

表 1：ランドマークの分類

フォーカス ランドマーク Primary Landmark	空との境界部分を含む建物の角を中心とするランドマーク
周辺ランドマーク Secondary Landmark	フォーカスランドマークの周辺に位置し、窓のような複雑な模様をもつランドマーク

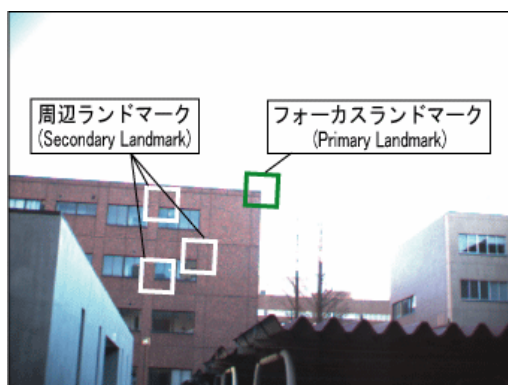


図 4：フォーカスランドマーク/周辺ランドマークの例

4.2.1. フォーカスランドマークの選択

補正処理開始時に算出した投影変換行列 $\mathbf{P}$ を基に、フォーカスランドマークの世界座標系における三次元位置を PDA カメラ画像中に投影する。このうち、画像中に投影されたものを探索用として選択する。

4.2.2. フォーカスランドマークの探索範囲決定

フォーカスランドマークの探索領域を、Shi-Tomasi の基準に基づく特徴点抽出法[6]と空領域抽出を組み合わせることで決定する。この探索領域とは、各ランドマークの重心が映り込んでいると予想される候補点の集合を指す。

ある点を中心とする一定範囲の全画素について、X、Y2 軸方向の輝度変化を調べ、その輝度変化の分布につ

いて直交方向に共に分散が大きければ、その点がフォーカスランドマークである可能性が高いことになる。そのため、Shi-Tomasi の基準に基づく特徴点抽出法を用いてフォーカスランドマークである建物の角部分を含むテクスチャを抽出することが可能である。一方、この手法のみでは他の直交成分を含むテクスチャも数多く検出される可能性がある。そこで空と建物の境界部分を抽出し、その境界部分に位置する画素でのみ Shi-Tomasi の基準に基づく特徴点抽出を行うことで、探索領域を減らしマッチングに要する計算コストを削減する。

図 5 左に示すように、補正処理開始時には PDA カメラ画像の上部に空領域が存在するようにする。全画素に対し求めた 2 軸方向の輝度変化の二乗和を求める。その結果が図 5 右のようになる。この結果に対し、図 6 左に示すように Y 軸正方向に前画素との差分を順次求める。この差分がある閾値を超えた時点での位置を空との境界点として抽出する。これを X 方向すべてに対して行い求めたのが図 6 右に示す境界部分である。

境界部分に対し Shi-Tomasi の基準に基づく特徴点抽出を行い、候補点を求めた結果が図 7 である。

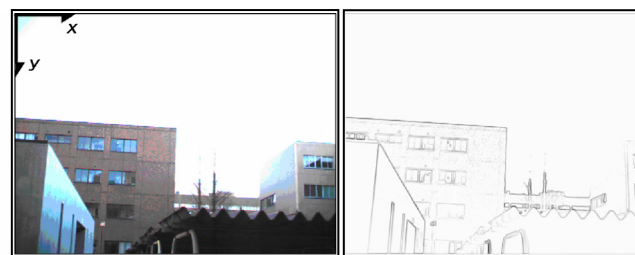


図 5：PDA カメラ画像（左）と輝度変化の二乗和を求めた結果画像（右）



図 6：空との境界部分の抽出

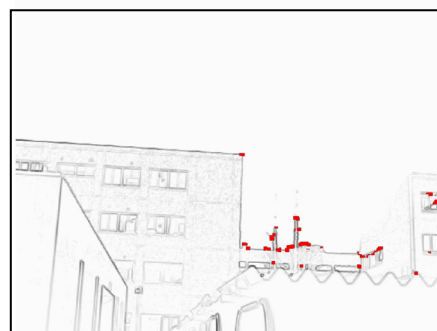


図 7：フォーカスランドマークの探索候補点（赤）

4.2.3. フォーカスランドマークの探索

前節で求めた候補点に対し、正規化相関に基づくテンプレートマッチングによって PDA カメラ画像中のフォーカスランドマークの位置を探索する。PDA カメラ画像中に映り込んでいる可能性のある F 個のフォーカスランドマークに対し、それぞれ探索を行い相関値 $c(-1 \leq c \leq 1)$ の最大値を求める。 F 個の中で最も相関値の高いフォーカスランドマークを第一フォーカスランドマークとして選択する。この第一フォーカスランドマークの相関値が最大となる PDA カメラ画像上の位置を $\mathbf{v}_0$ とする。

なお、他のフォーカスランドマークについては相関値の大きい順に第二、第三…とし、第一を用いた補正処理が失敗した場合に、順に利用する。

4.2.4. 周辺ランドマークの探索

次に選択された第一フォーカスランドマークに属する周辺ランドマークを探索する。投影変換行列 $\mathbf{P}$ を用いて第一フォーカスランドマークを PDA カメラ画像上へ投影した時の重心位置を $\mathbf{u}_0$ とする。 $\mathbf{d}_0 = \mathbf{v}_0 - \mathbf{u}_0$ だけずらした位置を中心とする一定範囲を探索領域とし、正規化相関に基づくマッチングによってそれぞれ探索を行う。相関値 c の最大値が閾値より大きい周辺ランドマークについて、信頼性の高いランドマークと見なし検出位置 $\mathbf{v}_i$ と投影位置 $\mathbf{u}_i$ ($i=1,2,3,\dots$) をそれぞれ求めておく。

4.3. ICP アルゴリズムによる補正行列の導出

PDA カメラ画像中で、フォーカスランドマーク 1 点と周辺ランドマーク 2 点以上検出した場合に、ICP (Iterative Closest Point) アルゴリズム[7][8]によって現在の位置姿勢の補正を行う。本研究では、4.2.3.及び 4.2.4.節で求めた検出位置 $\mathbf{v}$ と投影位置 $\mathbf{u}$ のずれ $\mathbf{d}$ を評価値とし、繰り返し演算によってこの評価値を最小化する位置姿勢の 6 自由度パラメータを算出する。この 6 自由度パラメータから式 3-2 の 4×4 の補正行列 $\mathbf{K}$ を求め、物理センサの計測値から得た $\mathbf{M}$ に掛け合わせる。

5. 実験

ランドマーク観測による位置姿勢補正と死角部分のリアルタイム提示を同時に実行する実験を行った。定点カメラを図 8 に示す実験環境中に 4 台設置し、すべての定点カメラ映像からオンラインで最新映像が取得できるようにシステムを構築した。定点カメラと PDA の間には、定点カメラ映像を管理し CG モデルテクスチャ及びランドマークテクスチャを生成するサーバを設けている。各定点カメラとサーバ間是有線で相互に通信を行い、サーバと PDA 間は無線 LAN (IEEE80

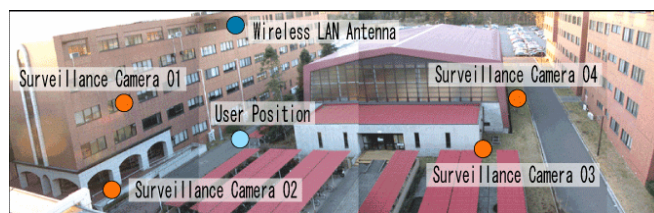


図 8：ユーザ位置と定点カメラの配置

2.11b) を経由して情報を伝送する。各定点カメラで撮影された画像は、JPEG 形式、704x480[pixel]の大きさでサーバに送信される。更新間隔は約 2.0[fps]である。以下に示す実験では、ユーザは図 8 中に示した位置にいるものとする。

5.1. ランドマーク観測による位置姿勢補正実験

構築した実験環境中で、PDA カメラ画像中からフォーカスランドマーク 1 点と周辺ランドマーク 2 点を検出し、ICP アルゴリズムによって PDA の位置姿勢を補正する実験を行った。補正処理の開始は、PDA カメラ画像中にフォーカスランドマークが映り込むようにユーザが PDA カメラの向きを合わせ、位置姿勢補正を要求した時点とした。

その結果、図 9 左に示すようにフォーカスランドマークである建物の角部分で 77[pixel]あったずれを、本補正手法によって図 9 右のように 1[pixel]まで補正することに成功した。補正処理の開始要求時から終了までに要した時間は約 8 秒である。表 2 に処理時間の内訳を示す。なお、一旦補正が行われれば、その後は得られた補正行列と慣性ジャイロセンサの出力に従うことで、PDA カメラを任意方向に向けても幾何的に一致した映像を提供できる。慣性ジャイロセンサの累積誤差が大きくなってきたら、再度補正処理を要求する。



図 9：ランドマーク観測による位置姿勢補正

表 2：位置姿勢補正処理時間の内訳

定点カメラ映像からランドマークテクスチャの切り出し	6.8-7.1[s]
選択された第一フォーカスランドマークと従属するランドマークの送受信	0.3-0.4[s]
ランドマークテンプレートの生成と探索が完了するまでの時間	1.8-1.9[s]
ICP アルゴリズムによる補正	0.1[s]

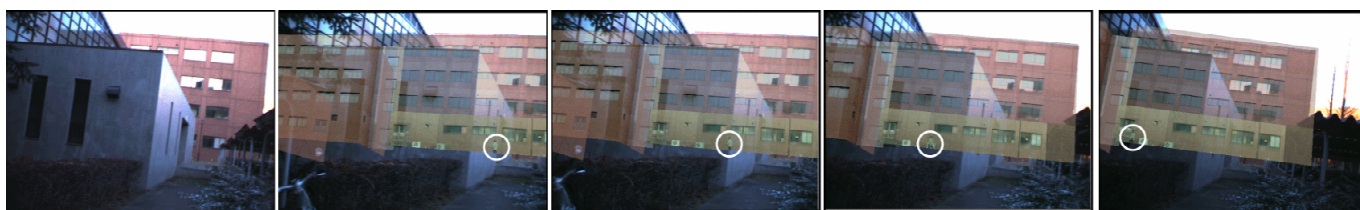


図 10：死角部分のリアルタイム提示

5.2. 死角部分のリアルタイム提示実験

次に、定点カメラ 4 台の映像から CG モデルを獲得し更新することでユーザに死角部分の状況をリアルタイムに提示する実験を行った。提示する CG モデルとオフラインで用意した初期の CG モデルテクスチャの情報は、PDA のシステム起動時にサーバから受信する。オンラインで更新する CG モデルテクスチャは図 11 左上に示すように事前に指定されている。サーバは、図 11 中の定点カメラ映像中の①～③に相当する領域から CG モデルテクスチャを更新しユーザに送信する。

実験結果の映像から、死角部分の様子を 1 秒程度の遅れで確認することができた。図 10 の実験結果例では、一番左に示すようにユーザからは見ることが出来なかった向こうの建物正面を別の歩行者が歩いている様子が確認されている。

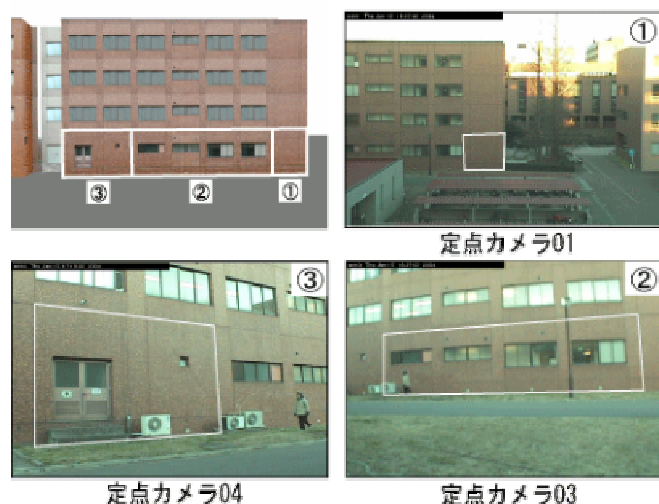


図 11：各定点カメラ映像から生成したテクスチャ

6. まとめと今後の課題

本研究では、死角に存在する建物の CG モデルについて、幾何的整合性を保ち位置ずれなく PDA カメラ映像上に重畳するために、物理センサによる PDA カメラの位置姿勢計測を定点カメラ映像を援用したランドマーク観測によって補正することを提案した。

本手法の特徴の一つは、ランドマークの探索において、物理センサの計測値を利用して探索するランドマ

ーク候補の絞り込みを行い、かつ、ランドマークをテクスチャの特徴によって階層化し検出が容易なランドマークから探索することによって、PDA カメラの位置姿勢補正に要する時間を削減していることである。

実際に定点カメラを屋外環境中に 4 台設置してシステムを実装し、ランドマーク観測による位置姿勢補正と死角部分のリアルタイム提示を同時に実行する実験を行い、提案手法の有効性を確認した。

今後の課題としては、広範囲に渡る屋外環境中でランドマークに相当する部分構造体を CAD データなどから自動的に抽出する方法や、ユーザに対する死角部分提示方法の評価などが挙げられる。

文 献

- [1] 天目陵平, 神原誠之, 横矢直和, “赤外線ビーコンと歩数計測を用いたウェアラブル型注視提示システム,” 信学技報, IE2002-54, 2002.
- [2] 佐藤清秀, 穴吹まほろ, 山本裕之, 田村秀行, “屋外装着型複合現実感のためのハイブリッド位置合わせ手法,” 日本バーチャルリアリティ学会論文誌, vol.7, No.2, pp.129-137, 2002.
- [3] 山下淳, 檜山敦, 雨宮智浩, 小林一郎, 広田光一, 廣瀬通孝, “ウェアラブル, およびモバイルコンピュータ用屋外仮想研究環境の構築,” ヒューマンインタフェース学会報告集, vol.4, No.5, pp.45-48, ヒューマンインタフェース学会 第 20 回研究会, 東京, 2002.
- [4] 武政泰輔, 亀田能成, 大田友一, “定点カメラ映像を用いた歩行者のための屋外型複合現実感システム,” 信学技報, PRMU2003-198, 2004.
- [5] 松崎誠, 大田友一, “定点観測カメラを援用した屋外における複合現実感のためのランドマーク検出,” 信学総大, No.D-12-83, 2001.
- [6] Jianbo Shi, Carlo Tomasi, “Good Features to Track,” IEEE, Computer Vision and Pattern Recognition (CVPR94), pp.593-600, 1994.
- [7] P. J. Besl and N. D. McKay, A Method for Registration of 3-D Shapes: “A Method for Registration of 3-D Shapes,” IEEE Trans. On PAMI, vol.14, No.2, pp.239-256, 1992.
- [8] 内山晋二, 山本裕之, 田村秀行, “複合現実感のためのハイブリッド位置合わせ手法—6 自由度センサと ICP アルゴリズムの併用—,” Proc.MIRU2002, pp.I.107-I.112, 2002.