

推薦論文

歩行者群の移動軌跡情報を用いた モバイルカメラ画像内の人物位置推定手法

岩橋 宏樹^{1,a)} 樋口 雄大^{1,b)} 山口 弘純^{1,c)} 東野 輝夫^{1,d)}

受付日 2014年5月14日, 採録日 2014年11月10日

概要: 本論文では、レーザレンジスキャナ (LRS) により検出した群衆の移動軌跡情報と、その群衆内の歩行者がモバイルカメラで撮影した周辺環境の視覚情報をもとに、撮影者自身およびカメラ画像中に写る周辺人物の現在位置を特定する手法を提案する。提案手法では、撮影された静止画に対して画像解析アルゴリズムを適用し、モバイル端末上で、撮影者と画像内の人物との相対位置を推定する。これらの局所的な人物相対位置情報をサーバ上へ収集し、LRS の計測値に基づく高精度な群衆位置情報とマッチングすることで、撮影者およびカメラ画像中に写る各人物の絶対位置を特定している。以上により得られる位置情報をモバイル端末へフィードバックすることで、拡張現実感 (AR) を用いた高度な歩行者ナビゲーション等の実現を目指している。シミュレーション実験により提案システムの性能を評価した結果、カメラ画像内の人物群と LRS により検出された歩行者位置との対応関係を、最大 83% の精度で推定できることが分かった。

キーワード: 群衆センシング, 画像解析, レーザレンジスキャナ, 拡張現実感

Utilizing Crowd Trajectories for Mobile AR Applications

HIROKI IWAHASHI^{1,a)} TAKAMASA HIGUCHI^{1,b)} HIROZUMI YAMAGUCHI^{1,c)} TERUO HIGASHINO^{1,d)}

Received: May 14, 2014, Accepted: November 10, 2014

Abstract: This paper presents an algorithm to recognize the current locations of mobile device users by fusion of human trajectories obtained by accurate crowd tracking sensors and images that are captured with built-in cameras in their devices. By providing the resulting position estimates to the mobile devices, we aim to support augmented reality (AR) applications such as advanced pedestrian navigation. Through simulation experiments with realistic sensor models, we show that our system can estimate locations of the device holders with high accuracy.

Keywords: crowd sensing, image analysis, laser range scanners, augmented reality

1. はじめに

我々の研究グループでは、これまで、都市の様々な共有空間における人や環境のセンシングに関する研究を実施してきた。文献 [1] ではレーザレンジスキャナ (LRS) を

利用した群衆センシングプラットフォーム「ひとがつながるなび」の設計開発について述べ、文献 [2] ではその上位概念として、群衆やその行動および思考、ならびに周辺環境のセンシングデータといった複数のセンシングソースを集約・解析し、商業ビルやオフィス、イベント会場といった共有空間における混雑ナビゲーションやソーシャルコミュニケーション支援、あるいは人々の存在や行動に応じた省電力制御といった様々な応用システムに活用するためのプラットフォーム「ひとなび」の設計開発と商業施設での実証実験について述べている (図 1)。LRS を用いたトラッ

¹ 大阪大学大学院情報科学研究科
Graduate School of Information Science and Technology,
Osaka University, Suita, Osaka 565-0871, Japan

a) h-iwahashi@ist.osaka-u.ac.jp

b) t-higuti@ist.osaka-u.ac.jp

c) h-yamagu@ist.osaka-u.ac.jp

d) higashino@ist.osaka-u.ac.jp



図 1 グランフロント大阪におけるひとたびシステムの実行画面
Fig. 1 Crowd navigation system in the Grand Front Osaka.

キングでは物体検出距離誤差がきわめて小さい（一般に数 cm 程度）ため、群衆の移動軌跡情報を高精度に出力できることが分かっている。

こうした群衆位置情報を歩行者ナビゲーション等に应用する場合、以下の要求があると考えられる。

(1) 群衆内での自己位置の特定：パーソナルナビゲーションでは、各ユーザの現在位置からの視点で群衆情報等を可視化したい場合が多い。このためには、そのユーザの正確な位置を取得する必要がある。LRS による群衆位置情報群は精度が高いため、ユーザ位置がそのうちのいずれに相当するかを特定すればそのユーザの正確な位置が分かるが、その特定のためには何らかの追加情報を要する。我々はこれまでに、ユーザ間の近接性を Bluetooth 等の無線 PAN から取得し、LRS の群衆位置情報との整合性をとることで群衆内での位置を特定する方法 [3] や、無線 LAN による位置推定結果から同様に位置を特定する方法 [4] を提案し、簡易な実証実験を実施している。しかし、前者は周辺人物の無線端末保持率に大きく影響を受けるため使用状況が限定され、後者では無線 LAN 位置推定の精度上の問題から群衆内の位置を一意に特定するには至っていない。

(2) 拡張現実感 (AR) への活用：(1) において、各スマートフォンユーザの位置が正確に特定されることで、たとえば、スマートフォンやスマートグラス等に標準搭載されたカメラ（モバイルカメラ）を介した AR 技術により、カメラ画像上の人物に対しその人物のスマートフォンが発信する情報を重畳することもできる。また、(1) で得られる正確な位置情報を付与した Tweet を、カメラ画像内の人物に重畳して表示させることも可能である（図 2）。このためには、カメラ画像内の人物と LRS によって検出された人物位置との対応関係を得る必要がある。

こうした要求に対し、本研究では、LRS により検出した歩行者位置群の中から、歩行者がモバイルカメラで撮影した画像中に写る人物像を検出し、それを用いて撮影者自身の位置を特定する手法を提案する。提案手法では、我々が開発している群衆センシングシステム「ひとたび」のようなシステムを用い、パーティ会場やイベント会場等の共有



図 2 拡張現実感 (AR) への活用
Fig. 2 AR-based human navigation.

空間での群衆の動きを LRS により匿名トラッキングし、群衆ナビゲーションに利用する状況を考える。このもとで、モバイルカメラから得られる画像を既存の顔検出ならびに人物像検出アルゴリズムを用いて解析し、周辺人物群への相対距離を事前学習により構築した距離予測モデルに基づき推定する。これにより得られる周辺人物との相対的な位置関係と最も整合性の高い LRS の群衆位置をその撮影者の位置とすることで、撮影者ならびに周辺人物の位置特定を行う。これらの位置情報は共有空間での高精度パーソナルナビゲーション等に利用できるほか、カメラ画像内の周辺人物にそのユーザの情報を重畳することで、人物属性情報やその人物の発信情報を直感的にユーザへ提示する AR ナビゲーションが実現できる。

630 枚の実写真をもとに構築した人物距離推定モデルを用い、20m × 20m の領域において 25~150 人が閉空間を移動する環境を想定したシミュレーションにより提案手法を評価した結果、最大正解率 83%での周辺人物の位置特定が実現できることを確認している。

2. 関連研究と提案手法の位置づけ

2.1 モバイル端末向け屋内測位技術

スマートフォン等市販のモバイル端末向けの屋内測位技術としては、Wi-Fi フィンガープリントに基づく手法が広く利用されている。RADAR [5] や Horus [6] は、建物内の各地点においてアクセスポイントからの受信電波強度を事前計測し、モバイル端末上で観測された受信電波強度と事前学習データとをマッチングすることで、端末の位置を推定する。また近年では、EZ localization [7] や Zero Configuration localization [8] のように、事前のキャリブレーションのコストを抑えた手法も提案されている。これらの手法は、既存の Wi-Fi 通信インフラを活用して低コストに屋内測位を実現できるという利点があるが、歩行者の

本論文の内容は 2013 年 12 月のマルチメディア通信と分散処理ワークショップにて報告され、同研究会主査により情報処理学会論文誌ジャーナルへの掲載が推薦された論文である。

体による電波の減衰や信号のマルチパス伝搬等の影響で、同じ位置においても受信電波強度は大きく変動するため、数十 m 程度の大きな誤差が生じる場合も多い。また、モバイル端末に内蔵されたモーションセンサを用いて端末保持者の移動軌跡を推定する歩行者自律航法 (PDR; pedestrian dead reckoning) 技術も活発に研究されている。スマートフォン等の市販デバイスを想定した手法としては、加速度センサを用いた歩行ステップの検出と、地磁気センサやジャイロセンサによる移動方向の検出を組み合わせるアプローチが一般的である [9], [10]。PDR は、端末保持者の位置の相対的な変化量を取得できるが、端末の絶対位置を得るためには、Wi-Fi 測位等他の測位技術を併用する必要がある。また、センサノイズや端末の想定外の動き等に起因して、時間の経過とともに軌跡推定誤差が急速に蓄積されるという問題がある。このように、市販のモバイル端末を対象とした従来の位置推定技術によって、モバイル AR 等のアプリケーションに供するに足る十分な位置推定精度を実現することは難しい。

2.2 歩行者トラッキング技術

イベント会場や商業施設等の公共空間において、群衆の移動特性や環境内の状況を正確に把握することを目指して、設置型センサを用いた歩行者トラッキング手法が活発に研究されている。なかでも、防犯カメラ等の映像に画像解析アルゴリズムを適用することで歩行者の数や移動軌跡を推定する映像ベースのトラッキング技術は、既存のインフラを活用できるといった利点から、様々なアプローチが検討されている [11]。背景差分に基づく歩行者検出手法 [12], [13] は、カメラから取得されたフレーム画像を、歩行者がいない時間に撮影した背景画像と比較し、画素値の差異が大きい領域を歩行者として検出する。これらの手法は、少ない計算量で歩行者のトラッキングを実現できるという利点がある一方、背景の変化による誤検出をいかにして軽減するかが課題となる。これに対し、オブジェクトセグメンテーション [14] は、グラフ理論等を応用することにより、背景差分の計算を行うことなく、画像中から人の形状を直接抽出する。また、大規模な画像データベースからオブジェクトの外観を表す特徴量を抽出し、パターンマッチングによって歩行者の顔や体を検出する手法も活発に検討されている。文献 [15] では、2つの領域間の輝度差 (Haar-like 特徴量) をもとに画像内の局所的なエッジ成分をとらえ、AdaBoost アルゴリズムを用いて、カスケード型の分類器を構成している。文献 [16] は、画像の局所領域から輝度の勾配と強度のヒストグラム (HOG 特徴量; Histograms of Oriented Gradients) を抽出し、サポートベクタマシンとの組合せにより、高精度な歩行者検出を実現している。

こうした映像ベースの手法に加えて、レーザレンジス

キャナ (LRS) を利用した群衆トラッキング技術も近年高い注目を集めており、防犯システム等の用途で商用利用が進められている [19]。最も一般的なアプローチは、LRS を人の腰の高さに合わせて地面と水平に設置し、各時刻の計測情報の差分をもとに歩行者の移動軌跡を抽出する方式である [3], [20], [21]。また、センサ周辺を通行する歩行者等がレーザ光を遮蔽すること (オクルージョン) によって、センサから遠い位置にいる歩行者の検出率が低下することを可能な限り避けるため、人の足首の高さに LRS を設置する方式も検討されている [22], [23]。

LRS を用いた歩行者トラッキングで推定できるのは、環境内の各地点における人の存在情報のみである。映像ベースのトラッキング技術では、これに加えて、衣服の色といった歩行者の視覚的な特徴も同時に取得することができるものの、いずれの手法も、トラッキングによって得られた歩行者群の位置情報と、歩行者が保持するモバイル端末との対応関係が未知であることから、1章で述べたようなモバイル AR アプリケーションに直接適用することは難しい。

2.1 節で述べたモバイル端末向けの屋内測位技術とは異なり、映像および LRS を用いた歩行者トラッキング手法は、いずれも誤差 1 m 以内の高精度な位置推定が実現可能である。また、いずれの手法においても、オクルージョンによる歩行者の検出漏れは発生するものの、カメラや LRS から直接捕捉可能な位置に存在する歩行者については、高い精度で検出することができる [3], [11]。したがって、測位精度や歩行者の認識率の観点では、2つのアプローチの間に大きな優劣はないといえる。一方で、カメラの視野角には一般に厳しい制約があり、映像ベースのトラッキング手法で歩行者の移動軌跡を正確に推定するためには、環境内に多数のカメラを設置する必要がある。たとえば、文献 [24] では、広角カメラを高さ 2.1 m の天井に設置することで歩行者のトラッキングを実現しているが、歩行者の全身を捕捉できる領域は、センサの設置位置を中心とする 3 m × 4 m の狭い領域に限られている。これに対し、LRS は、1 台のセンサで半径数十 m の領域をカバーできるため (たとえば、市販の LRS デバイスである UTM-30LX [25] は、距離 30 m、中心角 270° の扇形領域を計測可能である)、インフラ構築やシステムの運用に要するコストを効果的に軽減することができる。このため、本論文では、LRS を用いた歩行者トラッキング技術を想定してアルゴリズムを設計する。

2.3 提案手法の位置づけ

本論文で提案するシステムは、LRS による広域の群衆の観測情報と、スマートグラス等のモバイルカメラから得られる局所的な観測情報を融合することで、画像の撮影者およびその周辺人物の位置を特定するこれまでにない手法である。いったん認識された位置情報は、LRS による継続的トラッキングにより維持できるため、群衆の移動に対し

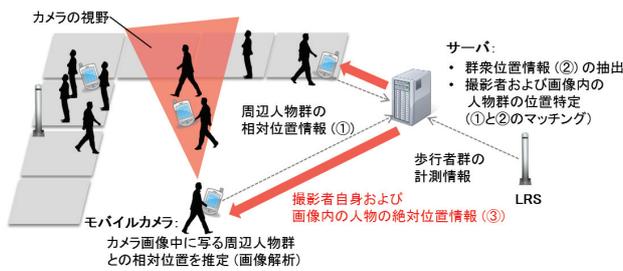


図 3 提案システムのアーキテクチャ

Fig. 3 Architecture of the proposed system.

ても有効であり，友人の位置トラッキング等にも応用できる．このように，モバイルカメラから得られる視覚情報とインフラベースの高精度トラッキング情報を組み合わせることで，高度な群衆センシングを実現する手法は，我々の知る限りこれまでに提案されておらず，新規性が高い．また，シミュレーション実験を通じて，最大 83% の正解率を達成する等，有用性を定量的に示している．

3. システムの概要と予備実験

3.1 システムの概要

提案システムのアーキテクチャを図 3 に示す．環境内には 1 台以上の LRS が設置されており，一定の時間間隔で，スキャン領域内の物体との距離が計測される．これらの計測情報は，ネットワークを介して定期的にサーバへ収集され，サーバ上で文献 [20], [21] 等の歩行者検出アルゴリズムを適用することで，スキャン領域内の歩行者群の現在位置および移動軌跡が抽出される．

また，環境内に滞在する一部の歩行者はスマートグラスやスマートフォン等のモバイルカメラ端末（モバイル端末と呼ぶ）を利用して，周辺の歩行者群の画像を定期的に取り得る．モバイルカメラは，撮影したカメラ画像に対して Haar-like 特徴量および HOG 特徴量に基づく人物検出アルゴリズム（後述）を適用し，カメラの視野内の周辺人物との相対位置を推定する．これらの画像解析アルゴリズムは，いずれも計算量が比較的小さく，スマートフォン等市販のモバイル端末上でのリアルタイム処理が可能である．そこで，本システムでは，画像の解析をモバイル端末上で行い，モバイル端末群とサーバとの間で共有される情報は，近隣の歩行者群との相対位置データに限定する．これにより，サーバ上で，顔認識に基づく個人特定や，推定年齢・性別といったプライバシー性の高い情報の収集が行われることを回避できるほか，カメラ画像をクラウドへアップロードし，画像解析処理をクラウド上へオフロードするアプローチと比較して，通信オーバーヘッドの大幅な軽減が可能となる．

サーバ上では，LRS から得られる移動軌跡情報と，モバイル端末群から収集した周辺人物の相対位置情報を統合することで，撮影者である各モバイル端末の現在位置およびカメラの姿勢を推定する．さらに，カメラ画像内の人物と

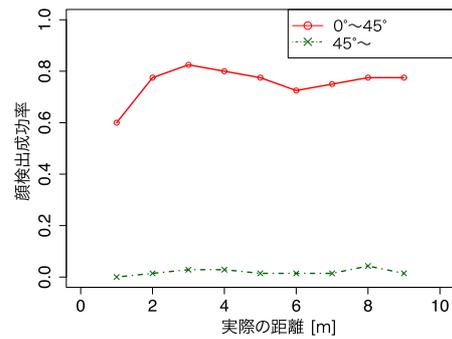


図 4 顔検出成功率

Fig. 4 Face detection rate.

LRS の計測結果に基づく歩行者群の位置情報との対応付けを行う．撮影者自身およびカメラ画像中の周辺人物の現在位置の推定結果は，撮影者のモバイル端末へとフィードバックされ，AR ナビゲーションをはじめとするモバイル端末向けの位置情報サービスに活用される．

3.2 予備実験

カメラ画像に基づく周辺人物の検出率および相対位置推定精度をモデル化するため，スマートフォン端末（Google Nexus S）を用いた予備実験を行った．

はじめに，同一の地点で直立する被験者を，モバイル端末の内蔵カメラを用いて様々な距離・角度から繰り返し撮影することで，人物画像のデータセットを収集した．撮影環境は，蛍光灯で照らされた室内とし，被験者の背景は白色の壁とした．カメラの高さは撮影者の目線に合わせ，約 1.5m とした．カメラから被験者までの距離は 1~9m (1m 刻み) とし，カメラに対する被験者の相対的な方位角を 0° (対面)~90° (横向き) の範囲で 15° 刻みで変化させた．10 名の被験者に対して，すべての距離・角度の組み合わせで撮影を行い，合計 630 枚の人物画像を取得した．これらの画像に対して，顔検出アルゴリズムおよび体検出アルゴリズムを適用し，被験者の顔および体の検出率と，画像内で顔や体として検出された領域（検出フレーム）のサイズを評価した．

なお，顔検出および体検出には，それぞれ，OpenCV の標準ライブラリで実装されている Haar-like 特徴量 [26] および HOG 特徴量に基づく検出アルゴリズムを適用した．これらのアルゴリズムは，計算量が少なく，スマートフォン等のモバイル端末上でリアルタイムな検出処理を実現できることから，顔および体検出の標準的な手法として広く利用されている [27]．

はじめに，被験者からの距離と顔検出の成功率との関係を図 4 に示す．顔検出の成功率は，カメラから見た被験者の相対的な方位角に依存する．被験者の向きが対面から ±45° 以下の場合には，平均で約 80% の検出率が得られているのに対し，対面状態からの角度のずれが 45° を超える

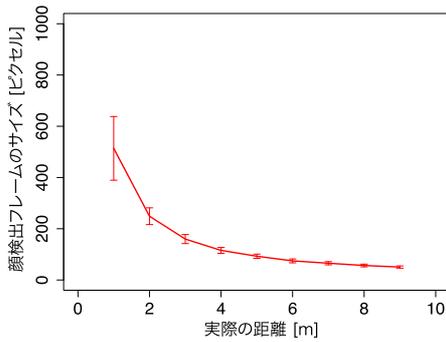


図 5 顔検出フレームのサイズ

Fig. 5 Size of the face detection frames.

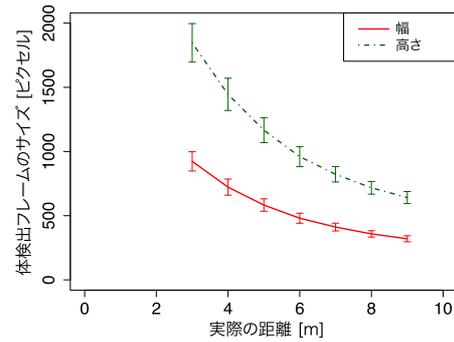


図 7 体検出フレームのサイズ

Fig. 7 Size of body detection frames.

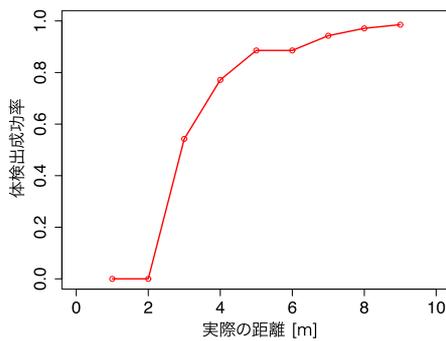


図 6 体検出成功率

Fig. 6 Body detection rate.

と、検出率が急激に低下している。

また、被験者からの距離と顔検出フレームのサイズとの関係を図 5 に示す。なお、OpenCV の顔検出アルゴリズムでは、検出された顔領域が正方形で出力されるため、ここでは、顔検出フレームの 1 辺の長さ（ピクセル数）をフレームサイズと定義する。また、グラフ内のエラーバーはフレームサイズの標準偏差を表している。

被験者からの距離が長くなるにつれ、検出フレームのサイズは徐々に小さくなる傾向がある。特に、被験者からの距離が 3m 以内の場合には、距離の増加にともないフレームサイズが大きく低下しており、顔領域のフレームサイズをもとに、被験者との距離を比較的高精度に推定できると考えられる。一方、被験者からの距離が 3m を超えると、距離の変化に対するフレームサイズの変化が小さくなり、顔領域のサイズに基づく距離の判別は困難となる。

次に、被験者との距離と体検出の成功率の関係を図 6 に示す。被験者からの距離が 3m 以下の場合には、被験者の体全体が画像内に収まらないケースが頻繁に発生し、検出率が低くなっている。一方、被験者との距離が 3m を上回る場合には、80%を超える高い検出率が得られている。また、顔検出とは異なり、被験者からの距離が同一であれば、カメラに対する被験者の方位角によらず、ほぼ一定の検出率が得られることが分かった。

最後に、被験者からの距離と検出フレームのサイズとの関係を、図 7 に示す。OpenCV の体検出アルゴリズムで

は、体として検出された領域が長方形で出力されるため、同図では、検出フレームの幅と高さの平均および標準偏差をそれぞれピクセル単位で示している。顔検出と同様に、被験者からの距離が離れるにつれて検出フレームのサイズが小さくなる傾向が確認できる。また、顔検出フレームに基づく距離の判別が難しい、3m より遠くにいる歩行者についても、距離の増加に応じて一定の割合でフレームサイズが変化していることが分かる。

このように、歩行者からの距離と顔・体検出のフレームサイズの間には強い相関があるため、カメラ画像に対して顔検出および体検出のアルゴリズムを適用し、検出フレームのサイズを評価することで、カメラの視野内の歩行者との距離を推定することができる。特に、撮影者と歩行者との距離が 3m 以下の場合には顔検出に基づく距離推定、3m を上回る場合には体検出に基づく距離推定がそれぞれ効果的であるといえる。

体検出に基づく距離推定においては、検出フレームの幅に比べて、高さ方向のサイズの方が対象物との距離の差に対する変化量が大いものの、一般に、検出対象の人物の身長差等に起因して、同じ距離に存在する人物の間でも検出フレームの高さには比較的大きなばらつきが生じることが想定される。また、対象物の一部が他の歩行者や障害物によってカメラの視界から頻繁に遮られることが想定される展示会場のような環境では、ばらつきがより大きくなる可能性が高い。以上の理由から、提案手法では、体検出フレームの幅をもとに、距離推定を行うものとする。

3.3 顔および体の検出結果に基づく相対位置推定

3.2 節の予備実験によって得られた、カメラ視野内の歩行者からの距離と、顔検出および体検出のフレームサイズとの関係を用いれば、カメラ画像をもとに、周辺人物との距離を推定することができる。提案手法では、顔検出フレームの 1 辺の長さ（図 5）と、体検出フレームの幅（図 7）をもとに、カメラ視野内の歩行者との距離 d を推定する。画像のサイズを $X \times Y$ 、カメラ画像中における検出フレームの中心点の位置を (x, y) 、カメラの水平画角を Θ_{cam} とす

ると、カメラの位置を原点、カメラの視線を x 軸とするカメラ座標系における周辺人物の位置 $\tilde{\mathbf{v}}$ は式 (1) によって推定することができる。

$$\tilde{\mathbf{v}} = \left(d \cos \left\{ \left(\frac{x}{X} - 0.5 \right) \Theta_{cam} \right\}, d \sin \left\{ \left(\frac{x}{X} - 0.5 \right) \Theta_{cam} \right\} \right) \quad (1)$$

提案手法では、カメラ画像に対して前述の顔検出アルゴリズムおよび体検出アルゴリズムを適用し、検出フレームの位置およびサイズに基づき、それぞれ、歩行者位置を推定する。そのうえで、顔検出アルゴリズムに基づく周辺人物の相対位置推定結果から、推定距離が 3m を上回る歩行者を除外するとともに、体検出アルゴリズムに基づく相対位置推定結果から推定距離が 3m 以下となる歩行者を取り除き、残った周辺人物の相対位置情報を重畳することで、最終的な相対位置結果を出力する。

4. システム設計

4.1 問題の定式化

対象領域に存在するすべての歩行者の集合を \mathcal{S} とする。対象領域内には 1 台以上の LRS が設置されており、一定の時間間隔で、スキャン領域内に存在する歩行者群の絶対位置が高精度に取得される。ここで、LRS によって検出された歩行者位置の集合を **LRS ビュー** と呼び、 $\mathcal{L} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ で表す。LRS による歩行者トラッキングでは、オクルージョンによって一部の歩行者に対する検出漏れが発生する可能性があるため、LRS ビュー内の歩行者数は、一般に、対象領域内に実在する歩行者の総数よりも少なくなる ($|\mathcal{L}| \leq |\mathcal{S}|$)。また、一部の歩行者 $\mathcal{S}_{cam} \subseteq \mathcal{S}$ は、スマートフォンやスマートグラス等のカメラ機器を保持しており、カメラから取得される周辺人物の画像に対して画像解析アルゴリズムを適用することで、カメラの視野内に存在する人物との相対位置を推定できるものとする。本論文では、歩行者 $s \in \mathcal{S}_{cam}$ のカメラから検出された周辺人物の相対位置の集合を歩行者 s の**カメラビュー** と呼び、 $\tilde{\mathcal{C}}_s = \{\tilde{\mathbf{v}}_0\} \cup \{\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2, \dots, \tilde{\mathbf{v}}_m\}$ で表す。ここで、カメラビュー内の歩行者位置 $\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2, \dots, \tilde{\mathbf{v}}_m \in \tilde{\mathcal{C}}_s$ は、 s のカメラ座標系から見た周辺人物の推定位置である。また、 $\tilde{\mathbf{v}}_0$ は同座標系における s 自身の位置を表し、つねに $\tilde{\mathbf{v}}_0 = (0, 0)$ とする。

いま、現時刻の LRS ビュー \mathcal{L} および歩行者 s のカメラビュー $\tilde{\mathcal{C}}_s$ が与えられたとする。このとき、カメラビュー内の各人物 $\tilde{\mathbf{v}}_i \in \tilde{\mathcal{C}}_s$ に対応する LRS ビュー内の歩行者 $\mathbf{u}_j \in \mathcal{L}$ を同定するためのマッピング関数 $f: \tilde{\mathcal{C}}_s \rightarrow \mathcal{L} \cup \{\phi\}$ を求めることが提案手法の目的である。ここで、 $f(\tilde{\mathbf{v}}_i) = \phi$ は、オクルージョン等に起因する歩行者の検出漏れによって、 $\tilde{\mathbf{v}}_i \in \tilde{\mathcal{C}}_s$ に対応する歩行者位置が LRS ビュー \mathcal{L} に含まれないことを表す。

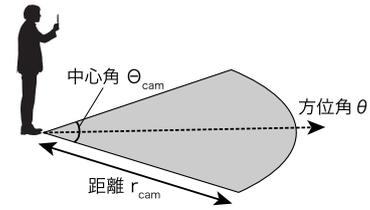


図 8 カメラモデル

Fig. 8 Camera model.

4.2 提案手法の概要

提案手法では、カメラの視野を、図 8 のような半径 r_{cam} 、中心角 Θ_{cam} の扇形の領域でモデル化する。このもとで、 s のカメラの視野に相当する扇形領域内の歩行者位置 $\mathcal{L}_s \subseteq \mathcal{L}$ を LRS ビューから切り出し、 s の現在位置 \mathbf{p}_s を原点、カメラの方位角を x 軸とするカメラ座標系に変換する。本論文では、変換後の歩行者位置の集合 $\tilde{\mathcal{L}}_s = \{\tilde{\mathbf{u}}_1, \tilde{\mathbf{u}}_2, \dots, \tilde{\mathbf{u}}_{n'}\}$ を、 s から見た**ローカル LRS ビュー** と呼ぶ。前述のカメラモデルより、 s の現在位置 \mathbf{p}_s およびカメラの方位角 θ_s が与えられれば、 \mathcal{L}_s および $\tilde{\mathcal{L}}_s$ は一意に定まる。一方、初期状態では、カメラの位置 \mathbf{p}_s および方位角 θ_s はいずれも未知である。そこで、提案手法では、 \mathbf{p}_s および θ_s を確率変数とみなしてローカル LRS ビューの候補 $\tilde{\mathcal{L}}_s^{(k)}(\mathbf{p}_s, \theta_s)$ ($k = 1, 2, \dots, N_{view}$) をランダムに生成し、カメラビュー $\tilde{\mathcal{C}}_s$ との整合性が最も高い候補 $\tilde{\mathcal{L}}_s'$ を、ローカル LRS ビューの推定値とする。ローカル LRS ビューの推定方法については、4.3 節で述べる。

その上で、カメラビュー $\tilde{\mathcal{C}}_s$ とローカル LRS ビューの推定結果 $\tilde{\mathcal{L}}_s'$ を比較し、それぞれのビューに含まれる歩行者位置間のユークリッド距離をもとに、最も確からしいマッピング $\tilde{f}: \tilde{\mathcal{C}}_s \rightarrow \tilde{\mathcal{L}}_s' \cup \{\phi\}$ を求める。さらに、ローカル LRS ビュー $\tilde{\mathcal{L}}_s'$ の各要素と LRS ビュー \mathcal{L} に含まれる歩行者位置との対応関係 $f_0: \tilde{\mathcal{L}}_s' \rightarrow \mathcal{L}$ は、ローカル LRS ビューの生成過程より既知である。これら 2 つの関数を結合することで、 $\tilde{\mathcal{C}}_s$ から \mathcal{L} へのマッピング関数 $f = f_0 \circ \tilde{f}$ が得られる。カメラビューから LRS ビューへのマッピング関数の導出については、4.4 節で述べる。

4.3 ローカル LRS ビューの推定

本節では、歩行者 $s \in \mathcal{S}_{cam}$ のカメラビュー $\tilde{\mathcal{C}}_s$ をもとに、LRS ビュー \mathcal{L} から、 s のカメラの視野に相当する領域を切り出し、ローカル LRS ビュー $\tilde{\mathcal{L}}_s$ を求める方法について述べる。以降では、議論の簡単化のため、対象領域を等間隔のグリッドに分割し、領域内の座標をグリッド上のセルで離散的に表現する。また、対象領域内のすべてのセルからなる集合を \mathcal{U} とする。

$\tilde{\mathcal{L}}_s$ を算出するためには、歩行者 $s \in \mathcal{S}_{cam}$ の正確な現在位置 $\mathbf{p}_s \in \mathcal{U}$ とカメラの方位角 θ_s が必要となるが、前述のとおり、初期状態においてこれらはいずれも未知である。

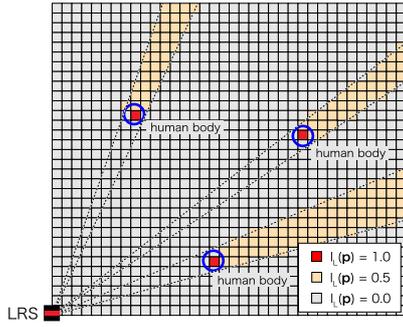


図 9 歩行者存在尤度 $l_{\mathcal{L}}(\mathbf{p})$ の空間的分布
 Fig. 9 Spatial distribution of $l_{\mathcal{L}}(\mathbf{p})$.

そこで、提案手法では、 \mathbf{p}_s および θ_s を確率変数とみなして、 N_{view} 通りの $\tilde{\mathcal{L}}_s$ の候補をランダムに生成する。 \mathbf{p}_s については、LRS から得られる計測結果を事前情報として利用することが可能である。LRS の計測値からは、検出された歩行者群の位置情報 \mathcal{L} に加えて、オクルージョンまたは LRS のカバレッジの制約によっていずれの LRS にも計測されなかったセルの集合 $\mathcal{O}_{\mathcal{L}}$ を求めることができる。これらの情報をもとに、対象領域内のそれぞれのセル $\mathbf{p} \in \mathcal{U}$ に歩行者が存在する尤度 $l_{\mathcal{L}}(\mathbf{p})$ を式 (2) により定義する。

$$l_{\mathcal{L}}(\mathbf{p}) = \begin{cases} 1.0 & \text{if } \mathbf{p} \in \mathcal{L} \\ 0.5 & \text{if } \mathbf{p} \in \mathcal{O}_{\mathcal{L}} \\ 0.0 & \text{otherwise} \end{cases} \quad (2)$$

ここで、LRS の計測値に基づく歩行者存在尤度 $l_{\mathcal{L}}(\mathbf{p})$ の空間的な分布の例を図 9 に示す。図内の丸印は、対象領域内において実際に歩行者が存在する位置を表している。LRS によるトラッキングでは、レーザ光を用いた正確な測距に基づき、歩行者の存在および位置を高い精度で検出できる。そこで、LRS からのレーザ光が到達しているセルについては、そのセル内に歩行者が検出された場合 $l_{\mathcal{L}}(\mathbf{p}) = 1.0$ 、そうでない場合 $l_{\mathcal{L}}(\mathbf{p}) = 0.0$ と考える。一方、対象領域内の歩行者や障害物等によって LRS からのレーザ光が届かない領域については、LRS の計測結果から歩行者の有無を判断することができない。このため、LRS から見て歩行者の影になっているセルについては、 $l_{\mathcal{L}}(\mathbf{p}) = 0.5$ と定義している。提案手法では、この分布を式 (3) により正規化することで、 \mathbf{p}_s の事前分布 $P(\mathbf{p}_s)$ を定義する。

$$P(\mathbf{p}_s) = \frac{l_{\mathcal{L}}(\mathbf{p}_s)}{\sum_{\mathbf{p} \in \mathcal{U}} l_{\mathcal{L}}(\mathbf{p})} \quad (3)$$

一方、カメラの方位角 θ_s に関する情報は LRS の計測結果からは得られないため、 θ_s の事前分布は式 (4) で与えられるものとする。

$$P(\theta_s) = \frac{1}{2\pi} \quad (4)$$

式 (3) および式 (4) の事前分布より、確率変数 \mathbf{p}_s 、 θ_s の実現値 $\mathbf{p}_s^{(k)}$ 、 $\theta_s^{(k)}$ をそれぞれサンプリングすることで、ローカル LRS ビューの候補 $\tilde{\mathcal{L}}_s^{(k)}$ が生成される。ここで、歩行

者の位置 \mathbf{p}_s と方位角 θ_s が互いに独立であると仮定すると、 $\tilde{\mathcal{L}}_s^{(k)}$ の事前分布 $P(\tilde{\mathcal{L}}_s^{(k)})$ は、式 (5) により与えられる。

$$P(\tilde{\mathcal{L}}_s^{(k)}) = P(\mathbf{p}_s = \mathbf{p}_s^{(k)}) \cdot P(\theta_s = \theta_s^{(k)}) \quad (5)$$

以上により生成された N_{view} 通りのローカル LRS ビューの候補の中から、最も確からしい候補を選択することで、ローカル LRS ビューの推定値 $\tilde{\mathcal{L}}'_s$ を求める。一般に、 $\mathbf{p}_s^{(k)}$ および $\theta_s^{(k)}$ が実際のカメラの位置および方位角に近いほど、 $\tilde{\mathcal{L}}_s^{(k)}$ とカメラビュー $\tilde{\mathcal{C}}_s$ に含まれる歩行者位置の類似度が高まると考えられる。そこで、提案手法では、それぞれの候補 $\tilde{\mathcal{L}}_s^{(1)}, \tilde{\mathcal{L}}_s^{(2)}, \dots, \tilde{\mathcal{L}}_s^{(N_{view})}$ とカメラビュー $\tilde{\mathcal{C}}_s$ との整合性に基づき、 $\tilde{\mathcal{C}}_s$ が与えられたときの $\mathcal{L}_s^{(k)}$ に対する事後確率 $P(\tilde{\mathcal{L}}_s^{(k)} | \tilde{\mathcal{C}}_s)$ を算出する。

カメラビューの誤差要因としては、主として (i) 画像解析の失敗に起因する歩行者の検出漏れ、(ii) カメラビュー内の歩行者位置の誤差、および (iii) オクルージョンに起因する歩行者の検出漏れの 3 つが考えられる。3 章の予備実験の結果より、画像解析による近隣の歩行者 $s' \in \mathcal{S}$ の検出率は、撮影者から s' までの距離 $d_{s'}$ に依存し、関数 $\gamma(d_{s'})$ で与えられる。同様に、 s' との距離およびカメラビュー上における偏角の推定誤差 e_d 、 e_θ の分布も、 $d_{s'}$ に依存して、それぞれ $\alpha(e_d | d_{s'})$ 、 $\beta(e_\theta | d_{s'})$ でモデル化される。さらに、カメラビュー内の歩行者位置 $\mathbf{v}_i \in \tilde{\mathcal{C}}_s$ より、カメラから見て歩行者の死角となっているセルの集合 $\mathcal{O}_{\mathcal{C}}$ を推定することが可能である。

上記の誤差モデルをもとに、カメラビュー内に検出された各歩行者位置 $\tilde{\mathbf{v}}_i \in \tilde{\mathcal{C}}_s$ について、 $\tilde{\mathbf{v}}_i$ に対応する歩行者 s_i の実際の位置 $\tilde{\mathbf{p}}_{s_i}$ の分布 $P(\tilde{\mathbf{p}}_{s_i} | \tilde{\mathbf{v}}_i)$ を式 (6) のように定める。なお、歩行者 s_i との実際の距離 d_{s_i} は未知であることから、ここでは、 d_{s_i} を s_i の推定位置との距離 $\|\tilde{\mathbf{v}}_i\|$ で近似する。

$$P(\tilde{\mathbf{p}}_{s_i} | \tilde{\mathbf{v}}_i) = \alpha(\|\tilde{\mathbf{p}}_{s_i}\| - \|\tilde{\mathbf{v}}_i\|) \|\tilde{\mathbf{v}}_i\| \cdot \beta((\arg(\tilde{\mathbf{p}}_{s_i}) - \arg(\tilde{\mathbf{v}}_i)) \|\tilde{\mathbf{v}}_i\|) \quad (6)$$

また、セル $\tilde{\mathbf{p}}$ に存在する歩行者が、画像解析の失敗によって検出されない確率は、 $1 - \gamma(\|\tilde{\mathbf{p}}\|)$ で与えられる。このもとで、提案手法では、カメラ視野内の各セル $\tilde{\mathbf{p}}$ における歩行者の存在尤度 $l_{\tilde{\mathcal{C}}_s}(\tilde{\mathbf{p}})$ を式 (7) のように定める。

$$l_{\tilde{\mathcal{C}}_s}(\tilde{\mathbf{p}}) = \begin{cases} \max \left\{ \max_{\tilde{\mathbf{v}}_i \in \tilde{\mathcal{C}}_s} \frac{P(\tilde{\mathbf{p}}_{s_i} = \tilde{\mathbf{p}} | \tilde{\mathbf{v}}_i)}{P(\tilde{\mathbf{p}}_{s_i} = \tilde{\mathbf{v}}_i | \tilde{\mathbf{v}}_i)}, 1 - \gamma(\|\tilde{\mathbf{p}}\|) \right\} & \text{if } \tilde{\mathbf{p}} \notin \mathcal{O}_{\mathcal{C}} \\ 0.5 & \text{if } \tilde{\mathbf{p}} \in \mathcal{O}_{\mathcal{C}} \end{cases} \quad (7)$$

さらに、ローカル LRS ビューの候補 $\tilde{\mathcal{L}}_s^{(k)}$ によって切り出された領域内の各セル $\tilde{\mathbf{p}}$ について、式 (2) と同様に、LRS の計測情報に基づく歩行者存在尤度 $l_{\tilde{\mathcal{L}}_s^{(k)}}(\tilde{\mathbf{p}})$ を求める。以上により算出されるカメラの視野内の歩行者存在尤

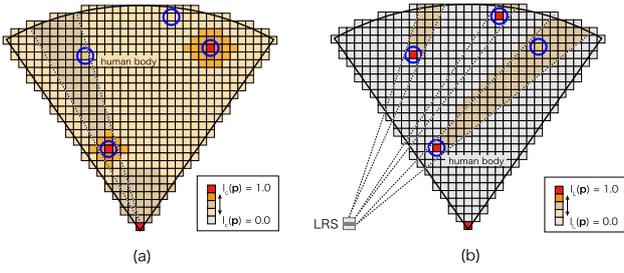


図 10 カメラ視野内の歩行者存在尤度の空間的分布 :
(a) $l_{\tilde{C}_s}(\hat{\mathbf{p}})$, (b) $l_{\tilde{L}_s^{(k)}}(\hat{\mathbf{p}})$

Fig. 10 Example likelihood distributions:
(a) $l_{\tilde{C}_s}(\hat{\mathbf{p}})$, (b) $l_{\tilde{L}_s^{(k)}}(\hat{\mathbf{p}})$.

度 $l_{\tilde{C}_s}(\hat{\mathbf{p}})$, $l_{\tilde{L}_s^{(k)}}(\hat{\mathbf{p}})$ の例を、それぞれ図 10 (a), (b) に示す。同図内の丸印は、それぞれのビュー内において実際に歩行者が存在する位置を表しており、カメラや LRS から見て歩行者の死角になっているセルでは、歩行者の検出漏れが生じている。また、カメラビュー上では、オクルージョンによる検出不能領域（オクルージョン領域）以外の地点においても、画像解析の失敗によって、一部の歩行者が検出されない場合がある。

このもとで、提案手法では、 $\tilde{L}_s^{(k)}$ と \tilde{C}_s の整合性を次のように定量化する。まず、カメラビューに基づく歩行者存在尤度 $l_{\tilde{C}_s}(\hat{\mathbf{p}})$ に対する $\tilde{L}_s^{(k)}$ の適合度 $Q_{\tilde{L}_s^{(k)} \rightarrow \tilde{C}_s}$ を式 (8) により算出する。

$$Q_{\tilde{L}_s^{(k)} \rightarrow \tilde{C}_s} = \frac{\sum_{\tilde{\mathbf{u}}_j \in \tilde{L}_s^{(k)}} l_{\tilde{C}_s}(\tilde{\mathbf{u}}_j)}{|\tilde{L}_s^{(k)}|} \quad (8)$$

同様に、ローカル LRS ビューの候補 $\tilde{L}_s^{(k)}$ に基づく歩行者存在尤度 $l_{\tilde{L}_s^{(k)}}(\tilde{\mathbf{u}}_j)$ に対する \tilde{C}_s の適合度 $Q_{\tilde{C}_s \rightarrow \tilde{L}_s^{(k)}}$ を式 (9) により定義する。

$$Q_{\tilde{C}_s \rightarrow \tilde{L}_s^{(k)}} = \frac{\sum_{\tilde{\mathbf{v}}_i \in \tilde{C}_s} \max_{\tilde{\mathbf{p}}} \frac{P(\tilde{\mathbf{p}}_{s_i} = \tilde{\mathbf{p}} | \tilde{\mathbf{v}}_i)}{P(\tilde{\mathbf{p}}_{s_i} = \tilde{\mathbf{v}}_i | \tilde{\mathbf{v}}_i)} \cdot l_{\tilde{L}_s^{(k)}}(\tilde{\mathbf{p}})}{|\tilde{C}_s|} \quad (9)$$

最後に、 $Q_{\tilde{L}_s^{(k)} \rightarrow \tilde{C}_s}$ および $Q_{\tilde{C}_s \rightarrow \tilde{L}_s^{(k)}}$ を式 (10) により統合することで、ローカル LRS ビューの候補 $\tilde{L}_s^{(k)}$ に対するカメラビュー \tilde{C}_s の尤度 $P(\tilde{C}_s | \tilde{L}_s^{(k)})$ を定義する。

$$P(\tilde{C}_s | \tilde{L}_s^{(k)}) = \min(Q_{\tilde{L}_s^{(k)} \rightarrow \tilde{C}_s}, Q_{\tilde{C}_s \rightarrow \tilde{L}_s^{(k)}}) \quad (10)$$

ベイズの定理より、カメラビュー \tilde{C}_s が与えられたときの $\tilde{L}_s^{(k)}$ の事後確率 $P(\tilde{L}_s^{(k)} | \tilde{C}_s)$ は式 (11) のように求まる。

$$P(\tilde{L}_s^{(k)} | \tilde{C}_s) = \frac{P(\tilde{C}_s | \tilde{L}_s^{(k)}) \cdot P(\tilde{L}_s^{(k)})}{P(\tilde{C}_s)} \propto P(\tilde{C}_s | \tilde{L}_s^{(k)}) \cdot P(\tilde{L}_s^{(k)}) \quad (11)$$

$\tilde{L}_s^{(1)}, \tilde{L}_s^{(2)}, \dots, \tilde{L}_s^{(N_{view})}$ の中から、 $P(\tilde{L}_s^{(k)} | \tilde{C}_s)$ が最大となるような候補を選択することで、ローカル LRS ビューの推定値 \tilde{L}'_s を決定する。

4.4 マッピング関数の導出

最後に、カメラビュー \tilde{C}_s とローカル LRS ビューの推定結果 \tilde{L}'_s を比較し、式 (12) により、最も確からしいマッピング $\tilde{f}: \tilde{C}_s \rightarrow \tilde{L}'_s \cup \{\phi\}$ を決定する：

$$\tilde{f}(\tilde{\mathbf{v}}_i) = \begin{cases} \arg \min_{\tilde{\mathbf{u}}_j \in \tilde{L}'_s} \|\tilde{\mathbf{v}}_i - \tilde{\mathbf{u}}_j\| & \text{if } \min_{\tilde{\mathbf{u}}_j \in \tilde{L}'_s} \|\tilde{\mathbf{v}}_i - \tilde{\mathbf{u}}_j\| < D \\ \phi & \text{otherwise} \end{cases} \quad (12)$$

このように、ローカル LRS ビューに含まれる歩行者位置 $\tilde{\mathbf{u}}_j \in \tilde{L}'_s$ の中で、 $\tilde{\mathbf{v}}_i \in \tilde{C}_s$ とのユークリッド距離が最小となるものを $\tilde{\mathbf{v}}_i$ に対応付ける。ただし、ローカル LRS ビュー上で、 $\tilde{\mathbf{v}}_i$ から距離 D 未満の領域に、LRS によって検出された歩行者位置が 1 つも存在しない場合には、 $\tilde{f}(\tilde{\mathbf{v}}_i) = \phi$ とする。本論文では経験的に $D = 2\text{m}$ とする。

5. ランドマークの利用

環境内に、あらかじめ正確な位置が分かっている目印（ランドマーク）が存在し、カメラ画像をもとにそれらのランドマークとの相対位置を推定できる場合には、ローカル LRS ビューの候補をより効率的に絞り込むことができる。

ローカル LRS ビューの候補 $\tilde{L}_s^{(k)}$ の視野内におけるランドマークの有無および各ランドマークとの相対位置は、ランドマークの正確な位置情報と、カメラの位置 $\mathbf{p}_s^{(k)}$ および方位角 $\theta_s^{(k)}$ の仮定から一意に定まる。ここで、 $\tilde{L}_s^{(k)}$ の視野内に存在するランドマークとの相対位置の集合を $\tilde{M}_{\tilde{L}_s^{(k)}}$ 、カメラ画像から検出されたランドマークとの相対位置の推定値の集合を $\tilde{M}_{\tilde{C}_s}$ とする。提案手法では、ローカル LRS ビューの候補集合 $\{\tilde{L}_s^{(1)}, \tilde{L}_s^{(2)}, \dots, \tilde{L}_s^{(N_{view})}\}$ から、式 (13) の条件を満たさない候補をあらかじめ取り除く。

$$\forall \tilde{\mathbf{p}}_c \in \tilde{M}_{\tilde{C}_s}, \exists \tilde{\mathbf{p}}_l \in \tilde{M}_{\tilde{L}_s^{(k)}}; \|\tilde{\mathbf{p}}_c\| - \|\tilde{\mathbf{p}}_l\| < D_{LM} \quad (13)$$

ここで、 D_{LM} はカメラビューとローカル LRS ビューの視野内におけるランドマークのマッチングの閾値であり、画像解析に基づくランドマークとの距離の推定精度に応じて決定する。

ローカル LRS ビューの候補の絞り込みに用いるランドマークとして、本論文では、LRS を設置するための金属製のポール（幅 0.14m、高さ 1.12m）を利用する。まず、ポールを様々な距離・角度から撮影した学習用の画像データを用意し、OpenCV の標準ライブラリとして実装されている、Haar-like 特徴量に基づく物体検出アルゴリズムを用いて、ポールの検出器を構築した。上記の検出器に対して、同様に様々な距離・角度からポールを撮影したテスト用の画像データ集合を入力し、ポールの検出率を評価した結果を図 11 に示す。ポールからの距離が 1m 未満の場合には、ポールの全体がカメラ画像内に収まらないケースが頻繁に発生し、検出率が約 50% となっている。一方、ポールからの距離が 2m 以上離れている場合には、70~80% 程

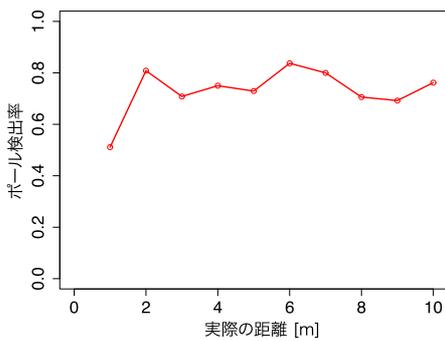


図 11 ポールの検出率

Fig. 11 Detection rate of LRS holders.

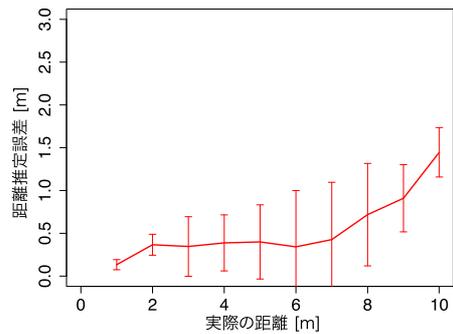


図 13 距離推定の誤差モデル

Fig. 13 Distance estimation errors.

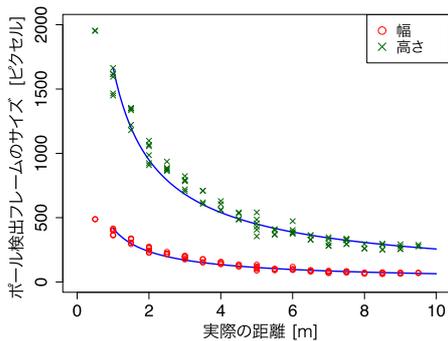


図 12 ポール検出に基づく距離推定モデル

Fig. 12 Ranging model for LRS holders.

度の高い検出率が得られていることが分かる。また、LRS ポール以外の領域が誤ってポールとして検出されたケースは、テスト用の画像データ集合全体のうち1%未満であり、高い適合性が実現されている。また、ポール検出器に学習用の画像データを入力し、ポールからの距離と検出フレームのサイズ（フレームの幅および高さのピクセル数）の関係を評価した結果を図 12 に示す。3.2 節の予備実験と同様、ポールからの距離に応じて検出フレームは徐々に小さくなっており、検出フレームのサイズをもとに、ポールまでの距離を推定することが可能である。本論文では、体検出に基づく周辺人物との距離推定と同様に、検出されたポール領域の幅をもとに、ポールからカメラまでの距離を推定する。最後に、前述の検出器を用いてテスト用の画像データ集合から LRS ポールを検出し、図 12 の距離推定モデルを用いてポールまでの距離を推定した場合の測距誤差の平均と標準偏差を図 13 に示す。ポールからの距離が 5m 未満の場合には、高い確率で誤差 1m 以内の高い測距精度が得られている。距離が 5m を超えると、距離の増加に対するフレームサイズの変化量が小さくなることから、測距誤差が急激に増加しているが、距離が 10m 以下の場合には、ほとんどのケースで、測距誤差は 2m 以内に抑えられている。以上の実験結果に基づき、6 章のシミュレーション実験では、 $D_{LM} = 2m$ とする。

6. シミュレーション

6.1 シミュレーション設定

提案手法によるマッピングの精度を検証するため、シミュレーションによる性能評価を行った。シミュレーションでは、20m × 20m の対象領域内に 25~150 名の歩行者をランダムに配置し、その中の 1 名がカメラを保持しているものとした。また、図 14 (b) に示した 4 地点に、それぞれフィールドの中央に向かって LRS を設置し、LRS のスキャン領域は、半径 30m、方位角 270° の扇形領域でモデル化した。なお、対象領域をグリッドに分割する際のセルの大きさは、0.25m × 0.25m とする。

LRS による歩行者トラッキングでは、人の体を円周 1m の円でモデル化し、LRS と歩行者との位置関係をもとに、オクルージョン領域を算出する。また、いずれかの LRS で検出された歩行者については、正しい現在位置に対応する LRS ビュー上のセルを特定できるものとした。

カメラビューについても、同様の方法でオクルージョン領域を算出し、この領域内の歩行者はカメラ画像からは検出できないものとする。カメラ画像に基づく歩行者の検出においては、視野内の各歩行者の距離および向きに依存して、3.2 節の検出率モデル（図 4、図 6）に応じた検出漏れが発生するものとする。また、検出に成功した歩行者との相対位置の推定誤差は、図 5 および図 7 の誤差モデルに基づき決定する。なお、Nexus S を用いた事前実験の結果に基づき、カメラモデルのパラメータは、 $r_{cam} = 10m$ 、 $\Theta_{cam} = 50^\circ$ とした。

なお、上記の想定環境において、ローカル LRS ビューの候補の生成数 N_{view} の値を変化させながら、事前に繰り返しシミュレーション実験を行ったところ、LRS ビュー内の各歩行者位置に対して 70 通り以上の候補を生成すれば、マッピング精度がほぼ一定の値に収束することが分かった。そこで本章の性能評価では、 N_{view} を LRS ビュー内の歩行者位置の数 × 70 とする。

以上の想定のもとで、提案手法によって、カメラビュー内の歩行者群を LRS ビュー内の歩行者位置へと正しくマッ

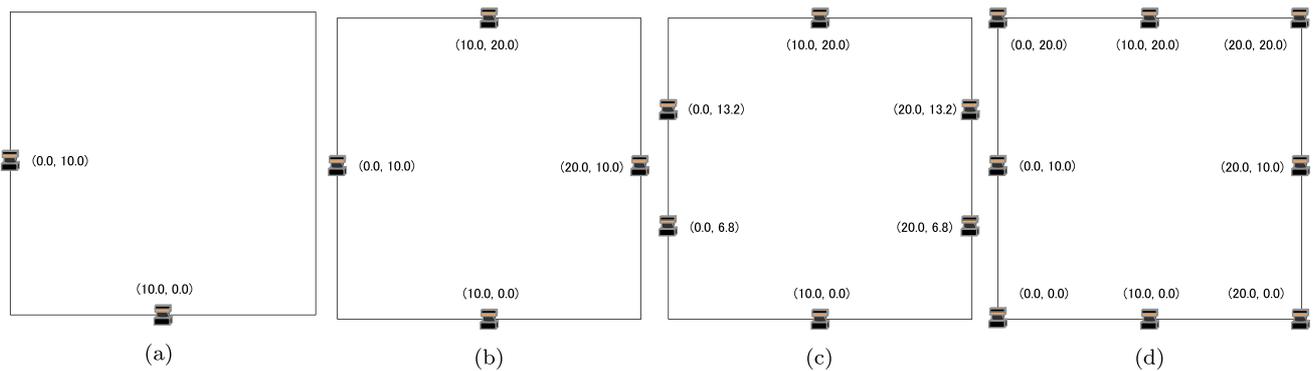


図 14 LRS の配置

Fig. 14 Sensor placement.

ピングできた割合（マッピング精度）を評価した。各シミュレーション実験においては、はじめに、対象領域内のランダムな位置に、ランダムな向きで歩行者群を配置する。カメラビュー内の歩行者位置と LRS ビュー内の歩行者位置とのマッピングは、それぞれのモバイルカメラごとに独立に実行されるため、何名の歩行者がモバイル端末を保持しているかは、提案手法の性能には影響しない。そこで、本実験では、対象領域内のすべての歩行者がカメラを保持していると想定して、それぞれの歩行者の視点から見た周辺歩行者群のマッピング精度を評価した。歩行者の位置および向きを変化させながら、それぞれのパラメータ設定について上記の実験を 250 回行い、すべての歩行者のカメラビューに対するマッピング精度の平均値を算出した。なお、提案手法は、カメラ画像の撮影時刻における歩行者間の相対的な位置関係をもとに前述のマッピングを実現していることから、カメラを保持した歩行者が移動していたとしても、マッピング精度への影響はないと考えられる。また、カメラビュー内に歩行者が 1 名も検出されなかった場合については、LRS ビュー内の歩行者位置と関連付けるべき情報そのものが存在しないことから、評価の対象外とした。

6.2 基本性能

はじめに、歩行者の数を 25 名または 50 名とし、ランドマーク（ポール）の検出結果に基づくローカル LRS ビューの候補のフィルタリングを行った場合と、行わなかった場合のマッピング精度をそれぞれ評価した。各シナリオにおけるマッピング精度の評価結果を図 15 に示す。ランドマークを用いた候補の絞り込みを行わない場合には、マッピング精度が約 50% となり、マッピングの誤りが比較的頻繁に発生している。特に、カメラビュー内に検出された歩行者の数が少ない場合には、歩行者位置のみからローカル LRS ビューの候補を一意に絞り込むことが難しく、誤ったマッピングの発生頻度が高くなることが分かった。一方、ランドマークの検出結果に基づきローカル LRS ビューの候

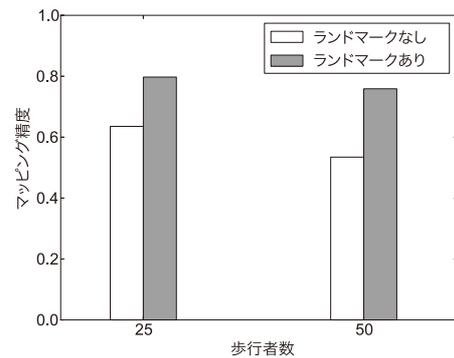


図 15 シミュレーション結果

Fig. 15 Basic performance.

補を事前に絞り込んだ場合には、マッピングの精度が 70~80% まで改善しており、誤った候補を効果的にフィルタリングできていることが分かる。このように、環境内の少数のランドマークを活用することで、精度の高いマッピングを実現することができる。なお、シミュレーション実験で生成した 250 通りのカメラビューのうち、1 名以上の歩行者が検出されたものの割合は、歩行者数が 25 の場合 58%、歩行者数が 50 の場合 73% となり、撮影者および周辺人物の位置を高い精度で推定できることを確認している。

6.3 歩行者密度の影響

歩行者の密度が提案手法の性能に与える影響を検証するため、歩行者数を 25~150 名の間で変化させながら繰り返しシミュレーション実験を行った。それぞれの歩行者数に対する提案手法のマッピング精度を図 16、カメラビュー内に 1 名以上の歩行者が検出される割合を図 17、カメラビューおよびローカル LRS ビュー（LRS ビューからカメラビューに相当する扇形領域を切り出したもの）内におけるオクルージョン領域の割合の平均値をそれぞれ図 18、図 19 に示す。なお、本実験では、いずれのシナリオにおいても、4 地点に設置されたランドマーク（ポール）の検出結果に基づき、ローカル LRS ビューの候補の絞り込みを行うものとした。歩行者数が 75 以下の場合には、70%を

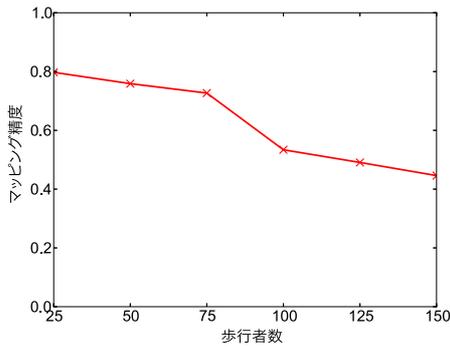


図 16 マッピング精度

Fig. 16 Mapping accuracy.

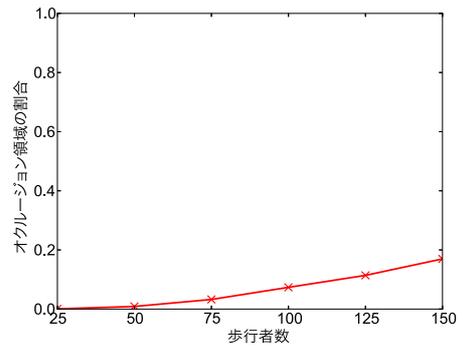


図 19 ローカル LRS ビュー内のオクルージョン領域

Fig. 19 Occluded area in local LRS views.

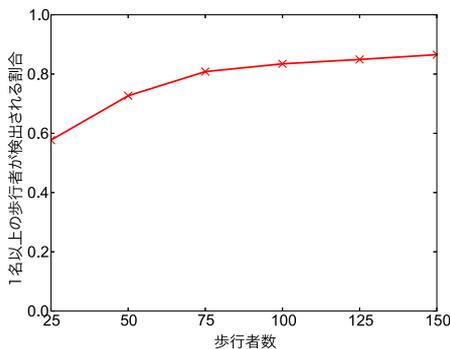


図 17 歩行者検出率

Fig. 17 Pedestrian detection rate.

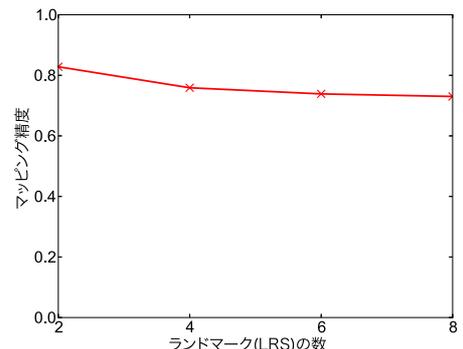


図 20 ランドマーク数の影響

Fig. 20 Impact of the number of landmarks.

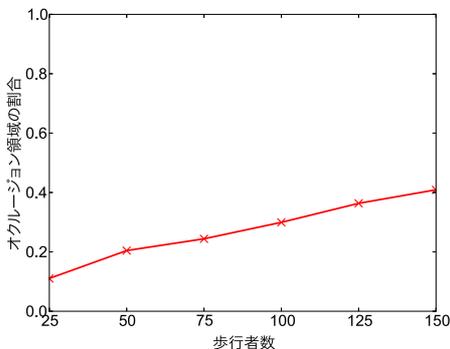


図 18 カメラビュー内のオクルージョン領域

Fig. 18 Occluded area in camera views.

超える高いマッピング精度が実現されている。一方で、歩行者数が増加すると、オクルージョンによる歩行者の検出漏れの頻度が徐々に高くなり、歩行者の総数が 100 以上になると、マッピングの精度が 45~53%まで急激に低下している。ローカル LRS ビュー内におけるオクルージョン領域の大きさは、想定される歩行者の密度に応じて LRS の設置数を増やすことで、ある程度の軽減が可能である。一方で、カメラビューは、各撮影者による単一視点の観測情報であるため、オクルージョンの発生を抑えることは本質的に困難である。これらの結果から、提案手法は、対象領域内の歩行者密度が 0.19 人/m² (75 人/400 m²) 未満である環境において有効な手法であるといえる。

6.4 ランドマーク密度の影響

最後に、環境内のランドマークの密度が提案手法の性能に与える影響を評価するため、LRS (ポール) の設置数を 2~8 の間で変化させて実験を行った。なお、本実験では、歩行者の数を 50 とし、図 14 (a)~(d) に示した位置に LRS を設置した。また、他の実験と同様、LRS はランドマークとしてだけでなく計測装置としても利用しているが、本節の実験シナリオでは、LRS の設置数が 2 の場合でもローカル LRS ビュー内のオクルージョン領域の割合は平均 1%未満となっており、計測装置としての LRS の設置数の増加がマッピング精度へ与える影響は無視できる程度に小さいと考えられる。シミュレーションによる評価結果を図 20 に示す。LRS の設置数が 2 の場合には 83%のマッピング精度が得られているのに対し、LRS の設置数を 8 とした場合のマッピング精度は 73%となっている。これは、ランドマークとしての LRS (ポール) が増えたことに起因している。たとえば、対象領域内に同種のランドマークが 1 つしかない場合には、カメラ画像内にそのランドマークがとらえられた時点で、撮影者の現在位置をそのランドマークの周辺に絞り込むことができる。一方、同種のランドマークの数が増えるにつれ、撮影者の現在位置の候補の数も増加するため、ランドマーク検出によるマッピング精度の改善効果は徐々に小さくなっていくことが想定される。このため、ランドマーク検出による撮影者の現在位置の絞

り込みの効果を高めるためには、環境内に設置された同種のランドマークの数が少ないことが望ましいといえる。

また、一般に、歩行者密度が高くなるにつれて、モバイルカメラから最も近い人物との距離が近くなり、視野内にあるはずのランドマークが見えなくなるケースが頻繁に発生することが想定される。こうした環境において高いマッピング精度を維持するためには、LRS のポールに加えて、展示ブースの看板等、異なる種類のランドマークを併用することで、カメラ視野内にランドマークがとらえられる可能性を高めることが有効であると考えられる。

7. まとめと今後の課題

本論文では、LRS により検出した群衆の移動軌跡情報と、その群衆内の歩行者がモバイルカメラで撮影した周辺環境の視覚情報をもとに、撮影者自身およびカメラ画像中に写る周辺人物の現在位置を特定する手法を提案した。シミュレーション実験による性能評価を通じて、カメラ画像内の人物を、最大 83% の精度で、LRS で検出された歩行者位置へとマッピングできることを示した。今後は、ランドマークに依存することなくマッピング精度を高めるためのアルゴリズムの改善や、実環境における有効性の検証等に取り組んでいく予定である。

謝辞 本研究の一部は、KDDI 財団ならび文部科学省国家課題対応型研究開発推進事業—次世代 IT 基盤構築のための研究開発「社会システム・サービスの最適化のための IT 統合システムの構築」(2012 年度～2016 年度)の助成を受けたものです。お礼申し上げます。ひとたびのデータを用いた実証実験に関しては、大阪大学東野研究室修士課程の上嶋祐紀氏、藤田和久氏、ならびに高藤巧氏にお世話になりました。ここに感謝します。

参考文献

- [1] 上嶋祐紀, 藤田和久, 樋口雄大, 廣森聡仁, 山口弘純, 東野輝夫, 下條真司: ひとつながるなび—位置と気持ちと空間の共有, マルチメディア, 分散, 協調とモバイル (DICOMO2013) シンポジウム論文集 (2013).
- [2] 山口弘純, 廣森聡仁, 樋口雄大, 内山 彰, 梅津高朗, 東野輝夫, 孫 為華, 下條真司, 山口容平, 下田吉之: 共有空間の「ひと」「くうき」「きもち」を一体化する「ひとつなび」のコンセプトとビッグデータに向けた実証実験, 電子情報通信学会技術研究報告, Vol.113, No.168, pp.13-18 (2013).
- [3] Wada, Y., Higuchi, T., Yamaguchi, H. and Higashino, T.: Accurate Positioning of Mobile Phones in a Crowd using Laser Range Scanners, *Proc. WiMob '13* (2013).
- [4] 和田悠佑, 山口弘純, 東野輝夫: レーザレンジスキャナと Wi-Fi Fingerprint を併用した歩行者の位置推定手法の提案, 情報処理学会研究報告, Vol.2013-MBL-65, No.26, pp.1-7 (2013).
- [5] Bahl, P. and Padmanabhan, V.N.: RADAR: An in-building RF-based user location and tracking system, *Proc. INFOCOM '00*, pp.775-784 (2000).
- [6] Youssef, M. and Agrawala, A.: The Horus WLAN location determination system, *Proc. MobiSys '05*, pp.205-218 (2005).
- [7] Chintalapudi, K., Padmanabha Iyer, A. and Padmanabhan, V.N.: Indoor localization without the pain, *Proc. MobiCom '10*, pp.173-184 (2010).
- [8] Lim, H., Kung, L.-C., Hou, J.C. and Luo, H.: Zero-Configuration, Robust Indoor Localization: Theory and Experimentation, *Proc. INFOCOM '06*, pp.1-12 (2006).
- [9] Woodman, O. and Harle, R.: Pedestrian Localisation for Indoor Environments, *Proc. UbiComp '08*, pp.114-123 (2008).
- [10] Li, F., Zhao, C., Ding, G., Gong, J., Liu, C. and Zhao, F.: A Reliable and Accurate Indoor Localization Method Using Phone Inertial Sensors, *Proc. UbiComp '12*, pp.421-430 (2012).
- [11] Enzweiler, M. and Gavrilu, D.: Monocular pedestrian detection: survey and experiments, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.31, No.12, pp.2179-2195 (2009).
- [12] Barnich, O. and Van Droogenbroeck, M.: ViBE: A powerful random technique to estimate the background in video sequences, *Proc. ICASSP '09*, pp.945-948 (2009).
- [13] Li, L., Huang, W., Gu, I.Y.H. and Tian, Q.: Foreground object detection from videos containing complex background, *Proc. Multimedia '03*, pp.2-10 (2003).
- [14] Rother, C., Kolmogorov, V. and Blake, A.: "GrabCut": interactive foreground extraction using iterated graph cuts, *ACM Trans. Graph.*, Vol.23, No.3, pp.309-314 (2004).
- [15] Viola, P., Jones, M. and Snow, D.: Detecting pedestrians using patterns of motion and appearance, *CVPR '03*, pp.734-741 (2003).
- [16] Dalal, N. and Triggs, B.: Histograms of oriented gradients for human detection, *Proc. CVPR '05*, Vol.1, pp.886-893 (2005).
- [17] Mikolajczyk, K., Schmid, C. and Zisserman, A.: Human detection based on a probabilistic assembly of robust part detectors, *Proc. ECCV '04*, pp.69-82 (2004).
- [18] Lowe, D.G.: Distinctive image features from scale-invariant keypoints, *Int. J. Comput. Vision*, Vol.60, No.2, pp.91-110 (2004).
- [19] 日立情報通信エンジニアリング: レーザ・センシングシステム LaserRadarvisionII.
- [20] Fod, A., Howard, A. and Mataric, M.J.: Laser-based people tracking, *Proc. ICRA '02*, pp.3024-3029 (2002).
- [21] 李 在勲, 金 容植, 川田浩彦, 大矢晃久, 油田信一: 測域センサを用いたセキュリティーシステムの開発—人物追跡・計数アルゴリズム, 第 24 回日本ロボット学会学術講演会論文集 (2006).
- [22] 中村克行, 趙 卉菁, 柴崎亮介, 坂本圭司, 大鋸朋生, 鈴木尚毅: 複数のレーザレンジスキャナを用いた歩行者トラッキングとその信頼性評価, 電子情報通信学会論文誌, Vol.88, No.7, pp.1143-1152 (2005).
- [23] Zhao, H. and Shibasaki, R.: A novel system for tracking pedestrians using multiple single-row laser-range scanners, *IEEE Trans. Systems, Man, and Cybernetics, Part A*, Vol.35, No.2, pp.283-291 (2005).
- [24] Teixeira, T., Jung, D. and Savvides, A.: Tasking networked CCTV cameras and mobile phones to identify and localize multiple people, *Proc. Ubicomp '10*, pp.213-222 (2010).
- [25] Hokuyo Automatic, Co., Ltd.: UTM-30LX.
- [26] Viola, P. and Jones, M.: Rapid object detection using a boosted cascade of simple features, *Proc. CVPR '01*, Vol.1, pp.I-511-I-518 (2001).

- [27] 山内悠嗣, 山下隆義, 藤吉弘巨: [サーベイ論文] 統計的学習手法による人検出, 電子情報通信学会パターン認識・メディア研究会 (PRMU) 技術報告, pp.113-126 (2012).

推薦文

本論文では, レーザレンジスキャナにより検出した群衆の移動軌跡情報と, 群衆内の歩行者が保持するモバイルカメラから得られる周辺人物の視覚情報を融合することで, 群衆内での現在位置および周辺人物の属性情報を認識する手法を提案している. 群衆の属性情報の取得という実用的な課題を, レーザレンジスキャナとモバイルカメラを用いて解決する斬新なアイデアを実現しており, 尤度を用いた位置推定やカメラの画像処理を組み合わせる等, 技術的にも高度である. また, 評価も, シミュレーションと実機評価により十分に行われている. 以上より, 本論文は推薦に値する.

(マルチメディアと分散処理研究会主査 勝本道哲)



岩橋 宏樹

平成 24 年大阪大学基礎工学部情報科学科中退. 平成 26 年同大学大学院情報科学研究科博士前期課程修了. 現在, 日本電信電話株式会社研究員. 在学中は, モバイルカメラ画像を用いた人物位置推定に関する研究に従事.



樋口 雄大 (正会員)

平成 22 年大阪大学基礎工学部情報科学科卒業. 平成 26 年同大学大学院情報科学研究科博士後期課程修了. 同年より同大学大学院情報科学研究科特任助教. 博士 (情報科学). モバイル/パーベイシブコンピューティングに関する研究に従事. IEEE 会員.



山口 弘純 (正会員)

平成 6 年大阪大学基礎工学部情報工学科卒業. 平成 10 年同大学大学院基礎工学研究科博士後期課程修了. 同年オタワ大学客員研究員. 平成 11 年大阪大学大学院基礎工学研究科助手. 平成 14 年同大学大学院情報科学研究科助手. 平成 19 年より同大学大学院情報科学研究科准教授. 博士 (工学). モバイルコンピューティング等に関する研究に従事. IEEE, 電子情報通信学会各会員.



東野 輝夫 (フェロー)

昭和 54 年大阪大学基礎工学部情報工学科卒業. 昭和 59 年同大学大学院基礎工学研究科博士後期課程修了. 同年同大学助手. 現在, 同大学大学院情報科学研究科教授. 博士 (工学). 分散システム, 通信プロトコル, モバイルコンピューティング等の研究に従事. 電子情報通信学会, ACM 各会員. IEEE Senior Member.