

推薦論文

# 俯瞰画像における動体領域面積に基づく 群衆人数推定法の提案

小島 颯平<sup>1,a)</sup> 内山 彰<sup>1,b)</sup> 廣森 聡仁<sup>1,c)</sup> 山口 弘純<sup>1,d)</sup> 東野 輝夫<sup>1,e)</sup>

受付日 2016年3月28日, 採録日 2016年10月4日

**概要:** 本研究では, クラウドソーシングによる都市の群衆把握を目的とし, スマートフォンを有する協力者がビルの2階など比較的高所から群衆を撮影した俯瞰画像内の群衆人数を推定する手法を提案する. 画像からの群衆人数推定法として, 特徴量に基づき人物形状の検出を行う方式があるが, 群衆内の人々は身体境界が明確に現れにくいいため, 高精度に特徴量を得ることは難しい. そこで本研究では, 群衆が存在する画像内領域をまず推定し, その面積に基づく群衆人数推定法を提案する. 提案手法では, 群衆が移動していることを仮定し, 短い時間間隔で撮影された2つの画像の差分をとることで群衆が存在する動体領域を抽出する. 3次元空間シミュレータ Unity を用いてあらかじめ構築した歩行者面積モデルを構築しておき, 動体領域における1人あたりの面積を用いて高精度な群衆人数推定を実現させる. Unity で再現した群衆画像および JR 大阪駅前で撮影した画像に対してそれぞれ提案手法を適用したところ, Unity によるシミュレーションでは絶対誤差が平均 13.1%となり, 人がすべて動いている前提であれば実環境の JR 大阪駅前では 12.0%となることを確認できた.

**キーワード:** 動体領域, 歩行者面積モデル, 群衆人数推定

## Image-based Crowd Counting with Perspective Geometry Using Smartphone

SOHEI KOJIMA<sup>1,a)</sup> AKIRA UCHIYAMA<sup>1,b)</sup> AKIHITO HIROMORI<sup>1,c)</sup>  
HIROZUMI YAMAGUCHI<sup>1,d)</sup> TERUO HIGASHINO<sup>1,e)</sup>

Received: March 28, 2016, Accepted: October 4, 2016

**Abstract:** In this research, we propose a method to estimate the number of people in crowds by using a smartphone camera. Our method uses a long shot of people crowds captured by a user above the ground level. A native approach is to count people by applying image recognition algorithms based on image features of a human. Actually, such an approach cannot achieve high accuracy because the edge of a human shape is often obscured by others in crowds. Therefore, we take a different approach using detection of moving object regions. The proposed method detects areas of moving objects by taking differences of consecutive two images captured at a short interval. We regard the detected areas as people crowds assuming that people are moving. We also build a model to estimate the occupied area size per person since it changes depending on camera height, and angle, positions in an image and moving directions due to perspective geometry. The model is constructed through 3D simulations of walking pedestrians using the 3D simulator Unity. For evaluation, we have applied the proposed method to images of people crowds simulated by on Unity and captured near JR Osaka station. The result shows that the absolute error is 13.1% in simulations and 12.0% those in the real world.

**Keywords:** moving object region, pedestrian area model, crowd people counting

<sup>1</sup> 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) s-kojima@ist.osaka-u.ac.jp

b) utiyama@ist.osaka-u.ac.jp

c) hiromori@ist.osaka-u.ac.jp

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

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

### 1. はじめに

近年, 都市街区などにおける人々の存在や行動を把握し, 混雑予想や大規模イベント推定に活用する事例がみられるようになってきている. たとえばモバイル空間統計 [1] では, 携帯電話ユーザの通信統計情報を活用し, 250mメッ

シュレベルでの人口推定を行っている。また、ジオタグ付きツイートから多数の人々が訪れるイベントを検出する試みなどがなされている [2]。しかし前者は大まかな街区レベルでの存在推定であり、後者は特定のイベントやキーワードに紐づいた特定ユーザの存在しか検知できない。これに対し、イベント会場や駅前、あるいは道路などの特定のスポット空間における群衆の存在を把握することができれば、交通機関の乱れによる混雑状況や屋外イベントにおける集客状況、帰宅難民による道路占有状況などがスポットごとに把握でき、迅速な代替輸送計画立案や災害時の避難者誘導、施設活性化など様々な応用が期待される。

群衆検出には映像を用いた方式 [3] が提案されているが、混雑したスポットをとらえるカメラが設置されている保証がないことから、様々な地点で動的に形成される群衆を可視領域が限定された固定カメラのみでとらえることは現実的でない。モバイル端末を用いた群衆検出手法も提案されているが [4], [5], [6], これらのいずれの方法も数百 m 四方程度の比較的広い範囲を対象としており、前述のようなスポット空間における群衆の存在把握には不十分である。

このような課題に対し、提案手法では、群衆が存在すると推定される領域の面積に基づき群衆人数推定法を提案する。群衆が移動していることを仮定し、経験的に得られた 0.7 秒の時間間隔で撮影された 2 つの画像の差分をとることで背景を除去し、動体領域を抽出する。ここで、画像内の各人物が画像領域に占める面積が分かれば、動体領域を占める人数の推定が可能となるが、実際には、近距離の被写体は画像内で大きく、遠距離の被写体は小さくなるため、同一人物であっても画像内の位置によって 1 人あたりの占有面積は異なる。そこで提案手法では、歩行者面積モデルを構築し、カメラの高度と傾度、画像内の被写体位置と移動方向に応じた 1 人あたりの占有面積を推定することで、高精度な群衆人数推定を実現させる。モデルの構築には、3次元空間シミュレータ Unity [7] により生成された網羅的な被写体のサイズを用いる。提案手法では、文献 [8] で用いられているスマートフォンの気圧センサから得られるカメラの高度と、加速度センサから得られる傾度に対応するモデルを用いることで、動体領域の検出と組み合わせた群衆人数の推定が可能となる。

性能評価のため、Unity で群衆の動きを再現し得られた画像と、実環境における性能評価として JR 大阪駅前撮影した動画から一定間隔で抽出した画像に対してそれぞれ提案手法を適用した、その結果、Unity によるシミュレーションでは絶対誤差が 13.1%、人がすべて動いている前提であれば実環境では 12.0% となり、高精度な群衆人数検出を実現できていることが確認できた。

## 2. 関連研究

### 2.1 画像に対する人検出手法

画像内の人を検出する手法として、人の特徴量を用いた検出手法が存在する。文献 [9] では、画像の局所領域から輝度の勾配ヒストグラムを作成し (HOG 特徴量: Histogram oriented gradient), サポートベクター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。実際に HOG 特徴量を AdaBoost [10] によって学習した例を示す。HOG 特徴量に基づく人検出は図 1 に示されるように、背景や人の持ち物を人として認識するといった誤検出が多いことがあげられる。このため、人数が少ない場合には誤検出率が非常に高くなる。また、群衆の場合は周辺に人がいるため背景と人の間に存在するような身体の境界である人特徴量が明確に現れにくいと、検出に失敗しやすくなるという問題がある。次に、3.2 節に示す背景除去法によって背景を除去した画像に対して HOG 特徴量を用いた場合の検出結果の例を図 2 に示す。単純に全領域に対し HOG 特徴量を用いて検出した場合と比べ、動体領域の抽出により誤検出は減っているものの、動体領域内では図 1 に類似し



図 1 HOG 特徴量による人検出の例

Fig. 1 Example of pedestrian detection using HOG features.



図 2 背景除去と HOG 特徴量による人検出の例

Fig. 2 Example of background subtraction and pedestrian detection using HOG features.

本論文の内容は 2015 年 7 月のマルチメディア、分散、協調とモバイル (DICOMO2015) シンポジウムにて報告され、モバイルコンピューティングとパーベイシブシステム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。

た誤検出や検出漏れが存在する。これらに示されるように HOG 特徴量に基づく人検出手法を俯瞰画像における群衆人数推定に用いることは容易でないことが分かる。

Integral Channel Features [11] ではグレースケール, RGB, HSV, LUV を含む色の勾配ヒストグラムを作成することによって人の特徴量を表す。得られた特徴量は AdaBoost [10], RealBoost [12], LogitBoost [13] の Boosting アルゴリズムを用いて, ソフトカスケード型の分類器を作成する。この手法は HOG で用いられる色の形式に HSV, LUV といった色の形式を追加しているため, HOG を内包した特徴量である。文献 [14], [15] では, 2つの領域間の輝度差 (Haar-like 特徴量) をもとに画像内のエッジ成分をとらえて, AdaBoost アルゴリズムを用いてカスケード型分類器を作成する。文献 [16] では, 固定カメラから得られる背景情報を用いて移動物体が通ることのできる場所を特定し, その場所について HOG や Omega-Shape といった特徴量を用いて人検出手法であるが, スマートフォンカメラを用いた場合には, 背景情報がないため適用が困難である。文献 [17] では, SIFT [18] をベースとした特徴量を用いて, 顔や肩などの体のパーツを人体モデルに基づいて組み立てることで歩行者を検出する。この方法により, 他の人物などの障害物による遮蔽 (オクルージョン) によって体の一部しか画像中に現れない場合でも人を認識することが可能となる。しかし HOG をはじめとした特徴量に基づく人検出手法では, 群衆の場合は周辺に人がいるため背景と人の間に存在するような身体の境界である人特徴量が明確に現れにくいいため, 高精度に特徴量を得ることが容易でないという課題がある。

また文献 [19] では, 固定カメラの映像から人の写っていない背景を用いて人のみを抽出し, その人のエッジ検出を行って人数増加にともなうエッジ長の増加の線形性を用いて人数推定を行っている。しかし文献 [19] が対象としている群衆人数よりも様々な高度や角度から撮影がされるため, さらに多くの群衆を撮影されることが想定される。これに対して本研究では, 撮影高度と撮影角度が一致すれば場所によらず利用できるモデルを構築し, 文献 [19] が対象としている 30 人程度よりもさらに多くの群衆人数の推定を可能としている。

## 2.2 スマートフォンを用いた群衆密度推定

本研究と同じように監視カメラやセンサ機器といったインフラに依存しない群衆密度の推定手法が検討されている。文献 [4], [20] では, 近隣に存在するモバイル端末間で Bluetooth による無線アドホック通信を行い, 通信範囲内の端末数や受信電波強度のばらつきをもとにユーザ周辺の群衆の密度を推定している。しかし, これらの方法は Bluetooth 端末をつねに周辺のデバイスから検出可能な状態にすることを想定しており, ユーザが意図していないと

ころで不正にトラッキングされたり悪意のあるユーザから不正アクセスを受けたりするといった, セキュリティ上の問題が考えられる。GPS 情報を用いて密度推定を行う手法 [5] では, 屋外かつ広域の推定が行われており, システムへの参加者の端末の絶対数による推定を行う。文献 [6] では, 携帯電話端末が接続している基地局の情報を利用して都市間における人の移動をモデル化することで, 低コストに群衆密度を推定することができる。しかしこの方法では数百 m 四方程度の比較的広い範囲を対象としており, 本研究の目的とする空間的粒度とは大きく異なる。

## 3. 群衆人数推定法

### 3.1 概要

提案手法の基本アイデアは, 群衆を 1 人 1 人数え上げるのではなく, 群衆の大きさ, すなわち画像内において群衆が写っている面積 (占有面積) から人数を推定することである。群衆内においては人物の境界が曖昧になり特徴量による認識が難しいため, そのように群衆を 1 つのまとまりとしてとらえ, おおよその人数を推定するアプローチが有効であると考えられる。提案手法では背景画像が除去された環境を想定しているため, 短い時間間隔で撮影された 2 枚の画像に対して差分をとった領域を群衆が存在する領域 (動体領域) として検出し, その面積を群衆の占有面積と見なす。ここで, 被写体の位置およびカメラの高度と傾度によって画像中における被写体の見かけ上の大きさが変化する。さらに, 同じ被写体であってもその移動方向によって動体領域の占有面積は変化する。そして, 群衆の動体領域の面積を, 歩行者面積で除することで, 動体領域内の人数推定をする。この 1 人の歩行者面積に基づく人数推定は, 1 人あたりの想定される占有面積から動体領域内の人数を推定する。歩行者面積モデルの構築では, 適切な 1 人あたりの占有面積を決定するため, 平均的な体格の人物 (標準被写体) と移動速度ならびに平均移動速度を仮定し, 事前にあらゆる高度, 傾度, 人物の位置, 移動方向に対する占有面積をシミュレーションにより計算する。モデルには身体の大きさだけでなく, 歩行時の足や腕の動きの大きさも含まれる。シミュレーションには, 3 次元空間シミュレータ Unity [7] を用いる。歩行者面積モデルの入力に必要な被写体の移動方向は, 連続する 2 枚の画像から得られるオプティカルフローによって推定する。

以降ではモデルに必要なカメラの画角, 撮影時の画像の縦横のピクセル数, 撮影時のカメラの高度および地面に対する傾きは既知であると仮定する。Android 端末におけるカメラの画角および画像の縦横のピクセル数は API により取得可能である。また, スマートフォンに搭載された気圧センサおよび温度センサや OpenWeatherMap などで得られる撮影場所の気温を用いることで, 撮影高度を推定する。文献 [8] では, 気圧および気温から誤差 1 m 以内で高



図 3 動体領域検出の例

Fig. 3 Example of moving area detection.

度推定を行うことができるとの報告がなされている。カメラの傾きは加速度センサーで検出した重力成分から容易に推定できる。

### 3.2 動体領域の検出

動体領域の抽出には図 3 (a) に示されるような画像を用いる。まず短い時間間隔で撮影された 2 枚の入力画像に対して、5 分の 1 の解像度に縮小を行う。これはスマートフォン上でこのアルゴリズムが実行されることを想定し、計算負荷を軽減させるためである。この 2 枚の入力画像に共通する背景を除去できれば動体領域を検出できる。まず手ぶれに対する補正に必要な撮影中の手ぶれによる画像の変位は POC (位置限定相関) [21] を用いて得られる。2 枚目の画像を POC で得られた変位の分だけ平行移動し、1 枚目の画像と重ね合わせる。

次に背景除去を行うため、画像差分計算および二値化処理を行う。画像差分計算では、このようにして得られた平行移動後の 2 枚目の入力画像と 1 枚目の入力画像の各ピクセルごとの画素値の差分の絶対値をとる。なお、手ぶれ補正したことによって生じる 2 枚の画像が重ならない領域は以降では無視する。これらの操作によって得られた画像が図 3 (b) である。

続いて二値化処理によって、動体領域とそれ以外の領域を白と黒で表し動体領域を抽出する。二値化処理とは各ピクセルの画素値を閾値に応じて、白と黒に分類する処理である。図 3 (b) では動体領域の部分は画素値が高く、2 つの画像の輝度差のために除去しきれなかった背景が低い画素値で存在している。差分画像である図 3 (b) に対してガウシアンフィルタをかけ、二値化処理を施すことによって動体領域である部分は白、そうでない部分を黒とすることができる。二値化をするにあたり、事前にガウシアンフィルタをかけ平滑化することで、ノイズを除去する。提案手法では判別分析法 (大津の二値化) [22] を適用することで、統計的な分布から画素値の高い部分と、低い部分を分割する最も適切な閾値を決定する。これにより、除去しきれなかった背景と動体領域を分ける閾値が得られる。図 3 (b) に対してガウシアンフィルタを適用し、判別分析法を適用



図 4 抽出された動体領域

Fig. 4 Extracted moving area.

した結果が図 3 (c) である。なお適用したガウシアンフィルタのサイズは  $25 \times 25$  ピクセルで、シグマの値は  $x$  方向、 $y$  方向ともに 4.1 である。また入力画像に対して適切に移動領域が検出できたかを確認するため、図 3 (c) と入力画像 1 を重畳した画像が図 4 である。動体領域は人以外の領域をなるべく含まない一方で、人である部分をできるだけ多く包含していることが望ましく、これらはパラメータにより調整可能なトレードオフの関係にある。適切に動体領域を推定するため、提案手法では目視により適切に動体領域を抽出できるようなパラメータ設定を決定した。図 4 は提案手法のパラメータ設定により抽出された動体領域と元画像を重畳した例である。

### 3.3 歩行者面積モデル

歩行者面積モデルでは、カメラの高度と傾度、被写体の移動方向、ならびに画像内の位置を入力として 1 人の占有面積を推定する。十分に短い時間において、歩行者の移動速度に大きな差異はないと考えられるため、それは一定であると見なし、移動はほぼ直進しかしないものとする。これらの各組合せにおいてシミュレーションを実施し、そのときの動体領域面積を実測することでモデルを構築するが、以下で説明する幾何的な特性により 0 度から 90 度の範囲のみを対象にモデルを構築すればよい。なお、本研究における 0 度の方向は人が画像の上方に移動する方向を 0 度とする。また画像の位置の原点は左上を原点とし、右方向を  $x$  軸、下方向を  $y$  軸とする。

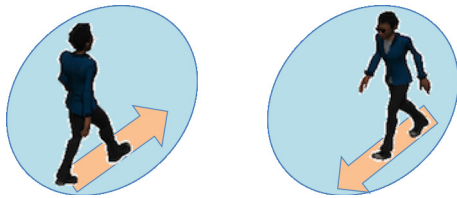


図 5 移動方向が 180 度異なる場合の動体領域  
Fig. 5 Moving area with 180° rotated vectors.

図 5 に示すような、移動方向が 180 度異なる 2 人の歩行者を考える。図 5 の左右の人の足下の座標は異なるが、右の人の移動方向を 180 度回転させ、足下の座標を左の人の座標と同じ地点にすることで、左と右の人の動きは同じものであるととらえることができるため、移動方向が 180 度異なる 2 人の動体領域面積は同じであると見なすことができる。また画像の  $y$  軸方向の位置が同じで、移動方向が左右に対称である人は、カメラからの距離がほぼ同一であると見なすことができるため被写体の大きさは同一となる。したがって、これらの理由から画像内の対称な移動方向の動体領域面積は等しい。このため移動方向は 360 度存在するが、0 度から 90 度のみの範囲を対象にモデルを構築すれば、すべての移動方向に対して歩行者面積推定が実現できる。

先に述べたように、画像の対称な移動方向の動体領域面積は等しく、 $x$  軸方向の位置には依存しない。よってモデルの作成には、足下の座標  $y$  と移動方向  $\theta$  ごとに得られる動体領域の面積に対して線形回帰を適用する。動体領域は 3.2 節で述べた方法に従って抽出する。得られた標準被写体の位置および移動方向と面積に対して線形回帰を適用することで歩行者面積モデルを構築する。歩行者面積モデルを利用する際には、画像の垂直方向を  $y$  軸としたときに移動方向が  $y$  軸となす角のうち小さい角  $\theta$  を算出し、歩行者面積モデルに与える。移動方向に関して、 $[0, 90]$  度の範囲であらゆる移動方向についてモデルを構築することが望ましいが、本研究では簡単のため作成するモデルの移動方向を 0 度、30 度、60 度、90 度の 4 つとし、移動方向のうち最も近い角度を与える。撮影高度および撮影傾度を  $h_c$  および  $\theta_c$ 、また移動方向を  $\theta$ 、人の座標として足下の位置を  $y$  とする。また本研究の評価のために作成された移動方向が 4 種類、撮影高度が 4 種類および撮影傾度が 6 種類の組合せである計 96 種類のモデルから、高度、傾度、および移動方向を入力として最もこれらの組合せに値が近い 1 つのモデルを選択する関数を  $m(\theta_c, h_c, \theta)$  とする。ここで、建物の高所から撮影することを想定しフロア間の高さを 3.5m とする。高度については、歩行者がある程度の大きさで撮影可能な 1 階から 5 階までを想定し、3.5m から 14m まで 3.5m 刻みで 4 種類の高度についてモデルを用意する。傾度については、事前実験によって本研究が想定するような比較的高い場所から群衆全体をとらえられるように撮影を行った結



図 6 0 度のときの動体領域

Fig. 6 Moving object area of 0° moving direction.

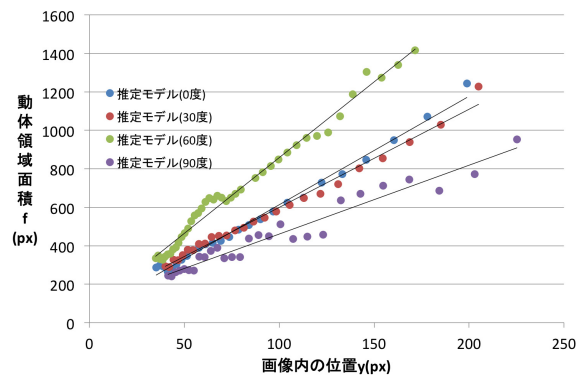


図 7 歩行者面積モデルの例 ( $h_c = 7.5$  m,  $\theta_c = 25$  度)

Fig. 7 Example of moving area models ( $h_c = 7.5$  m,  $\theta_c = 25$  degrees).

果、20 度から 45 度の範囲となった。そこで 20 度から 45 度の間で簡単のため 5 度間隔の 6 種類でモデルを用意する。 $m(\theta_c, h_c, \theta)$  により、用意された 96 種類のモデルから算出されたオプティカルフローおよびセンサによって得られた高度と傾度に最も値が近いモデルを選択する。これらによって得られる歩行者面積モデルを  $f(y, m(\theta_c, h_c, \theta))$  とすると次のように表される。 $a(h_c, \theta_c, \theta)$ ,  $b(h_c, \theta_c, \theta)$  は係数である。

$$f(y, m(\theta_c, h_c, \theta)) = a(h_c, \theta_c, \theta)y + b(h_c, \theta_c, \theta) \quad (1)$$

例として、以下に高度 7.5 m、傾度 25 度の歩行者面積モデルを示す。図 6 は Unity 上で 0 度の方向に移動する標準被写体を撮影した画像と抽出された動体領域を重畳した画像である。図 6 のように 1 人の歩行者を様々な移動方向に移動させ、各移動方向について足下の座標  $y$  と動体領域面積をプロットしたものが図 7 である。それぞれの移動方向における歩行者面積を推定する関数は次のようになった。

$$f(y, m(25, 7.5, 0)) = 5.6427y + 48.8519 \quad (2)$$

$$f(y, m(25, 7.5, 30)) = 5.1063y + 88.604 \quad (3)$$

$$f(y, m(25, 7.5, 60)) = 7.8796y + 70.218 \quad (4)$$

$$f(y, m(25, 7.5, 90)) = 3.5816y + 102.52 \quad (5)$$

様々な高度、傾度、移動方向ごとに上記のような一次関

数の式があり,  $m(\theta_c, h_c, \theta)$  によっていずれか1つの式が選択される. さらに歩行者の座標  $y$  を入力することで歩行者面積を得ることができる.

### 3.4 歩行者面積モデルを用いた群衆人数推定

動体領域と背景領域を表す二値化画像と, 3.3 節で導出した歩行者面積モデルを用いて群衆人数推定を行う. 歩行者面積モデルに必要な被写体の移動方向の方向はオプティカルフローにより推定する. これにより, 抽出された動体領域に対して歩行者面積モデルを適用することができる. 群衆人数の推定は図 8 に示す青い四角形で囲われた動体領域ごとに行う. この1つ1つの独立した動体領域を群衆ブロックと呼ぶ. この群衆ブロックは, 二値化画像に対してラベリング処理を施し, 同一ラベルの動体領域ごとのバウンディングボックスを算出することで得られる.

各群衆ブロック  $c$  のどの部分に人が存在するか分からないので,  $c$  の重心  $(g_x, g_y)$  に足元があるものとする.  $c$  の移動方向としては  $c$  内で検出されたオプティカルフローの平均方向を採用する. オプティカルフローの検出には, ORB 特徴量による特徴点抽出と, Lucas-Kanade 法 [23] による各特徴点の軌跡推定を用いる. 得られたオプティカルフローを0度から90度の範囲の同一面積となるベクトルに変換し, それらの平均をとる. 図 9 に得られたオプティカルフローの例を示す.

平均化されたオプティカルフローのベクトルを  $v_c$  とすると,  $v_c$  は群衆ブロック内の平均的な移動方向を表していると考えられるため, これに対応する歩行者面積モデルを用いて群衆ブロック  $c$  内の人数を推定する. すなわち,  $c$  内の人数  $N_c$  は以下の式で表される.

$$N_b = \frac{S_c}{f(g_y, m(\theta_c, h_c, \Theta(v_c)))} \quad (6)$$

ここで  $S_c$  は群衆ブロック  $c$  の面積,  $\Theta(v_c)$  は移動方向  $v_c$  の移動方向とする.

こうして得られた重心位置における人の面積によって群衆ブロックの面積を除することで1つの群衆ブロックの人数を推定することができる. 最終的に, それらの総和をとることで画像内の人数を推定する.

事前実験として, Unity を用いて群衆の動きを再現し, カメラ位置を任意の高度, 傾度に設定して得られる映像を経験的に得られた0.7秒の間隔で画像を撮影し擬似的に写真の撮影を行い, 3.4 節の手法を適用したところ図 10 の結果が得られた. これはモデルの構築時の人の位置は足の位置となっているが, 群衆ブロック内の平均的な人の基準位置を動体領域の重心としているため, 適切な位置が選択されていないことおよび, 人同士の重なりによる誤差である. 図 10 のプロットしている値について横軸を推定人数, 縦軸を実人数とした図が 図 11 であり, 補正前の歩行者面積モデルによって推定された人数と実人数の関係を表してい



図 8 群衆ブロック検出例

Fig. 8 Example of crowd block detection.



図 9 Unity の実験画像を用いたオプティカルフローの検出例

Fig. 9 Example of optical flow detection in Unity simulation.

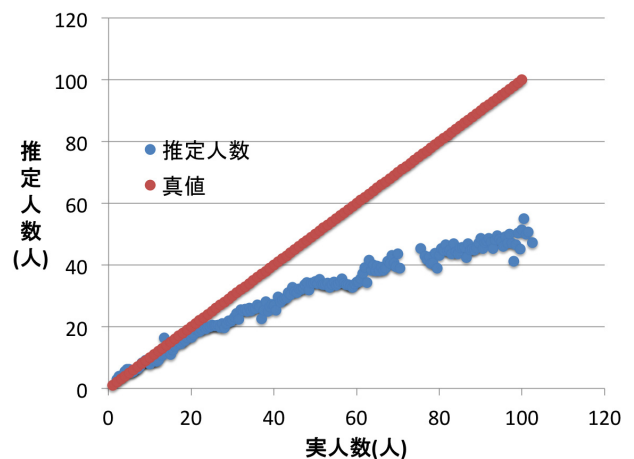


図 10 実人数に対する推定人数

Fig. 10 Actual # of people vs. estimated # of people.

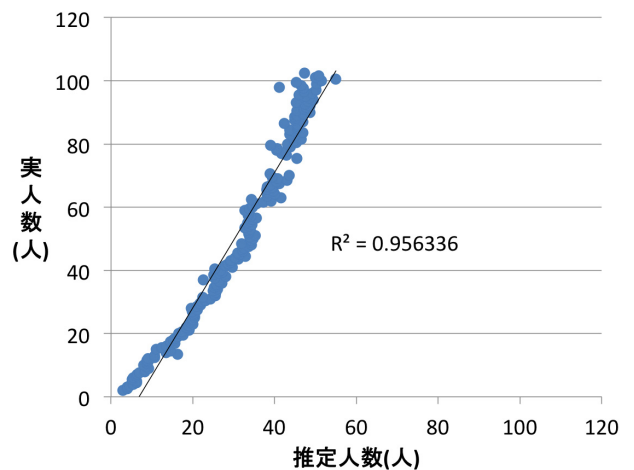


図 11 推定人数に対する実人数

Fig. 11 Estimated # of people vs. actual # of people.

る。図 11 中の  $R^2$  値は、線形回帰を行った式の決定係数を表す。実人数の増加にともないオクルージョンの影響が大きくなるため、歩行者面積モデルによって推定された人数は実人数よりも少なくなるが、実人数が増加すれば推定人数も線形的に増加する傾向は保たれている。したがって、歩行者面積モデルにより算出された推定人数に対して、適切な補正式を決定することによって、推定精度の向上を図る。この補正式は、事前にシミュレーションによって様々な実人数に対する歩行者面積モデルによる推定人数のデータを集め、推定人数を説明変数、実人数を目的変数とした線形回帰によって決定する。図 11 はシミュレーションによって撮影された画像を用いて求めたもので、人数ごとに異なるオクルージョンの起こりやすさに応じて歩行者面積モデルで推定された人数の補正を実現している。

#### 4. 性能評価

##### 4.1 実験環境

提案手法の性能評価を行うため、Unity を用いて群衆の動きを再現し、そのカメラから見られる映像を経験的に得られた 0.7 秒の間隔で画像を記録することで擬似的に写真の撮影を行った。評価では補正モデルをシミュレーションにより構築し、補正モデルの構築に使った結果とは異なるデータを用いてシミュレーションに対する評価を行う。また実環境における評価として、JR 大阪駅前の交差点において歩行者信号が青の期間を対象に Android 端末 Nexus 5 を用いて撮影した動画から 0.7 秒間隔で画像を抽出することで入力画像を得た。各入力画像について目視により人の頭部の数を数えることで真値を得た。提案手法の実装は Java および画像処理ライブラリ OpenCV2.4.10 を用い、Mac OS10.9.4 上で動作させた。実環境においても、シミュレーションにより構築した補正モデルを適用した。

##### 4.2 撮影高度および角度の影響

提案手法の撮影高度と撮影角度による性能の変化を示すため、ビルの 2 (3.5 m), 3 (7.0 m), 4 (10.5 m), 5 (14.0 m) 階において、20, 25, 30, 35, 40, 45 度を想定し、歩行者面積モデルを作成しシミュレーション結果を用いて性能評価を行った。シミュレーションでは 40 m 四方の平面の上を 0 人から 200 人の歩行者がランダムに歩行する。歩行者はランダムに選択した方向に 1 [m/s] の速度で直進し、それぞれの歩行者は簡単のため人同士の衝突や人が近くにいることによる速度の減少などは行わずつねに一定の速度で歩行する。カメラの画角の設定は実環境の実験で用いた Android 端末と同じ設定にするため、横の画角は 60°, 縦の画角は 45°, 画素は 1,920 × 1,080 の設定で行った。

表 1 に示すとおり、高度 3.5 m の場合を除く環境について表 1 の誤差の平均をとったところ 13.1% となり、撮影角度および高度ごとにモデルを作成することで、高精度な人

表 1 異なる角度と高度における人数推定の絶対誤差 (%)  
Table 1 Absolute error (%) in different camera height and angle.

傾度 \ 高度	3.5 m	7.0 m	10.5 m	14.0 m
20 度	59.5	16.2	11.8	10.0
25 度	80.8	11.8	10.3	9.79
30 度	-	11.7	9.85	8.39
35 度	-	13.0	8.69	8.45
40 度	-	29.8	7.84	9.30
45 度	-	43.3	18.2	9.41

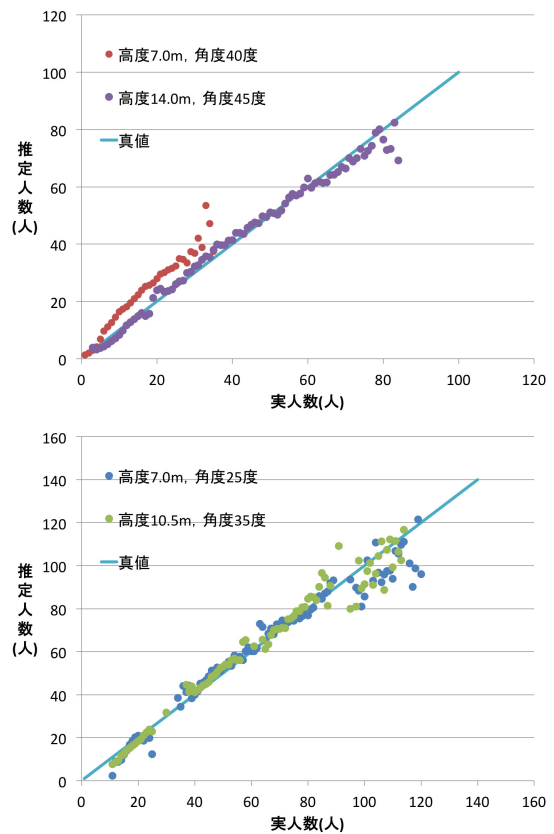


図 12 高度と傾度の影響

Fig. 12 Effect of camera altitude and tilt.

数推定が実現できた。Yahoo!地図アプリ [24] では混雑度を混雑指数という 5 段階の分類でヒートマップとして表現している。本研究は特定のスポット空間における混雑状況のマッピングを目的としており、こういった利用を想定した場合に十分な精度が得られていると考えている。表 1 を見ると、高度が 7.0 m、傾度が 40 度のとき誤差が 29.8%と大きくなっている。これはこの撮影高度で角度が大きい場合、他の場合に比べて撮影時に写る人数が少なくなるため、割合で見たときの実人数に対する推定人数の誤差が大きくなってしまったためである。高度が 7.0 m、傾度が 45 度の場合も同様である。図 12 に示すように、高度が 7.0 m、角度が 40 度のときも他のパターンの推定結果同様に、真値に近い推定結果となっていることが確認できる。また同様に

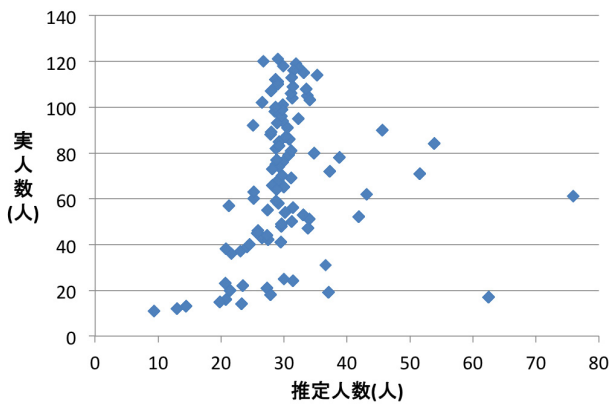


図 13 撮影高度 3.0m における補正例  
Fig. 13 Correction example in 3.0m camera height.



図 14 入力画像と動体領域を重畳した図  
Fig. 14 Overlapped image of input and detected moving object areas.

図 12 において異なる撮影高度や角度を比較した場合、異なるパターンにおいても実人数が同じ部分はほぼ同じ推定値であり、高度と傾度に応じてモデルを選択することの有効性が確認できる。2階 (3.5m) から撮影した場合、表 1 に示すとおり誤差がきわめて大きくなっている。なお、高度 3.5m の角度が 30 度以上の場合、さらに誤差が大きくなっていくため性能評価を行っていない。高度 3.5m のときの歩行者面積モデルを構築する際の、図 10 と同様の補正を行う線形回帰を行った結果を図 13 に示す。図 13 に示すように図 10 に見られるような線形的な傾向が確認できなかった。これは図 14 に示すように、オクルージョンの影響が補正の限界を超えてしまい、図 13 に示すように実人数が増加するにつれて動体領域面積が増加する傾向をとらえることが難しくなったためである。この結果から、3階以上の高所においてはオクルージョンが存在する場合であっても、実人数の増加にともなう動体領域面積の増加傾向をとらえることができ、提案手法が有効であることが確認できた。

#### 4.3 実環境での評価

実環境の評価環境として横断歩道の歩行者を撮影した画像を入力とした。横断歩道では人の移動する範囲が限定さ

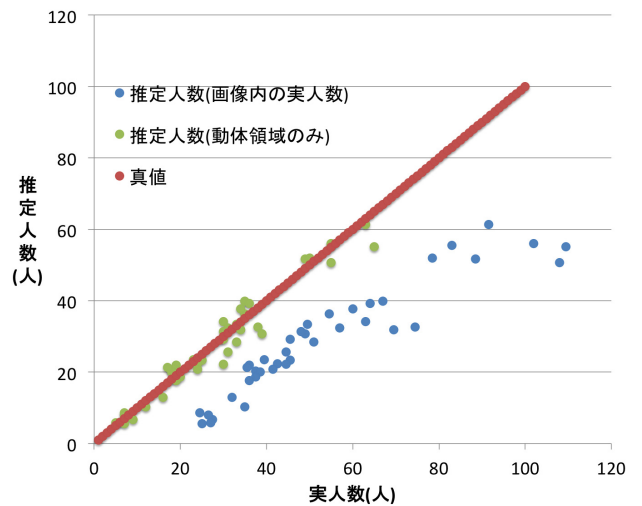


図 15 実環境における実人数に対する推定人数  
Fig. 15 Result of crowded counting in JR Osaka.



図 16 実環境における動体領域抽出例  
Fig. 16 Example of detected moving areas in JR Osaka.

れており、それぞれの歩行者の歩行速度もほぼ一定であることからオクルージョンが発生しやすい環境であると考えられる。このため、提案手法の精度低下の大きな要因であるオクルージョンが発生しやすい横断歩道の環境で評価することにより、提案手法の有効性を示す。

実環境における性能変化の結果を図 15 に示す。画像内の実人数に対する絶対誤差の平均は 48.0% となった。図 15 はシミュレーション時の結果と比較して、実人数よりも少ない人数として推定されていることが分かる。これは図 16 に示されるように、本来人が存在しており動体領域として抽出されるべき部分において人の動きが小さかったため、うまく検出できていなかったことによるものである。そこで抽出された動体領域に含まれる人のみを対象として群衆人数の真値を定義し、評価した結果を図 15 の推定人数 (動体領域のみ) として示す。この場合の推定人数の絶対誤差は 12.0% となり、群衆内の人が全員動いているという前提であれば、提案手法により高精度な人数推定が実現できることが分かった。

#### 5. おわりに

提案手法では、短い時間間隔で撮影された 2 つの群衆俯瞰画像の差分をとることで群衆が存在すると推測される動



体領域を抽出し、その面積に基づき群衆人数推定を行う。1人あたりの占有面積は、画像内の位置や移動方向によって異なるため、カメラの高度と傾度、画像内における被写体の位置と移動方向のパターンを網羅した歩行者面積モデルを構築することで群衆人数推定を実現する。さらに、高密度な場合に発生する人同士の重なりを線形回帰で補正することで、精度の向上を図る。

性能評価のため、Unityで再現した群衆画像およびJR大阪駅前で撮影した画像に対してそれぞれ提案手法を適用したところ、Unityによるシミュレーションでは評価環境において絶対誤差が平均13.1%、人がすべて動いている前提であれば実環境のJR大阪駅前では12.0%となり、高精度な群衆人数推定を実現できた。

今後の課題として、実環境において様々なシチュエーションに対して性能評価を行い幅広い環境での提案手法の有用性を示すことがあげられる。

謝辞 本研究の一部はJSPS科研費JP26220001, JP15H02690, JP26700006の助成を受けたものです。

#### 参考文献

- [1] 寺田雅之, 永田智大, 小林基成: モバイル空間統計における人口推計技術 (社会・産業の発展を支える「モバイル空間統計」: モバイルネットワークの統計情報に基づく人口推計技術とその活用), NTT DoCoMo テクニカル・ジャーナル, Vol.20, No.3, pp.11-16 (2012).
- [2] 風間一洋, 鳥海不二夫, 榊 剛史, 篠田孝祐, 栗原 聡, 野田五十樹: 東日本大震災時のTwitterデータを用いた単語間の関係の時系列変化の分析, 人工知能学会全国大会 (第26回) 論文集 CD-ROM, Vol.26, pp.1-4 (2012).
- [3] 宮崎真次, 宮野博義, 池田浩雄: 群衆行動解析技術を用いた混雑推定システム (社会の安全・安心を支えるパブリックソリューション特集)-(安全・安心な暮らし), NEC 技報, Vol.67, No.1, pp.82-85 (2014).
- [4] Weppner, J. and Lukowicz, P.: Collaborative Crowd Density Estimation with Mobile Phones, *Proc. 9th ACM Conference on Embedded Networked Sensor Systems*, pp.193-200 (2011).
- [5] Reades, J., Calabrese, F., Sevtsuk, A. and Ratti, C.: Cellular Census: Explorations in Urban Data Collection, *IEEE Pervasive Computing*, Vol.6, No.3, pp.30-38 (2007).
- [6] Isaacman, S., Becker, R., Cáceres, R., Martonosi, M., Rowland, J., Varshavsky, A. and Willinger, W.: Human Mobility Modeling at Metropolitan Scales, *Proc. 10th International Conference on Mobile Systems, Applications and Services (MobiSys '12)*, pp.239-252 (2012).
- [7] Unity-Technologies: Unity-Game Engine, available from (<http://japan.unity3d.com/>).
- [8] 岩波慶一郎, 新井イスマイル: スマートフォン搭載気圧センサを用いた移動経路推定手法における気圧センサ値の評価と補正手法の検討, 情報処理学会, マルチメディア, 分散, 協調とモバイル (DICOMO2014) シンポジウム論文集, Vol.2014, pp.1620-1626 (2014).
- [9] Dalal, N. and Triggs, B.: Histograms of oriented gradients for human detection, *Proc. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, pp.886-893 (2005).
- [10] Viola, P. and Jones, M.: Robust real-time object detec-

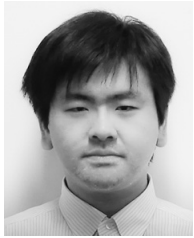
- tion, *International Journal of Computer Vision*, Vol.4, pp.51-52 (2001).
- [11] Dollar, P., Tu, Z., Perona, P. and Sivic, B.: Integral Channel Features, *Proc. British Machine Vision Conference*, pp.91.1-91.11 (2009).
- [12] Freund, Y. and Schapire, R.E.: A decision-theoretic generalization of on-line learning and an application to boosting, *Proc. Computational Learning Theory*, pp.23-37 (1995).
- [13] Friedman, J., Hastie, T., Tibshirani, R., et al.: Additive logistic regression: A statistical view of boosting (with discussion and a rejoinder by the authors), *The Annals of Statistics*, Vol.28, No.2, pp.337-407 (2000).
- [14] Viola, P., Jones, M.J. and Snow, D.: Detecting pedestrians using patterns of motion and appearance, *Proc. 9th IEEE International Conference on Computer Vision (ICCV '03)*, pp.734-741 (2003).
- [15] Corvee, E., Bremond, F., et al.: Haar like and LBP based features for face, head and people detection in video sequences, *Proc. International Workshop on Behaviour Analysis and Video Understanding (ICVS 2011)*, pp.10-20 (2011).
- [16] Li, M., Zhang, Z., Huang, K. and Tan, T.: Estimating the number of people in crowded scenes by mid based foreground segmentation and head-shoulder detection, *Proc. 19th International Conference on Pattern Recognition, ICPR 2008*, pp.1-4, IEEE (2008).
- [17] Mikolajczyk, K., Schmid, C. and Zisserman, A.: Human detection based on a probabilistic assembly of robust part detectors, *Proc. 8th European Conference on Computer Vision (ECCV '04)*, pp.69-82 (2004).
- [18] Lowe, D.G.: Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision*, Vol.60, No.2, pp.91-110 (2004).
- [19] Davies, A.C., Yin, J.H., Velastin, S., et al.: Crowd monitoring using image processing, *Electronics & Communication Engineering Journal*, Vol.7, No.1, pp.37-47 (1995).
- [20] Ghose, A., Bhaumik, C. and Chakravarty, T.: BlueEye: A System for Proximity Detection Using Bluetooth on Mobile Phones, *Proc. 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication*, pp.1135-1142 (2013).
- [21] Kuglin, C.: The phase correlation image alignment method, *Proc. International Conference Cybernetics and Society*, pp.163-165 (1975).
- [22] 大津展之: 濃度分布からの閾値決定法 (1977).
- [23] Lucas, B.D., Kanade, T., et al.: An iterative image registration technique with an application to stereo vision, *IJCAI*, pp.674-679 (1981).
- [24] ヤフー株式会社: Yahoo!地図アプリ, 入手先 (<http://promo.map.yahoo.co.jp/map/>).

#### 推薦文

本論文で提案する「俯瞰画像からの群衆人数推定モデル」は、交通機関の輸送計画立案、施設活性化施策など社会問題解決への応用が高く期待される。また、群衆の俯瞰画像から1人1人を検出し数えることは難しいという課題に対し、人々の動線方向から複数の動体領域を推定し、動体領域の面積と人数の関係から全体の人数を求めることにより解決しており、動体領域に着目した点に新規性を有する。

以上のことから本論文を推薦する。

(モバイルコンピューティングとパーベシブシステム  
研究会主査 稲村 浩)



小島 颯平 (学生会員)

平成 27 年大阪大学基礎工学部情報科学研究科卒業。同年同大学大学院情報科学研究科博士前期課程進学。スマートフォンを用いた群衆推定技術に関する研究に関心。



内山 彰 (正会員)

平成 20 年大阪大学大学院情報科学研究科博士後期課程修了。同年イリノイ大学客員研究員。平成 21 年大阪大学大学院情報科学研究科特任助教。平成 25 年同大学院情報科学研究科助教。博士 (情報科学)。モバイルセンシングや状況認識、ヘルスケアに関する研究に従事。電子情報通信学会, IEEE 各会員。



廣森 聡仁 (正会員)

平成 16 年大阪大学大学院基礎工学研究科博士後期課程修了。平成 17 年株式会社エヌ・ティ・ティ・ドコモ入社。平成 20 年大阪大学大学院情報科学研究科助教。平成 26 年同大学院情報科学研究科講師。平成 28 年より同大学院情報科学研究科准教授。博士 (工学)。モバイルアプリケーションやモバイルネットワークの設計および性能評価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 年同大学院情報科学研究科助手。平成 19 年より同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。電子情報通信学会, IEEE 各会員。



東野 輝夫 (正会員)

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, 本会フェロー。