

スマートフォン画像を用いたクラウドソーシングによる 群衆人数推定システム

小島 颯平¹ 内山 彰¹ 廣森 聡仁¹ 山口 弘純¹ 東野 輝夫¹

概要：本研究では、交差点や歩道、公共広場のような都市空間におけるスマートフォンの画像を用いた群衆人数推定手法を提案する。提案手法では、比較的高所な建物や歩道橋からスマートフォンを用いて対象領域を撮影し、群衆人数推定を行う。推定結果をクラウドサーバに送信し、様々な局所領域における混雑度マップを構築する。群衆内において画像処理により人を個別に検出するのは人特徴量が明確に表れないため難しい。そこで本研究では経験的に得られた 0.7 秒という短い時間間隔で撮影された 2 枚の画像から背景差分法により動体領域を抽出する。動体領域面積、撮影高度、撮影角度を入力としてあらかじめ構築したモデルを用いて動体領域内に存在し得る最大人数と最小人数ならびに動体領域に存在する群衆人数の最尤値を推定する。100 人以上の群衆を含む様々なシチュエーションにおいて大阪駅周辺で実験したところ、平均人数誤差が 11.3 人となった。

A Crowd Detection System Based on Crowd Sourcing by Using Smartphone Images

Sohei Kojima¹ Akira Uchiyama¹ Akihito Hiromori¹ Hirozumi Yamaguchi¹ Teruo Higashino¹

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、大規模イベント推定、都市計画、災害時の救援に役立てる事例が重要視されている。モバイル空間統計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。ソーシャルメディアもこういった目的で利用することができ、例えばジオタグ付きツイートを解析することで多数の人々が訪れるイベントを検出する試みがなされている [1]。しかし前者は大まかな街区レベルでの存在推定であり、後者はソーシャルメディアによる位置情報やユーザが発信する内容には偏りがあるため [2]、十分な数のサンプルが集まらないという問題が存在する。

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] が多数提案されているが、混雑したスポットを捉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捉えることは現実的でない。固定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推定法として、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推定を行う方法がある [4]。この方法は、Bluetooth 端末を常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るといった、セキュリティ上の問題が考えられる。その他にも同じような情報源として、協力してくれるユーザの GPS 軌跡 [5]、Wi-Fi ビーコン検出 [6]、携帯基地局を用いた方法 [7] が存在するが、これらの精度

¹ 大阪大学大学院 情報科学研究科
Graduate School of Information Science and Technology, Osaka University

はエリア内のシステム参加者数に依存する。

このような課題に対し、本研究では交差点や歩道、公共広場のような都市空間におけるスマートフォンの画像を用いた群衆人数推定手法を提案する。本手法をスマートフォンアプリケーションに実装し、それらの空間を捉えることができるような建物や歩道橋にいるユーザによってクラウドソーシングで群衆情報を取得する。群衆の把握や人流を把握するために動画を撮影することは有効であるが、動画を撮影するためにはその場所でしばらくとどまる必要があり、ユーザの協力を得る事は難しい。加えて、動画を撮影し処理することは、スマートフォンの多くのリソースを必要とする。そこで、我々は経験的に得られた0.7秒の短い時間間隔で撮影された2枚の画像を利用し、群衆人数の推定を行う。既存手法としてHOG特徴量を用いて個別に人を検出する取り組みがなされているが[8]、群衆内の人の特徴量が表れにくい本本研究が目的とする群衆に対して適用するのは難しい。そこで本研究では短い時間間隔で撮影された2枚の画像から背景差分法[9]により動体領域を抽出する。文献[10]において、あらかじめ用意したモデルから群衆人数推定、およびさらに精度をあげるために補正モデルも用意し推定を行う手法の提案を行った。文献[10]における補正モデルはランダムに人が歩くシミュレーションを元に構築したが、実環境においては横断歩道、ショッピングモールなどシチュエーションによってモビリティが異なるため人がランダムに歩いたシミュレーションを用いた補正は適切ではないことが分かった。さらに群衆内の人の特徴量を捉えることが難しいことから、オプティカルフローのような特徴量を捉えてモビリティを解析することも難しいため、様々なモビリティを用意しシチュエーションに応じた補正モデルを適用することも難しいことが分かった。そこで、動体領域の情報、スマートフォンの気圧センサおよび加速度センサによって取得できる撮影角度と撮影高度を入力とし、動体領域内に存在し得る最大人数と最小人数を算出する。さらに人のフローに依存しないモデルを構築することで最大人数と最小人数を用いて動体領域内の群衆人数の最尤値を推定するロバストな手法を提案する。本手法はスマートフォン上で実行できる十分軽量の手法であり、スマートフォン上で人数推定することでサーバ上の計算量や通信のトラフィックを減少させることが可能である。

本手法の評価のため、スマートフォン上で動作するアプリケーションを実装した。実験にはUnityを使用して人の歩行を再現したシミュレーションと実環境の評価のために大阪市において撮影された画像を用いた。シミュレーション上においては平均誤差が11.1%となり、実環境においては100人以上の人数を含む画像に対して性能評価を行ったところ平均人数誤差が11.3人となった。これにより本手法の有効性が確認することができた。



図 1: HOG 特徴量を用いて人検出した例

2. 関連研究

群衆を検出する手法として、人特徴量に基づき画像内の人を検出することが考えられる。文献[11]はHOG特徴量[8]や色情報など複数の画像チャンネルを用いた手法が提案されている。文献[12]ではHaar wavelet特徴が画像内の人のエッジを様々なパターンで表すことが可能であることを示している。さらに文献[13]では、人の形状だけでなく体の部位についてもHaar wavelet特徴を適用する方法が示されている。また同じように体の部位の特徴を表すことが可能であるSIFT特徴量を用いて、それらのパーツを組み合わせることで人を検出する方法が存在する[14]。しかし、群衆内的人是は部分的にしか特徴量を得られないため、これらの手法は群衆内の人に対して適用することは難しい。図1にHOG特徴量をAdaBoostによって学習し、群衆内の人を検出した例を示す。例を見ると、群衆内の人の特徴が明確に表れにくい、正しく人が検出されていないことが分かる。

人検出以外の方法として、時間的に連続する複数画像に対して人をトラッキングすることによって群衆人数を求める方法が存在する[15–17]。文献[16]は連続した画像からエッジ特徴とテクスチャ特徴を用いて群衆内の人数を推定する方法である。しかし、これらの特徴は連続した画像間の類似性に基づく特徴量であり、高密度な群衆の場合に適用することは難しい。

他に画像処理を用いた手法として、1つまたは複数のカメラを用いて人の頭部をトラッキングする方法がある[18–20]。このような映像を用いたトラッキング技術は、撮影エリアを通過した歩行者の数や軌跡を正確に検出するのに有効である。しかし、カメラの撮影できる範囲は限られているため、十分な性能を得るためには広いエリアを撮影できるようにあらかじめ多くのカメラを配置する必要がある。よって、ショッピングモールのような広いエリアに存在する群衆の検出には適さない。

近年、LRS(レーザーレンジスキャナ)を用いた群衆トラッキングも注目を集めている[21]。LRSは270°の扇形

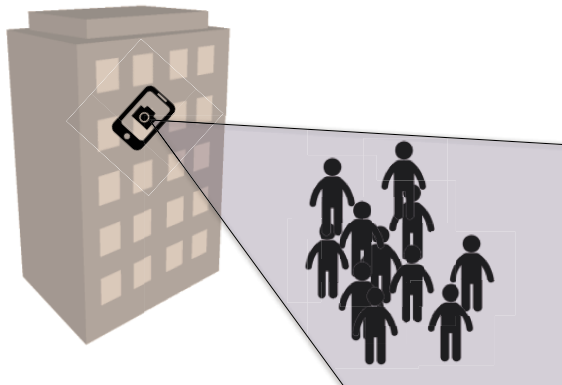


図 2: システム概要

上で半径 30m をスキャンすることができ [22], ToF(time of flight) 法によって周辺の物体の正確な距離測定が可能である。先行研究により複数の LRS を用いて歩行者の足首や腰を捉える事ことで正確な歩行者の軌跡を得られることが示されているが、高密度な群衆を想定した場合には歩行者同士で遮蔽が発生し LRS による測定を遮るため、高い精度で群衆を検出することは難しい。

より費用対効果の高い群衆密度推定手法として、スマートフォンを用いる方法がある。文献 [5] [23] [7] では、携帯ネットワークのトラフィックを利用して人のモビリティを推定し可視化する方法を示している。しかし、これらは街区レベルの群衆や人検出を目的としている。

文献 [24] では、近隣の携帯電話同士でオーディオピーコンを交換することで群衆内の携帯利用者数を推定する方法が述べられている。オーディオピーコンは MAC アドレスのような端末を識別できる要素を含んでおり、これによって近隣の端末を識別する。近隣の端末間でピーコンの交換を繰り返すことで、携帯端末は群衆内に存在する端末数を認識することができる。しかし、この手法はスピーカーとマイクを占有し、システムに参加しているユーザのみの識別しか行うことができない。文献 [4] は、Bluetooth により狭い範囲のアドホックネットワークを形成し、群衆密度センシングをモバイル端末で行う方法を提案している。しかし、Bluetooth が利用できる端末の割合によって群衆人数の推定精度が変化してしまう。

3. 画像処理による人数推定

3.1 システム概要

本システムでは、歩道橋や図 2 のような 2 階以上の比較的高所から群衆の俯瞰画像をスマートフォンで撮影することを想定している。システムの利用者は本手法を実装したアプリケーションを用いる事で撮影対象領域の人数推定を行う。

推定精度を向上させるために、結果をクラウドサーバに



図 4: 動体領域の検出例

集め統合する。このような場合、撮影対象領域に存在する群衆人数の最尤値だけでなく、その領域に存在し得る群衆人数の最小値と最大値も収集し、確率分布として提示できることが望ましい。例えば、2 人のユーザから推定結果が送信された場合、それぞれの推定結果の統合が容易にできる。このような理由から、本研究では対象領域に存在する人数の最小値、最大値ならびに最尤値を推定する。

3.2 動体領域検出

本研究では、群衆人数に関するロバストな特徴量として、移動体が存在し得る領域 (動体領域) を用いる。本研究では、画像内に写る人以外の移動体を想定していないが、実環境では車のような様々な他の移動体が存在する。人以外の移動体の除去には物体認識の適用を想定している [8, 11, 12]。このような人以外の移動体は、人よりも大きく特徴的な形をしているため、高精度に認識が可能である。

動体領域を検出するために、図 3(a) に示すような 2 枚の短い時間間隔で撮影された画像を用いる。時間間隔は人の歩行速度やカメラと人の距離に依存し、動体領域を検出するために適切な値を設定する。本研究では実環境においては 1 秒、シミュレーションでは 0.7 秒という経験的に得られた値を用いる^{*1}。2 枚の画像を撮影した後、入力画像に対しリサイズ、グレースケール化、手ぶれ補正を行う。計算時間を短縮するため、まず 1/5 に画像を縮小する。手ぶれ補正には POC (Phase Only Correlation Method) [25] を用いることで、2 枚の画像に座標値でどれだけの変化があったかが分かり、その値を用いてアフィン変換することで手ぶれの補正を行う。その後、各ピクセルごとに差分をとりその絶対値を用いて図 3(b) に示す差分画像を生成する。

動体領域を検出するため、差分画像に対して二値化を行う。その際にノイズを除去するためにガウシアンフィルタを適用し、二値化には大津の手法と呼ばれる判別分析法を用いる [26]。判別分析法は画像内の各ピクセルの値に基づ

^{*1} 実環境とシミュレーションで設定時間が異なるのは、実環境においては人が移動する際に他の人の影響を受け、シミュレーションではそのような影響を考慮をしていないため、シミュレーション環境よりも実環境のほうが人の歩行速度が遅くなるためである。

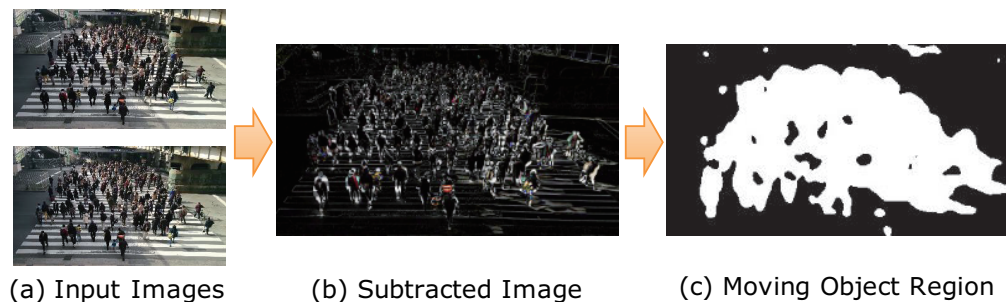


図 3: 動体領域抽出

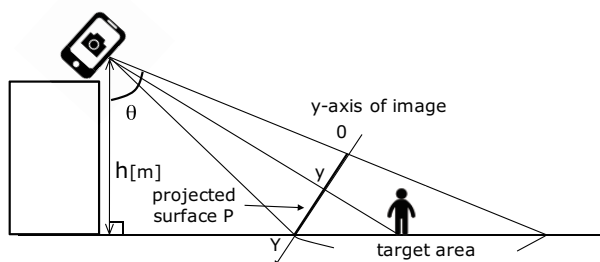


図 6: カメラの撮影姿勢

き、二値化の閾値を決定する。図 3(c) の白領域が検出した動体領域であり、図 4 に元画像と検出された動体領域を重ねた結果を示す。

3.3 群衆人数推定

本節では、動体領域内に存在し得る人数の最小値と最大値を推定する方法について述べる。先に述べたように、動体領域内の移動体として人のみが存在しているとする。直感的には動体領域面積は群衆人数と相関があると考えられるが、図 5 に示すように同じ動体領域でも異なる密度である可能性が考えられる。これは動体領域内では様々な人の分布と移動パターンが考えられるためである。

3.3.1 最小人数推定

図 5(a) に示すように動体領域 m において 1 人 1 人の動体領域が重なり合わない場合に、推定人数が最小となる。

よってこの場合の m における最小人数 $n_{min,m}$ は以下の式で推定できる。

$$n_{min,m} = \frac{s_m}{f_I(\theta, h, y)}, \quad (1)$$

s_m は m の面積であり、 $f_I(\theta, h, y)$ は図 6 のように撮影高度 $h[m]$ 、撮影角度 $\theta[rad]$ から撮影したときに、1 人の人が画像内の y 軸の位置 y において占有する動体領域面積である。

実際には群衆内のどの位置に人がいるか分からないため、位置 y として m の重心の座標を用いる。3D ゲームエンジン Unity を用いることで現実の人の動きを再現し、想定される θ 、 h 、 y の組み合わせについてあらかじめ $f_I(\theta, h, y)$



図 7: 異なる移動方向ごとの動体領域

をモデル化する。このモデルを歩行者面積モデルと呼ぶ。詳細は次節にて述べる。

加速度センサや気圧センサを用いることで、撮影時の角度と高度を推定し、撮影時の適切なモデル選択を行う。撮影角度は端末にかかる重力加速度を加速度センサから検出することで、容易に推定することができる。高度推定は文献 [27] で述べられているように、高度の変化に応じて気圧が変化する法則に基づいて行う。この法則によれば、同じ気温の環境においては $0.5[hPa]$ 変化するごとに $4m$ の高度変化があることが知られている。提案手法で必要なのは撮影対象領域から撮影地点までの高度であり、対象領域の高度推定を行うためには 2 地点間の気圧変化を得る必要がある。そのため、本研究ではクラウドソーシングによって様々な地点の気圧を収集する。ここで文献 [28] で述べられているように、スマートフォンの気圧センサは端末ごとに異なる特性があるため、文献 [29] で述べられている SBC (Scalable Barometer Calibration) を用いる。SBC は多くのスマートフォンの気圧センサに対しクラウドソーシングにより自動的にキャリブレーションを行う手法である。

3.3.1.1 歩行者面積モデル

歩行者面積モデルは平均的な体格の歩行者を再現した 3D モデルを用い構築する。歩行者を方向 $\vec{v}$ に向かって y 軸の位置 y から歩かせることで歩行を再現する。ここで、 x 軸方向にはカメラと歩行者の距離は大きく変化しないので、動体領域の x 軸方向に関する影響はないものとする。歩行を再現したシミュレーションで 2 枚の画像を撮影し、前述の動体領域検出を適用することで、動体領域面積を求める。

図 7 は異なる移動方向ごとの動体領域を示しており、移動方向ごとにその面積は異なる。すなわち、動体領域

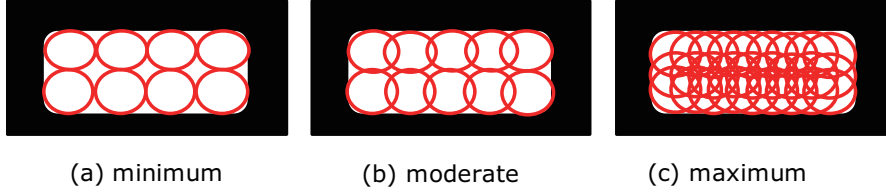


図 5: 同じ動体領域で密度が異なる場合

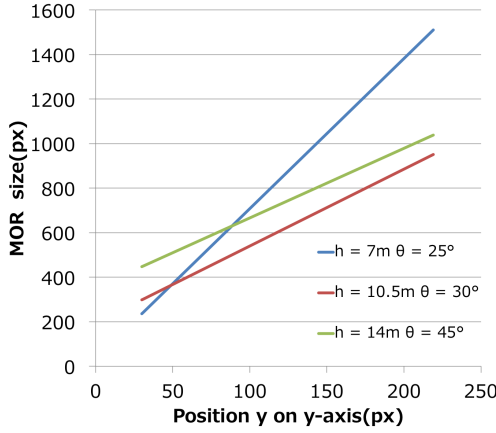


図 8: 歩行者面積モデル

面積は移動方向に依存することが分かる。Lucas-Kanade 法 [30] を用いたオプティカルフローの推定は一般的に 2 枚の画像を用いて物体の移動方向を推定するものである。オプティカルフローの推定には 2 枚の画像間における特徴点の追跡を行うが、特に群衆内に存在する人の特徴点を追跡することは難しいので、群衆のオプティカルフローを推定するのは困難である。よって我々は全ての移動方向に対する動体領域面積の平均を用いる。

想定される θ , h , y と移動方向に対するシミュレーション結果から、歩行者面積モデル $f_I(\theta, h, y)$ を構築する。1 人の占有面積と θ , h , y の相関を求めるために線形回帰を行い、以下のような式を求める。

$$f_I(\theta, h, y) = a_{\theta, h, y} \cdot y + b_{\theta, h, y}, \quad (2)$$

$a_{\theta, h, y}$ と $b_{\theta, h, y}$ は回帰係数である。図 8 はモデルの例である。

3.3.2 最大人数推定

最小の人数よりも動体領域内に人が多く存在している場合が多数存在する。文献 [31] によれば、物理的な群衆の密度の限界は $2.5[\text{人}/\text{m}^2]$ である。この事実に基づき動体領域内の最大人数 $n_{\max, m}$ を動体領域 m の実空間における面積 $s_{g(m)}$ を用いて以下のように推定する。

$$n_{\max, m} = 2.5s_{g(m)} \quad (3)$$

上式の $g(m)$ は画像中の動体領域を撮影対象領域の平面上に投影した領域である。カメラで撮影された画像に対し

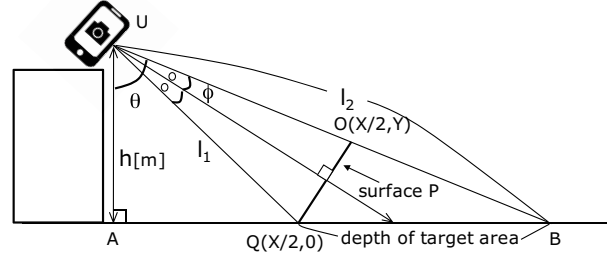


図 9: 横方向から見た元画像とアフィン変換後の画像

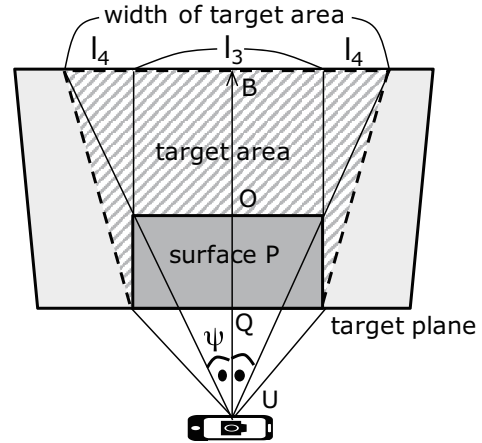


図 10: 真上から見た元画像とアフィン変換後の画像

てアフィン変換 [32] を適用することで、鳥瞰図に変換し、 $g(m)$ を得る。

図 9, 10 はそれぞれ横から見た場合と真上から見た場合の、カメラによって撮影された画像の投影面 P とアフィン変換後の画像を平面上に表した場合の図である。カメラの撮影角度を θ 、高度を h 、垂直画角を ϕ 、平行画角を ψ とする。図 9, 10 にある撮影対象領域の実空間における横幅と奥行きを幾何計算により算出する。

図 9 の θ , h , ϕ から実空間における l_1 の長さを求める。 $\triangle UAB$ と $\triangle UAQ$ から撮影対象領域の奥行きを算出し、 $\triangle UAB$ から l_2 を得る。同様に図 10 において、 UO と ψ の長さから l_3 が求まる。 l_4 の長さは ψ と OB の長さが $l_2 - l_1$ であることから求める事ができる。よって撮影対象領域の横幅は $2l_4 + l_3$ となる。

アフィン変換を行うためには、投影面 P の四隅の点の座標が必要である。これは図 10 の実空間の長さ l_3 が撮影した画像の x 軸方向の大きさ X に対応することを用いて算出できる。

実空間における撮影対象領域の奥行き d と横幅 w が求まるので、 d と w を掛け合わせることで、実空間における面積が算出できる。最終的に、以下の式で動体領域の実空間における面積が算出できる。

$$g(m) = dw \cdot r(m) \quad (4)$$

上式の $r(m)$ は、アフィン変換後の画像の撮影対象領域全体に対する動体領域 m の占有率である。

3.4 群衆人数推定

提案手法では先に述べた通り動体領域内に存在しうる最小人数と最大人数に基づき群衆人数推定を行う。文献 [31] によると歩行者の歩行速度と密度には相関があるが、既に述べたように群衆内の人に対してオプティカルフローを推定することは難しい。そこで、群衆内の密度は均一であるものとし、以下の式で定義される画像 J における群衆占有率 o を用いる。

$$o = \sum_{m \in J} \frac{s_m}{s_J}, \quad (5)$$

ここで s_J は J の移動可能な領域の面積である。

画像 J における全ての動体領域を内包する矩形領域の面積を s_J として得る。撮影対象領域の密度に群衆占有率 o は関係しているため、 o は密度を表す特徴量となる。

幾何的な関係を考慮すると、 o と密度の関係はカメラの撮影姿勢 h, θ による影響を受ける。カメラの撮影姿勢と密度の関係を考慮するため、撮影対象領域の密集度 $\beta(\theta, h, o)$ を以下の式で表す。

$$\beta(\theta, h, o) = \alpha_{\theta, h} \cdot o. \quad (6)$$

適切な係数 $\alpha_{\theta, h} (0 \leq \alpha_{\theta, h} \leq 1)$ を求めるため、様々な θ, h, o について網羅的にシミュレーションを行った。このシミュレーションでは他の遮蔽物がない広い空間を歩行者がランダムに移動することで、様々な人密度を再現する。シミュレーションにより以下の式の値が最も適切になる係数を選択する。

$$n_m = \beta(\theta, h, o)n_{max, m} + (1 - \beta(\theta, h, o))n_{min, m}, \quad (7)$$

n_m は動体領域 m の人数を表す。

最終的に 2 枚の連続した画像内の群衆人数 n を推定する。

$$n = \sum_{m \in J} n_m, \quad (8)$$

表 1: 様々な撮影高度と角度における絶対誤差 (%)

	7.0m	10.5m	14.0m
20°	11.4	9.31	10.3
25°	12.1	8.93	10.9
30°	11.9	8.76	8.57
35°	13.2	8.24	8.84
40°	16.2	10.5	8.70
45°	22.5	11.4	8.91

4. シミュレーションにおける性能評価

実環境における実験の前に、提案手法の性能を理想環境で調べるためにシミュレーションを行った。シミュレーションには Unity を用いて人の動きを再現し、シミュレータ内でスクリーンショットを撮影する。シミュレーションでは提案手法の群衆検出と群衆人数推定の性能を調べるために、人以外の物体は置かずに行った。実環境における実験は次章で述べる。画像処理は Java と OpenCV 2.4.10 を用いて提案手法を実装し、MacOS10.9.4 上で動作確認を行った。

まずはじめに、建物のフロア間の高さを 3.5m とし、7m, 10.5m, 14m の撮影高度と 20°, 25°, 30°, 35°, 40°, 45° の異なる撮影角度についてそれぞれの組み合わせのシミュレーション画像を撮影した。すべての高度と角度の組み合わせについて前述したモデルをあらかじめ構築した。

表 1 の結果はそれぞれの高度と角度のペアについて画像内の人を 5 人から 120 人の間で変化させたときの平均誤差を示している。全ての場合について、平均誤差が 11.1% となり、本研究が目的とする群衆人数推定においては高精度な人数推定を実現できた。7.0m-45° において誤差が 20% 以上となっている理由は他の高度と比べて地面との距離が近く、さらにこの角度の場合は他の角度に比べて遠い地点が写っていないため、画像内に写る全体の群衆人数が少なくなり、相対的に誤差が大きくなってしまったためである。撮影角度の影響は地面との距離が最も近い高度が 7.0m のときに顕著に見られ、他の高度の場合はややその傾向が現れる結果となった。

さらに結果を詳細に見るために、(7.0m, 45°), (10.5m, 30°), (14.0m, 25°) の場合について実人数と推定人数をプロットしたものを図 11 に示す。図を見ると分かるように、(7.0m, 45°) の場合にはほぼ結果が真値と一致していることが分かる。3 つの場合について同じ実人数におけるそれぞれの誤差はほぼ同じものであり、高度と角度に基づいて構築した提案手法のモデルが異なる高度と角度の場合の差異を吸収できていることが確認できた。

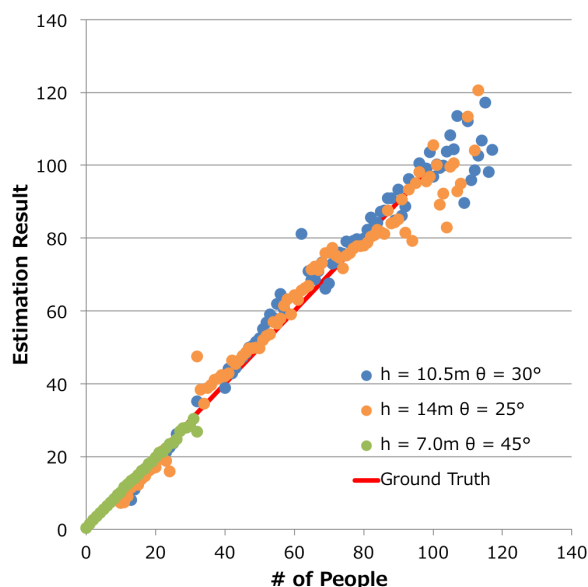
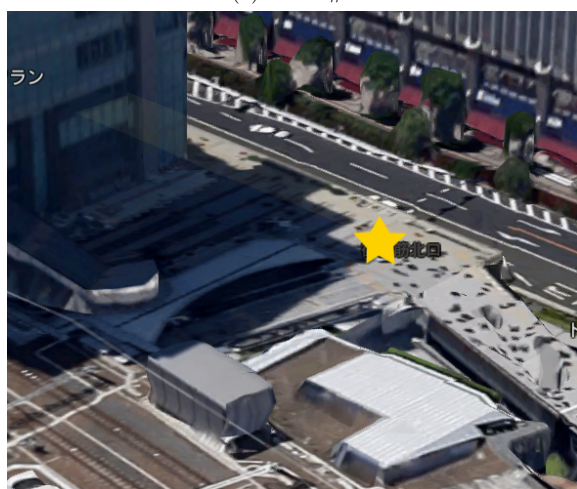


図 11: シミュレーションにおける人数推定結果



(a) Place #1



(b) Place #2

図 12: 撮影場所



図 13: 画像上部に高密度な群衆が存在する場合

5. 実環境における性能評価

5.1 実験環境

実環境における性能評価のため、Android5.1.1 上で OpenCV を用いてアプリケーションを作成した。本アプリケーションのカメラを起動し撮影ボタンを押すと自動的に 1.0 秒の時間間隔で 2 枚の画像を撮影し、同時に気圧センサと加速度センサの値も保存する。撮影高度は撮影対象領域と撮影地点の間の気圧差によって推定を行う。撮影対象領域の気圧は前述した通りクラウドソーシングを用いて GPS の位置情報に基づき取得する。本アプリケーションを用いて図 12 に示すような高所から 12 人のボランティアによって約 300 枚の画像を撮影した。実験は 2 日間の休日と 1 日の平日の計 3 日間に行った。既存手法との性能を比較するため、人の頭部が写っている 5,322 枚の正解画像と 13018 枚の非正解画像を用意し、HOG 特徴量を AdaBoost によって学習しカスケード型分類器を生成した。

実環境にはシミュレーション環境には無かった、性能に影響を与える要因が存在している。例えば、撮影対象領域がカメラから非常に遠いところに高密度な群衆がある場合は難しい。このような高密度な群衆の場合には、カメラからの距離が遠いため、手ぶれの影響を受けやすい。その例を図 13 に示す。この図を見ると分かる通り、遠い場所になっている画像上部の領域に高密度な群衆が存在している。本システムの参加者が異なる場所からこの群衆を撮影することを考慮すると、このような群衆は誤った推定結果を出さないために除去すべきである。そこで、人数推定に誤りが生じる可能性が高い画像の上から 1/3 を除去し、人数推定を行う。

5.2 実験結果

実環境における実験結果を図 17 に示す。推定結果は真値に近く、平均人数誤差が 11.3 人となることが分かった。この結果から実環境においてもシミュレーション環境とほぼ同様の結果となり、本システムが実環境においても有効

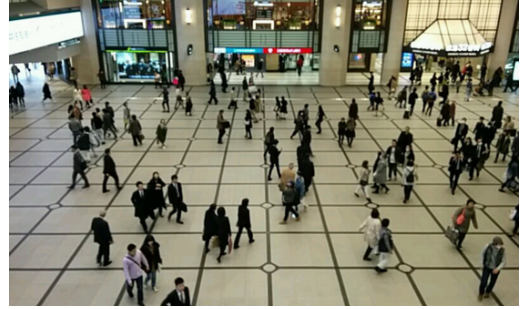


(a) 元画像



(b) 動体領域

図 14: 横断歩道における例

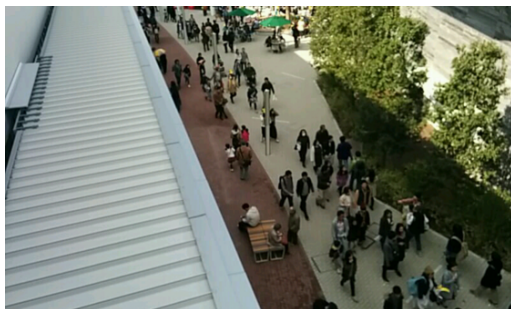


(a) 元画像

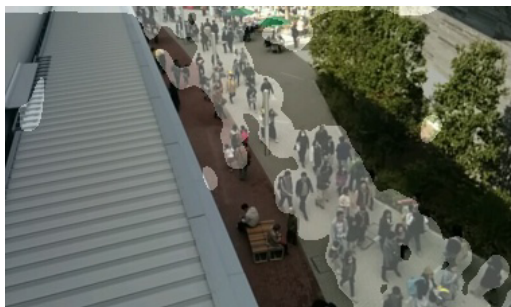


(b) 動体領域

図 16: ショッピングモールの例



(a) 元画像



(b) 動体領域

図 15: 屋外の歩道の例

であることが分かった。極めて高密度な場合には、群衆人数の増加に伴う動体領域面積の増加があまり見られず、大きい誤差の要因となる場合がある。実環境における異なる地点の例を図 14, 15, 16 に示す。

また、HOG 特徴量に基づいた既存手法との比較結果を

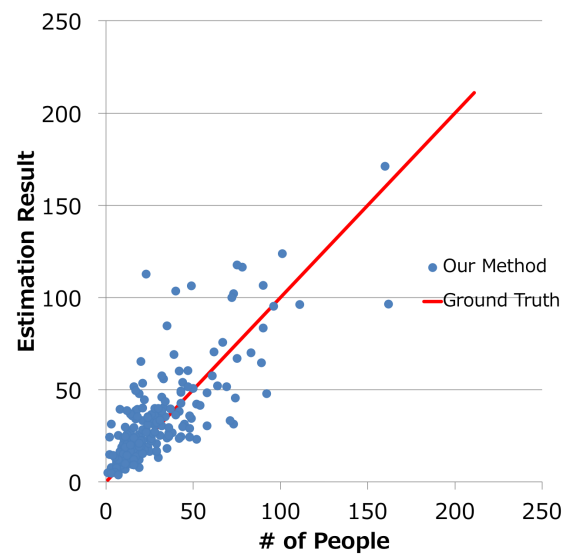


図 17: 実環境における結果

図 18 に示す。群衆人数が少ない場合には既存手法の方が精度が高い場合も存在するが、群衆人数が多くなった場合には、既存手法の誤差が非常に大きくなっている。一方、提案手法では群衆内の人数に関わらず精度の高い人数推定が実現できており、ロバスト性が高いことが分かる。

最後に、スマートフォン上における提案手法の実行時間について述べる。Nexus5 上で実行したところ平均実行時間は 320ms となった。対して HOG 特徴量に基づく手法では 220ms となり提案手法よりも実行速度が早かった。し

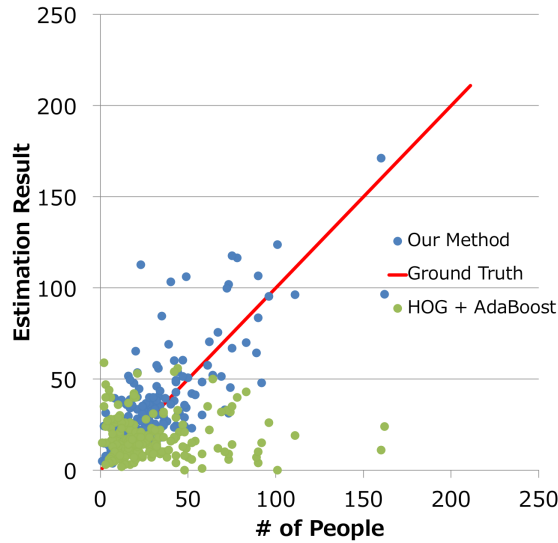


図 18: HOG 特徴量を用いた既存手法との比較

かし、ほぼ同じ実行速度で実行でき、群衆人数推定における精度が高いことから、提案手法は実行時間と精度を合わせ持った手法であると言える。

6. アプリケーション実装

本研究では (i) 推定した群衆人数をアップロードする機能、(ii) Google Maps API を用いて群衆情報をマッピングする機能をスマートフォン上に実装した。推定結果をアップロードする際には位置情報や撮影姿勢も同時にアップロードし、対応する位置のマップ上に群衆人数をプロットする。JR 大阪駅周辺で実験したときのアプリケーションのスクリーンショットを図 19 に示す。推定結果を円で表し、さらに推定結果を分かりやすく可視化するためにヒートマップのように推定結果によって円の色を変化させた。詳細にマッピングする人の存在領域は 3 章で述べた動体領域のアフィン変換後の画像を用いる。動体領域をアフィン変換後の実空間における幅と奥行きを用いて、マップに対する縮尺を合わせる。また撮影地点の位置情報は GPS から取得することができ、撮影対象領域の位置情報を撮影高度、撮影角度から算出する。撮影地点と撮影対象領域の経度緯度をそれぞれ $A(x_1, y_1)$, $B(x_2, y_2)$ とすると、以下の式で A から B の方角 ω を求めることが可能である。

$$\Delta x = x_2 - x_1 \quad (9)$$

$$\omega = 90 - \text{atan2}(\sin \Delta x, \cos y_1 \cdot \tan y_2 - \sin y_1 \cdot \cos \Delta x) \quad (10)$$

以上の位置情報と方角を用いて実際に撮影した領域に正確に重ね合わせる。この結果、図 20 のように JR 大阪駅前の横断歩道に対する推定結果および動体領域を正確にマッピングできていることが確認できる。

今後、最大推定値や最小推定値と群衆人数推定値を用い



図 19: 推定結果のマッピング

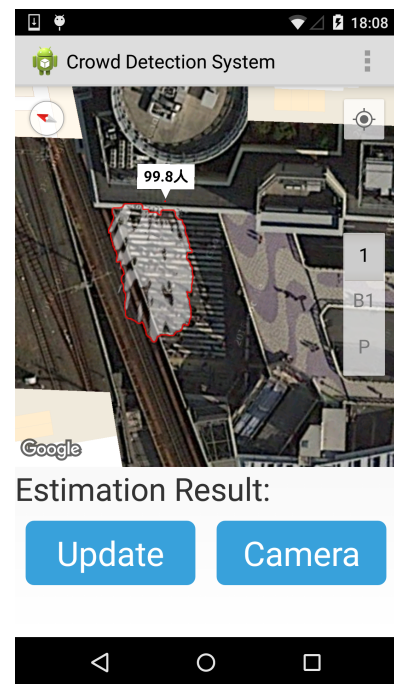


図 20: 航空写真に動体領域を重ね合わせた Google Map

る事で、異なる円同士が重なっている部分の情報を統合する機能を追加する予定である。

7. まとめ

本研究では、高い建物や歩道橋から撮影された画像を用いることで軽量の群衆人数推定を実現した。提案手法は、(i) スマートフォンで動作し、動画に基づく手法と比較し

てシステム参加者の負担が小さく、(ii) 個別に人を検出するよりも高精度な人数推定を行えるため、クラウドソーシングによる群衆密度の収集には有用である。動体領域、撮影高度と撮影角度を入力として動体領域の面積をモデル化し、そのモデルによって動体領域内の群衆人数推定を行う。実験結果では約 10% の誤差で人数推定が実現できており、高精度な群衆人数推定が実現できることが分かった。

今後、Yahoo!クラウドソーシング等のサービスを利用することで、より多くの群衆人数推定結果を収集し、時空間的に異なる群衆分布のデータベースを構築したいと考えている。

参考文献

- [1] 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.
- [2] Brent Hecht and Monica Stephens. A tale of cities: Urban biases in volunteered geographic information. In *Proceedings of the AAAI International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, pp. 197–205, 2014.
- [3] Shinji MIYAZAKI, Hiroyoshi MIYANO, Hiroo IKEDA, and Ryoma OAMI. New congestion estimation system based on the “crowd behavior analysis technology”. *NEC Technical Journal*, Vol. 9, No. 1, pp. 78–81, 2015.
- [4] Jens Weppner and Paul Lukowicz. Collaborative crowd density estimation with mobile phones. In *Proceedings of the 9th ACM Conference. Embedded Networked Sensor Systems*, pp. 193 – 200, 2011.
- [5] Jonathan Reades, Francesco Calabrese, Andres Sevtsuk, and Carlo Ratti. Cellular census: Explorations in urban data collection. *IEEE Pervasive Computing*, Vol. 6, No. 3, pp. 30–38, 2007.
- [6] A. B. M. Musa and Jakob Eriksson. Tracking unmodified smartphones using wi-fi monitors. In *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems*, SenSys ’12, pp. 281–294. ACM, 2012.
- [7] Sibren Isaacman, Richard Becker, Ramón Cáceres, Margaret Martonosi, James Rowland, Alexander Varshavsky, and Walter Willinger. Human Mobility Modeling at Metropolitan Scales. In *Proceedings of the 10th International Conference on Mobile Systems, Applications and Services (MobiSys ’12)*, pp. 239–252, 2012.
- [8] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (CVPR ’05)*, pp. 886–893, 2005.
- [9] Kentaro Toyama, John Krumm, Barry Brumitt, and Brian Meyers. Wallflower: Principles and practice of background maintenance. In *Proceedings of the 7th IEEE International Conference on Computer Vision*, pp. 255–261, 1999.
- [10] 小島颯平, 内山彰, 廣森聡仁, 山口弘純, 東野輝夫. 俯瞰画像における被写体サイズ推定モデルに基づく群衆人数推定手法の提案. 情報処理学会, マルチメディア, 分散, 協調とモバイル (DICOMO2015) シンポジウム論文集, pp. 157–165, 2015.
- [11] Piotr Dollar, Zhuowen Tu, Pietro Perona, and Serge Belongie. Integral channel features. In *Proceedings of the British Machine Vision Conference*, pp. 91.1–91.11, 2009.
- [12] Paul Viola, Michael J. Jones, and Daniel Snow. Detecting pedestrians using patterns of motion and appearance. In *Proceedings of the 9th IEEE International Conference on Computer Vision (ICCV ’03)*, pp. 734–741, 2003.
- [13] Etienne Corvee, Francois Bremond, et al. Haar like and LBP based features for face, head and people detection in video sequences. In *Proceedings of International Workshop on Behaviour Analysis and Video Understanding (ICVS 2011)*, pp. 10–20, 2011.
- [14] Krystian Mikolajczyk, Cordelia Schmid, and Andrew Zisserman. Human detection based on a probabilistic assembly of robust part detectors. In *Proceedings of the 8th European Conference on Computer Vision (ECCV ’04)*, pp. 69–82, 2004.
- [15] Anthony C. Davies, Jia Hong Yin, Sergio Velastin, et al. Crowd monitoring using image processing. *Electronics & Communication Engineering Journal*, Vol. 7, No. 1, pp. 37–47, 1995.
- [16] A. B. Chan, Zhang-Sheng John Liang, and N. Vasconcelos. Privacy preserving crowd monitoring: Counting people without people models or tracking. In *IEEE Conference on Computer Vision and Pattern Recognition 2008 (CVPR 2008)*, pp. 1–7, June 2008.
- [17] Beibei Zhan, Dorothy N. Monekosso, Paolo Remagnino, Sergio A. Velastin, and Li-Qun Xu. Crowd analysis: a survey. *Machine Vision and Applications*, Vol. 19, No. 5, pp. 345–357, 2008.
- [18] Markus Enzweiler and Dariu M Gavrila. Monocular pedestrian detection: Survey and experiments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 31, No. 12, pp. 2179–2195, 2009.
- [19] Francois Fleuret, Jerome Berclaz, Richard Lengagne, and Pascal Fua. Multicamera people tracking with a probabilistic occupancy map. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 30, No. 2, pp. 267–282, 2008.
- [20] Robert T. Collins. Mean-shift blob tracking through scale space. In *Proceedings of Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp. II–234, 2003.
- [21] Huijing Zhao and Ryosuke Shibasaki. A novel system for tracking pedestrians using multiple single-row laser-range scanners. *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, Vol. 35, No. 2, pp. 283–291, 2005.
- [22] Hokuyo Automatic Co., LTD. Scanning range finder, UTM-30LX, 2010.
- [23] Francesco Calabrese, Francisco C. Pereira, Giusy Lorenzo, Liang Liu, and Carlo Ratti. *Pervasive Computing: 8th International Conference, Pervasive 2010, Helsinki, Finland, May 17-20, 2010. Proceedings*, chapter The Geography of Taste: Analyzing Cell-Phone Mobility and Social Events, pp. 22–37. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.
- [24] Pravein Govindan Kannan, Seshadri Padmanabha Venkatagiri, Mun Choon Chan, Akhihebbal L Ananda, and Li-Shiuan Peh. Low cost crowd counting using audio tones. In *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems (SenSys ’12)*, pp. 155–168. ACM, 2012.
- [25] C.D. Kuglin. The phase correlation image alignment method. In *Proceedings of the International Conference*

- Cybernetics and Society*, pp. 163–165, 1975.
- [26] Mehmet Sezgin and Bülent Sankur. Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*, Vol. 13, No. 1, pp. 146–168, 2004.
 - [27] Guangwen Liu, Masayuki Iwai, Yoshito Tobe, Dunstan Matekenya, Khan Muhammad Asif Hossain, Masaki Ito, and Kaoru Sezaki. Beyond horizontal location context: Measuring elevation using smartphone’s barometer. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, UbiComp ’14 Adjunct, pp. 459–468, 2014.
 - [28] Kartik Muralidharan, Azeem Javed Khan, Archan Misra, Rajesh Krishna Balan, and Sharad Agarwal. Barometric phone sensors: More hype than hope! In *Proceedings of the 15th Workshop on Mobile Computing Systems and Applications*, HotMobile ’14, pp. 1–6, 2014.
 - [29] Haibo Ye, Tao Gu, XianPing Tao, and Jian Lu. SBC: scalable smartphone barometer calibration through crowdsourcing. In *11th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services, MOBIQUITOUS 2014*, pp. 60–69, 2014.
 - [30] Bruce D Lucas, Takeo Kanade, et al. An iterative image registration technique with an application to stereo vision. In *IJCAI*, pp. 674–679, 1981.
 - [31] S. J. Older. Movement of pedestrians on footways in shopping streets. *Traffic Engineering and Control*, Vol. 10, No. 4, pp. 160–163, 1968.
 - [32] Richard Szeliski. *Computer Vision: Algorithms and Applications*. Springer, 2011.