

# 群衆をマクロにとらえる画像処理技術

～人物・群衆の幾何モデルに基づいた人数推定とその応用～

新井 啓之<sup>1</sup> 伊藤 直己<sup>1</sup> 谷口 行信<sup>1</sup>

**概要:** デジタルサイネージの広告効果測定をはじめとして、その場所にいる人数（画像内人数）またはその場所を通過した人数（通過人数）を計測／推定する画像処理技術が必要とされている。これまで提案されてきた多くの手法は個々の人物の検出や追跡を基本としており、画面内に数人程度が存在するような比較的混雑していない状況では実用的な精度を実現するが、大きな駅のコンコースで発生するような数 10～100 人規模の混雑状況に適用することは困難であった。本稿では、個人個人の検出を行うことなく群衆をマクロにとらえることで、画像内人数および通過人数を推定する手法を紹介する。本手法では校正された単眼の固定カメラを利用し、画像上での局所的な観測量から、各画素と実空間の対応関係および人物形状やオクルージョンの幾何的なモデルに基づき人数を推定する。さらに、これまで実施してきた実証実験の事例を交えながら実用上の課題についても紹介する。

## 1. はじめに

都市や公共空間において不特定多数の人（群衆）の動きや状態を把握する技術は、マーケティング、都市計画、安全管理等の分野において必要とされている。特にデジタルサイネージの広告効果測定は、近年、そのニーズが顕在化し、画像処理技術を用いた計測システムが利用され始めている。また、これと並行する形で、広告効果を示す指標の提案、標準化も進められている [1]-[3]。2001 年に屋外広告調査フォーラムが提案した指標である DEC(Daily Effective Circulation)[1] は、主として看板等の屋外広告をを対象としてしており、その屋外広告を見る可能性がある一日あたりの通行量（人数）として定義されている。また 2008 年には米国の業界団体 DPAA（旧 OVAB）が、デジタルサイネージの広告効果測定のガイドライン：Audience Metrics Guidelines[2] を提案している。この中では、広告を何人の人が何回見たかの積算値である Average Gross Unit Impression という指標が提案されている。DEC においてはその場に現れた人数が基本指標となり、一方、Average Gross Unit Impression では、顔認識等による視聴者数が基本指標となるが、対象エリアが広い場合など全ての人物の顔認識が難しい場合には、その場に現れた人数が重要な指標となる。一方、駅、空港、地下街等の人が多く集まる場所において群衆の状態をモニタリングし、状況に応じて

適切にコントロールする安全管理支援のニーズも根強く存在している。これらの利用場面においては、その場所にいる人数：画像内人数、またはその場所を通過した人数：通過人数を計測／推定する画像処理技術に対して大きな期待が寄せられている。

これまで既に実用化された技術も含め、人数計測／推定を実現する様々な画像処理技術が提案されているが、大雑把に言えば、これらは以下の 3 つの要素技術もしくはその組み合わせによって実現されている。1 番目は形状検出であり、全身、上半身、顔などの特定形状を検出するものである [4][5]。2 番目はトラッキング（追跡）であり、画像中の領域や特徴点の時間方向での対応付けに基づき、それぞれの人物の移動経路を求めるものである [6]-[9]。3 番目は特徴量の学習による人数推定であり、背景差分やテクスチャ解析などにより算出した画像全体または映像の所定の時間区間を代表する特徴量と人数との関係を事例学習により事前に獲得しておくことで人数推定を行うものである [10]-[13]。このうち形状検出は、画像内の人数が多くなり画像上で手前の人物が奥の人物の一部を遮蔽してしまう所謂オクルージョンが頻発する状況では、所定の形状が画像上で正しく観測されにくくなるという問題がある。トラッキングは、近年、混雑状況への適用性を高めようとする多くの改良がなされてきているものの、混雑時においては領域や特徴点を時間方向で対応付けることが本質的に困難となり、精度は低下する。さらに混雑時においては群衆内の人物の動き（歩行速度と方向）が揃ってくる傾向があり、このことが、

<sup>1</sup> NTT メディアインテリジェンス研究所  
Yokosuka-shi, Kanagawa, 239-0847, Japan  
arai.hiroyuki@lab.ntt.co.jp

個々の人物を分離することを一層難しくする。一方、特徴量の学習による人数推定は、混雑時においても特徴量の算出が可能であり、混雑時にも適用可能であるという特長を持つが、利用する場所毎、カメラ毎に大量の事例データを用意する必要があり、導入の手間（コスト）が大きくなってしまふ。

これら従来手法の問題を解決すべく、混雑状況にも適用可能でかつ事例データを必要としない画像内人数推定技術、さらにこれを発展させる形で通過人数推定技術を実現した。これらの技術では個人個人の検出を行わずに群衆をマクロにとらえる形で人数推定を実現している。以下、提案技術のフレームワーク、画像内人数推定技術、通過人数推定技術について具体的に説明するとともに、本技術の実用化に向けての取り組み、課題等についても紹介する。

## 2. 人数推定のフレームワーク

図1に本手法の基本的な処理の流れを示す。本手法は斜め下向きに設置された較正済みのカメラを利用する。図1の上側が画像内人数推定の流れ、下側が通過人数推定の流れを示している。いずれも画像上の局所的な観測量をカメラ較正データに基づき実空間に変換し、そこからマクロな物理量を推定、最後にオクルージョンの影響を見積り、補正することで最終的な物理量を得るという基本的な流れは共通している。一般に斜め下向きのカメラで床面上に立っている人物を撮影する場合、画像上の観測量を実空間での観測量に一意に対応付けることができないという問題が存在する。例えば画像上の面積と実空間の人物上の面積を対応付けようとしても、人物の立ち位置によってその対応関係は変わってくる。同様に、人物の位置が不明である場合、画像上で局所的に観測された動きベクトルを実空間の動きベクトルに一意に対応付けることはできない。以下に示す画像内人数推定および通過人数推定では、このような画像と実空間の対応の任意性をいかにして回避、解消するかがアルゴリズムの中核的なアイデアとなっている。以下、順を追って説明していく。

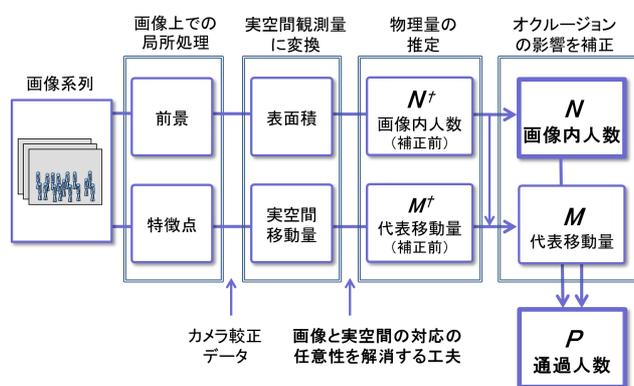


図1 本手法の処理の流れ

## 3. 画像内人数推定

ここでは画像内人数推定のアルゴリズムを示す。以下、本手法の前提条件、人数推定の基本的な考え方と課題、その課題を解決するために重要な役割を果たす基準表面積 (SSA: Standard Surface Area) の定義とその重要な性質を示し、この性質を利用した人数推定方法を示す。最後にオクルージョンの影響を見積り、補正する方法を示す。

### 3.1 前提条件

本手法は、以下の4つを前提条件としている (図2参照)。

- (a) 床面に対して較正済みの固定カメラを利用
- (b) 人物のシルエットを矩形で近似 (サイズ固定)
- (c) 画像上の人物幅  $\gg$  画素幅
- (d) 人物は床面上に乱雑に分布

ここで床面に対してカメラが較正済みというのは、画像上の各画素と実空間上の直線が対応づけられる状態にあることを指すものとする。

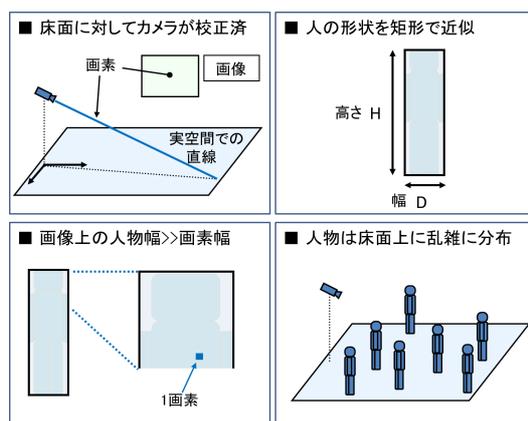


図2 アルゴリズムの前提条件

### 3.2 画像内人数推定の基本アイデア

本手法の基本アイデアは、画像から得られた前景 (人物のシルエット) と実空間における人物上の表面積とを対応付けることで画像内的人数を推定しようとするものである。人物のシルエットを高さ  $H \times$  幅  $D$  ( $H, D$  は既知の固定サイズ) の矩形で近似し、その人物 (矩形) が床面上の既知の位置  $(X, Y)$  に立っている場合を考える (図3参照)。ここで、画素  $(x, y)$  に対応する、床面上の位置  $(X, Y)$  に立つ人物上の表面積を  $s(x, y, X, Y)$ 、その人物の画像における前景画素 (シルエット) の集合を  $f_1$ 、実空間における人物の表面積 (画像上にシルエットを与える人物断面の実空間における面積) を  $S_0$  とすると以下の関係が成り立つ。

$$\sum_{(x,y) \in f_1} s(x, y, X, Y) = S_0. \quad (1)$$

人物のシルエットをカメラに正対する平面 (床面上に立つ

矩形)で近似する場合、 $s(x, y, X, Y)$ の値は画素 $(x, y)$ をカメラ校正データに基づき実空間上に逆投影することで計算できる。もし1人の人物の前景が画像上で完全に検出されていれば、その前景画素に対して $s(x, y, X, Y)/S_0$ の和をとれば、その値は理想的には1となる。つまり画素 $(x, y)$ が前景であることは、床面上の位置 $(X, Y)$ にいる人物の $s(x, y, X, Y)/S_0$ 人分の表面積が観測されたことを意味する。したがって、画像内でオクルージョンが発生していないのであれば、画像内人数 $N$ は次式、

$$N = \frac{\sum_{(x,y) \in F} s(x, y, X, Y)}{S_0}, \quad (2)$$

により推定することができる。ここで $F$ は画像内の前景画素全体の集合である。

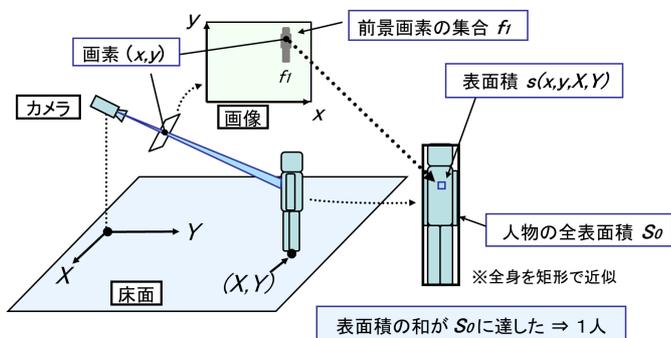


図3 画像内人数推定の基本的なアイデア

### 3.3 基準表面積 (SSA) の定義

前述したように人物の床面上位置 $(X, Y)$ が既知であるという前提は、とりわけ混雑時には現実的ではない。人物の床面上の位置が不明である場合、ある画素に対応する人物上の表面積は図4のように様々な値をとることになる。したがって、式(2)をそのまま適用することはできない。このような対応の任意性の問題を解決するため、基準表面積 (SSA: standard surface area: SSA) と呼ぶ概念を導入する。画素 $(x, y)$ は人物の様々な位置と対応しうるが、

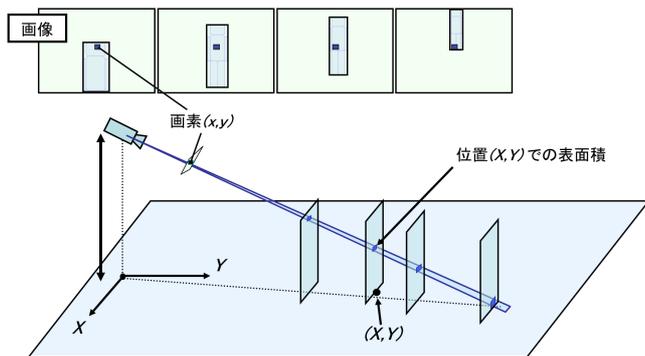


図4 画素と実空間上の表面積の対応の任意性

対応する表面積は実空間での対応位置の高さにのみ依存する。このことに注目して基準表面積 $\hat{s}(x, y)$ を、

$$\hat{s}(x, y) = s(x, y, X_\eta, Y_\eta), \quad (3)$$

と定義する。ここで、 $(X_\eta, Y_\eta)$ は、画素 $(x, y)$ の逆投影が床面から所定の高さ $\eta$ の位置になるような床面上の人物(矩形)の位置を表す。 $\eta$ は固定パラメータであるが、カメラの床面高さ $T_z$ に関して $\eta < T_z$ の範囲で任意に決めることができる。この基準表面積 $\hat{s}(x, y)$ は、その画素(の逆投影)が実空間においてどのように広がっていくのかを定量的に表現するものである。

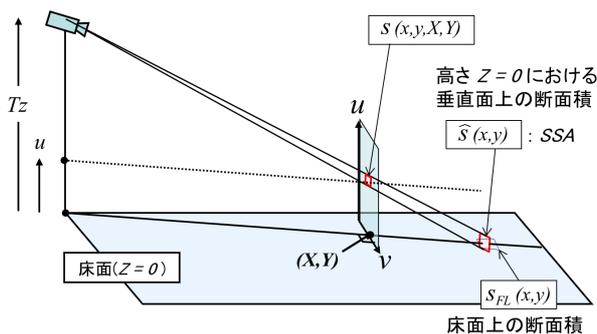


図5 基準表面積の定義(床面:  $\eta = 0$ で定義した場合)。

### 3.4 基準表面積の和の不変性

前述の式(1)と同様に、基準表面積 $\hat{s}_{x,y}$ の1人分の前景画素に対する和を考えてみる。記述の簡単化のため本稿では $\eta = 0$ の場合を考える。また人物は、実空間においてカメラ方向に正対する高さ $H$ 、幅 $D$ の矩形で近似する。前述の前提条件(c)は画像の解像度が十分に高いことを意味しており、この場合、 $\hat{s}(x, y)$ の和は次式

$$\sum_{(x,y) \in f_1} \hat{s}(x, y) = \iint_{F_1} \hat{s}(x, y) dx dy, \quad (4)$$

のように積分で近似できる。ここで $F_1$ は画像上の1人分の前景画素の集合 $f_1$ に対応する連続的な領域を表す。図5に示したように、床面上の位置 $(X, Y)$ でカメラ方向に正対する矩形における座標系 $(v, u)$ を取ると、 $s(x, y, X, Y)$ の定義から、

$$\frac{dudv}{dx dy} = s(x, y, X, Y) \quad (5)$$

が成り立つ。式(5)を式(4)に代入することで、

$$\sum_{(x,y) \in f_1} \hat{s}(x, y) = \iint_{F'_1} \frac{\hat{s}(x, y)}{s(x, y, X, Y)} dudv. \quad (6)$$

を得る。ここで上記の積分範囲 $F'_1$ は、 $(x, y)$ から $(v, u)$ の座標変換に伴い、矩形上の表面全体に置き換わっている。式(6)右辺の被積分関数は、図5において、 $s(x, y, X, Y)$ と

$\hat{s}(x, y)$  をそれぞれ底面とし、カメラを頂点とする四角錐を  
考えることにより、

$$\frac{\hat{s}(x, y)}{s(x, y, X, Y)} = \left( \frac{T_z}{T_z - u} \right)^2. \quad (7)$$

であることがわかる. 式 (7) を式 (6) に代入し、右辺の積  
分を実行することにより、

$$\begin{aligned} \sum_{(x,y) \in f_1} \hat{s}(x, y) &= \int_{F_1'} \int \frac{T_z^2}{(T_z - u)^2} dudv \\ &= \int_0^D \int_0^H \frac{T_z^2}{(T_z - u)^2} dudv \\ &= S_0 \left( 1 + \frac{T_z H}{(T_z - H) T_z} \right) \\ &= S_{unit}. \end{aligned} \quad (8)$$

を得る. 式 (8) を見ると、基準表面積  $\hat{s}(x, y)$  の 1 つの矩形  
上での和を計算した結果は  $x, y, X, Y$  のいずれにも依存せ  
ず、カメラの高さ  $T_z$ 、人物 (矩形) の実空間における高さ  
 $H$  および矩形の実空間での表面積  $S_0$  にも依存している  
ことがわかる. つまり基準表面積の和は、対象となる矩形  
の位置に依存しない不変量となっている (図 6 参照). こ  
の基準表面積の和を  $S_{unit}$  と表記する.

式 (2) と同様の考え方により、もし画像内でオクルージョ  
ンが発生していないとすれば、画像内人数は次式、

$$N^\dagger = \frac{\sum_{(x,y) \in F} \hat{s}(x, y)}{S_{unit}}, \quad (9)$$

によって推定することができる. ここで  $F$  は画像中の前  
景画素全体の集合である. 式 (9) により、前景検出結果の  
みから、個々の人物の位置を決めることなく画像内人数を  
推定することが可能となる. しかし、特に混雑時においては  
画像上でのオクルージョンが頻発するため、式 (9) によ  
る推定値は実際の人数より少ない値となる.

1人分の前景画素の集合

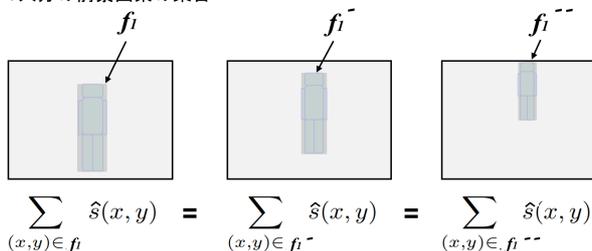


図 6 基準表面積の和の不変性

ここで基準表面積  $\hat{s}(x, y)$  の具体的な算出方法を簡単に  
説明しておく. まず注目する画素  $(x, y)$  について、周囲の  
4 点  $(\{x \pm 1/2\}, \{y \pm 1/2\})$  を床面上に逆投影し、得られた  
四角形 (台形) の面積  $S_{FL}$  を求める (図 5 参照). またこ  
の画素  $(x, y)$  が床面に対してなす角度  $\alpha(x, y)$  を算出する.

これらを用いて基準表面積  $\hat{s}(x, y)$  ( $\eta = 0$  の場合) は、

$$\hat{s}(x, y) = s_{FL}(x, y) \cdot \tan[\alpha(x, y)]. \quad (10)$$

として計算することができる.

### 3.5 オクルージョンの影響の補正

とりわけ混雑時においてはオクルージョンがしばしば発  
生することにより、画像上で観測される人物の面積が減少  
し、式 (9) による推定人数は、実際の人数より小さくなる.  
ここでは、オクルージョンの影響を見積り、推定人数を補  
正するアルゴリズムを紹介する.

オクルージョンの影響を定式化するために、重畳度  $R$   
と呼ぶ量を考える. 重畳度  $R$  は、1 つの画素に対応する実  
空間上の直線上に何人 ( $R$  人) の人がいるかの期待値とし  
て定義される. したがって  $R$  は、混雑度合、具体的には式  
(9) で得られる  $N^\dagger$  に依存する. また、それぞれの画素に  
よって床面となす角度が異なることから、本来  $R$  の値も画  
素毎に異なるが、ここでは大雑把な近似として、画像の中  
心付近の画素における重畳度  $R$  を画像全体を代表する値  
として考えることとする. 重畳度  $R$  の導出は先行文献 [15]  
に記述されているので、ここでは結果のみを記載する. 図  
7 に示すように、画像が撮影する床面上の面積を  $A$ 、画像  
の中心画素と床面とのなす角度を  $\phi$ 、人物を近似する矩形  
の実空間における幅を  $D$ 、高さを  $H$  とすると、重畳度  $R$   
は、カメラから見た場合に矩形の影になる領域に何人が入  
るのかを見積もることで、

$$R \cong \frac{1}{1 - N^\dagger \frac{DH \tan \phi}{A}}, \quad (11)$$

となる. オクルージョンの発生時には、重畳度  $R$  にした  
がって画像上で観測される表面積 (画素数) が減少すると  
考えることができるので、オクルージョンの影響を次式、

$$N \cong RN^\dagger. \quad (12)$$

により補正することができる. この  $N$  が、オクルージョ  
ンの影響を補正した画像内人数の最終的な推定値となる.

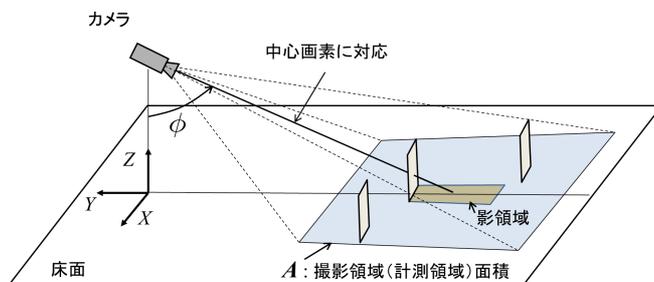


図 7 オクルージョンのモデル化

: カメラから見た影領域に入る人数で定量化

#### 4. 通過人数推定技術

次に、画像フレーム系列から、群衆全体としての移動量（代表移動量）を推定し、前述の画像内人数の推定結果とを合わせて通過人数を推定する方法を紹介する（処理の大まかな流れは図1参照）。本手法では、混雑時においては、群衆を構成する個人個人の動き（方向や速度）の自由度が下がり、全体として動きが揃ってくることに注目している。

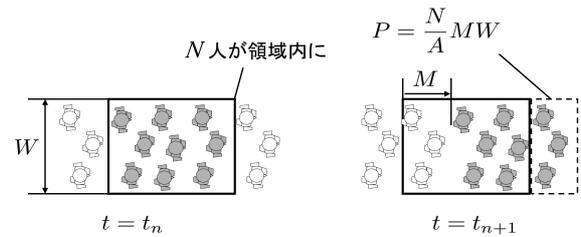


図8 通過人数計測の原理

##### 4.1 通過人数推定の基本アイデア

以下で示す手法は、流体の流量が断面積、密度、流速の積で求められることと同様の原理に基づいている。時刻  $t = t_n$  において、面積  $A$  の実空間領域内に  $N$  人が乱雑に存在しており、これらの人々が  $t = t_{n+1}$  までの間に全体として距離  $M$  だけ移動した場合を考える（図8参照）。この領域の断面長（移動方向に垂直な断面長）を  $W$  とすると、この時間内に領域から出た人数は、

$$P = \frac{N}{A} MW, \quad (13)$$

として推定することができる。領域の面積  $A$  と断面長  $W$  は固定値なので、領域内的人数  $N$  と所定時間内の群衆の代表移動量  $M$  を計測/推定することができれば通過人数  $P$  を得ることができる。 $N$  は前述の画像内人数を適用すればよいので、以下では代表移動量  $M$  を求める手法について述べる。代表移動量の推定においては、混雑時にも適用できるように、局所的な特徴点の対応付けのみを用いて群衆の代表移動量を推定していく。

##### 4.2 群衆の代表移動量の推定

斜め下向きカメラで複数の人物を撮影する場合、仮に複数の人物が実空間上で同じ方向、同じ速度で移動していても画像上での速度の見え方にはバリエーションが出てくる。具体的に言えば、よりカメラに近い位置にある頭部の動きは画像上では相対的に大きな動きとなり、足付近は小さい動きとなる。逆に画像上の同じ画素位置で同じ動きが観測された場合でも、その画素が人物の身体のどの高さを見ているかによって、対応する実空間速度は変わることになる（図9参照）。以下では、このような画像上の動きの見えたと実空間速度の対応の任意性を解消しながら代表移動量  $M$  を算出する方法について説明する。

まず、時間的に近い2枚のフレーム画像について、それぞれのフレーム画像において特徴点を抽出し、2枚の間で特徴点同士のマッチングを行うことで、局所的な動き量（フレーム間で対応付けられた特徴点对）を算出する。この際、人物以外の背景部の影響を除外するため、前景領域上の特徴点のみを利用する。次に、画像上で得られた局所的な動き量を実空間上の移動量に変換する。この際、前述したように、画像上の動き量は実空間の移動量に一意に対

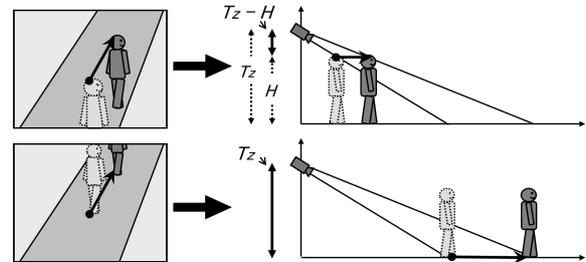


図9 実空間における移動量の任意性

応しないという問題があるが、この問題を解決するため投票原理を用いる。具体的には、画像上の動き量が実空間上の人物の足 ( $z = 0$ ) に起因するとした場合の実空間上の移動量と、実空間上の人物の頭部 ( $z = H$ ) に起因するとした場合の実空間上の移動量をそれぞれ求め、その間の移動量全てに1票投票する。図10の上側の例では3対の特徴点から、それぞれ投票を行った場合を示している。身体の高さの異なる場所から概ね均等に特徴点が得られている場合には、投票値が最大となる移動量を、群衆全体の移動量と考えることができる。

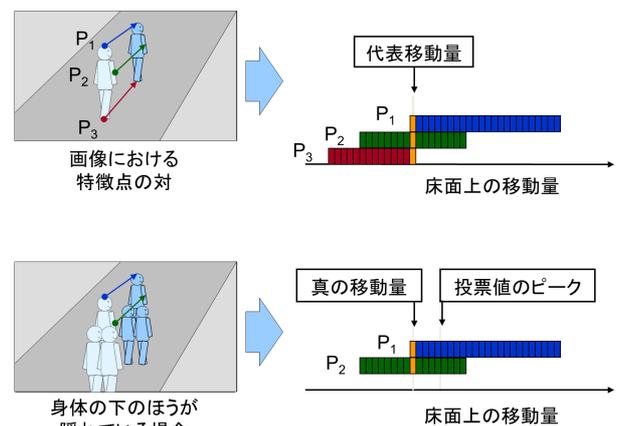


図10 移動量の投票

##### 4.3 代表移動量におけるオクルージョンの影響の補正

上述した投票原理による代表移動量に推定は、オクルージョンが発生している場合には正しい結果が得られない。図10の下側の例ではオクルージョンにより、頭部付近のみから2対の特徴点が得られている場合を示している。この例のように、オクルージョンの発生により、検出される

特徴点の対が身体の上部に偏った場合には、投票値が最大となる移動量は、実際の移動量より大きな値にずれていくことになる。

この投票値のずれを補正するため、前述の重畳度  $R$  を再び利用する。 $R$  は1つの画素に対して何人が重畳しているかを表しており、前章で説明したように前景検出結果から求めることができる。いま、処理対象のフレームにおける重畳度が  $R$  である場合に、前景として検出されている1つの画素について考えて見る。その画素に対応する人物位置の範囲を図11のように  $0 \sim L$  と表す（簡略化のため奥行方向のみ表示）。重畳度が  $R$  ということは、この範囲に  $R$  人がいると期待される。この場合、仮に1人の人物が図11の位置  $l$  に固定されているとした場合、残りの  $R-1$  人全員が、その1人の背後にいる確率  $q(l)$  は、

$$q(l) = \left(\frac{L-l}{L}\right)^{R-1}. \quad (14)$$

と書くことができる。したがって、 $R$  人のうちで最前面に立つ人物の位置が  $(l + \Delta l)$  である確率の確率密度関数  $Q(l)$  は、

$$Q(l) = \frac{q(l)}{\int_0^L q(l) dl}. \quad (15)$$

となる。図11において人物の位置  $l$  と、注目画素と対応する人物上の高さ  $h$  の関係は、

$$h = H \frac{L-l}{L}. \quad (16)$$

と表すことができるので、上記の  $l$  に対する確率密度関数の確率変数を  $h$  に置き換え、重畳度が  $R$  である場合の画像で観測される人物上の高さ  $h$  の期待値（平均値）： $E(h|R)$  を求めると、

$$E(h|R) = \int_0^H h \cdot P(h) dh, \quad (17)$$

$$P(h) = \frac{p(h) dh}{\int_0^H p(h) dh}, \quad (18)$$

$$p(h) = \left(\frac{h}{H}\right)^{R-1}. \quad (19)$$

となり、さらに右辺の積分を実行すると、

$$E(h|R) = \frac{R}{R+1} H. \quad (20)$$

となる。

$R$  が1、つまり混雑していない状況においては上記の期待値は  $H/2$ 、つまり身体の中心となり、一方  $R$  の値が大きくなるにつれて、つまり混雑するにつれて、その期待値は  $H$  に近づいていく。これは、混雑時においては人物の下部が画像上で隠ぺいされるようになり、観測される部位が頭

部付近に偏っていくことと対応している。ここで、図9を見ると、画像上で観測される移動量は、カメラの床面からの高さを  $T_z$  とすると、 $(T_z - h)$  に比例することがわかる。したがって、前述の投票値の最大値を与える移動量 ( $M^\dagger$  とする) を、上記のオクルージョンによる偏りを重畳度  $R$  に基づき補正する式は、

$$M = \frac{T_z - E(h|R)}{T_z - E(h|1)} M^\dagger = \frac{T_z - HR/(R+1)}{T_z - H/2} M^\dagger, \quad (21)$$

となる。ここで  $M$  が補正後の移動量（群衆全体としての代表移動量）となる。そして、この代表移動量  $M$  および式(12)における  $N$  を式(13)に代入することで、所定時間内の通過人数  $P$  を求めることができる。

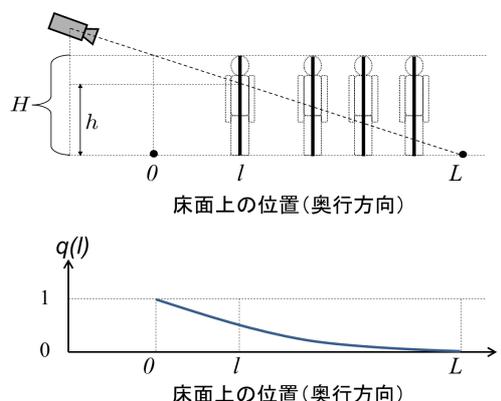


図11 1人が位置  $l$  に固定された場合に他の  $(R-1)$  人が後ろにいる確率;  $q(l)$

## 5. 実験結果

### 5.1 画像内人数の推定結果

前述の画像人数推定手法の有効性を確認するために実施した実験の結果を図12および図13示す。図12は、大学構内で撮影した評価用映像を用いたもので、本手法による推定結果と真値（目視）とは良好に一致していると言える。一方、図13は、首都圏のターミナル駅での実験結果であり、目視カウント結果と比べ、推定結果がやや小さい値となっているが、概ね一致していると言える。

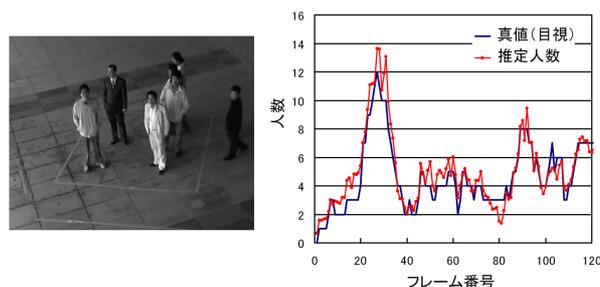


図12 画像内人数の推定結果（大学構内：評価用映像）

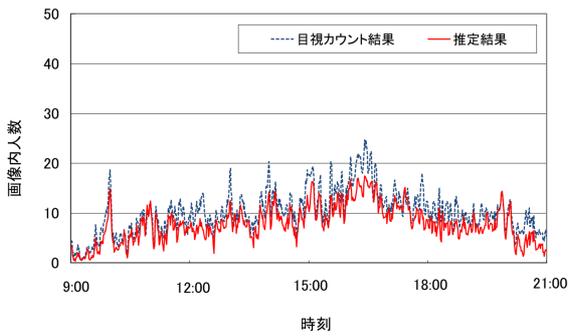


図 13 画像内人数の推定結果 (ターミナル駅の通路)

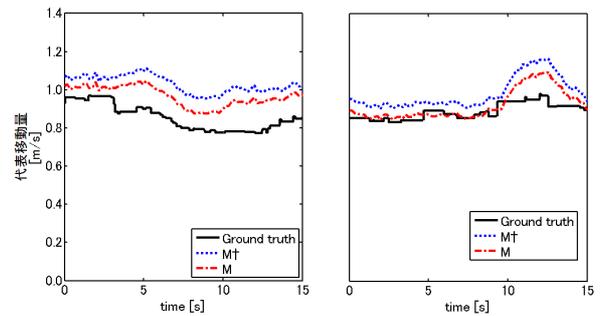


図 15 代表移動量の推定結果

## 5.2 通過人数の推定結果

前述の代表移動量推定および通過人数推定手法の有効性を確認するために行った実験の結果を以下に示す. 図 14 は実験に用いた画像系列の例である. 図 15 は, 前述した投票原理に基づく代表移動量の推定結果である. なお, 図 15 では時間間隔を 1 秒に換算して表記しており, 実空間における群衆の速度 [km/s] に対応する. これらを見ると, 代表移動量  $M$  を概ね正しく推定できているものの, 20%程度の誤差も発生している. この原因としては, 本手法では特徴点が身体の様々な部位から均等に検出されることを前提しているものの, 実際に検出される特徴点には偏りがあることが考えられる. 図 16 はトラッキングと本手法による推定結果を比較したものである. オクルージョンのほとんど発生しない閑散時と, 常態的にオクルージョンが発生している混雑時とに分けて, 誤差の大きさを比較している. トラッキングは閑散時の精度に優れるものの, 混雑時の誤差が非常に大きくなる. 一方, 本手法は, 閑散時に一定の誤差があるものの, 混雑時においてはむしろ精度が向上している. これは, 混雑時においては群衆を構成する個人個人の動きが揃うことで, 群衆全体としての流動状態が安定することによると考えられる.



図 14 実験に用いた映像の様子

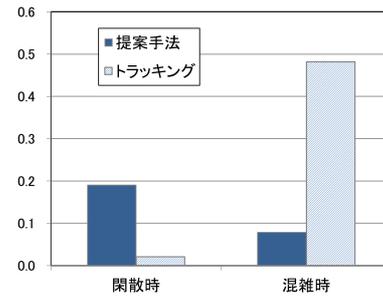


図 16 通過人数推定の誤差率:  $|推定値 - 真値| / 真値$

提としている. 具体的には, カメラの床面からの高さ, カメラの姿勢 (3 軸の回転) および焦点距離 (実空間とカメラ座標の拡大率) 等のパラメータを必要とする.

較正方法としては, 当初は, 印刷したドットパターンを床面に配置する方法や, 立体構造物を用いたもの [14] を適用していた. しかし実際の利用場所の条件は様々であり, 較正用の被写体を持ち込むことや現場に配置することが難しいケースが多く存在した. 特にその場所に通行者がいる状態で較正を行う場合には, 通行者に迷惑をかけずに作業を遂行することは困難であった.

この問題を解決するため, 人物をはじめとする鉛直被写体により較正を行う較正ツールを開発した. このツールでは, 鉛直方向の高さが既知である被写体の画像を複数枚取得し, その上端, 下端を画面上で指定するだけで較正が完了する. 高さが既知の鉛直な被写体は, 身長が既知の人物でも良いし, 巻尺を鉛直に垂らす等でも代用可能であり, 現場において通行者に迷惑をかけることなくスムーズに較正作業を行うことが可能となった.

## 6. 実用化における課題

本技術は, 評価実験および現場での実証実験を行いながら適宜改良を進めてきた. 本章では, その中で見えてきた実用上の課題およびそれに対する技術改良等について紹介する.

### 6.1 効率的なキャリブレーションの実現

本技術はカメラが床面に対して較正されていることを前

### 6.2 真下向きカメラへの適応

本手法は斜め下向きのカメラで広範囲を撮影する場合を想定しているが, 実際には現場におけるカメラ設置位置の制約から, 真下もしくは真下に近いエリアの計測が必要となるケースがあった. 本手法では人物のシルエットを直立する矩形で近似しているが, このモデルにおいては真下付近では矩形が見えなくなり, モデルと実際の画像との間に大きなずれが生じるため, 真下付近には適用できないという問題があった. この問題を解決するため, シルエットの

モデルを、床面上に傾斜配置した矩形に変更した。矩形を傾斜させたモデルにおいては、真下付近でもカメラから見える状態となり、また上述した計算と同様に解析的に解くことができる。これにより実際の画像上の人物の見え方をより安定的に近似できるようになった（詳細は文献 [16] 参照）。

### 6.3 トラッキングとのハイブリッド化

図 16 からわかるように、閑散時には本手法は従来のトラッキング手法と比較して精度が低くなる。しかし、デジタルサイネージの広告効果測定の用途においては、計測場所に現れる人が必ずしも多くない場所も多く存在する。また人が多い場所でも、時間帯によっては閑散としていることがある。閑散時の精度を確保することは実用上重要であると考え、トラッキングと前述の通過人数計測をハイブリッド化した（詳細は文献 [17] 参照）。画像内人数の推定結果に基づき、トラッキングの処理結果と、本手法による通過人数推定結果に所定の重みをかけた加重平均により、通過人数を算出しており、画像内人数が少ない場合にはトラッキングが、画像内人数が多い場合には本手法による通過人数計測がそれぞれ重視される形としている。これにより、閑散時から混雑時まで安定した精度での計測が可能となった。

### 6.4 プライバシー保護の問題

カメラで撮影されることについては、不快感を覚える方も多く、個人情報、パーソナルデータの保護の観点から注意が必要な問題となっている。現状では、画像データは一切残せない仕様とし、またネットワークを介した画像転送は行っておらず、いわば現場で動作する人数計測センサとして運用する形をとっている。画像撮影および画像データの扱いについては、各種法令やガイドラインの整備状況、世論の動向等を踏まえながら慎重に判断していく必要がある。

## 7. まとめ

デジタルサイネージの広告効果測定や安全管理支援への適用を目指し、混雑状況においても安定に動作する人数推定技術を紹介した。本技術では、個人個人の検出を行うことなく、画像上の局所的な観測量を、人物および群衆の幾何的なモデルに従い実空間上の物理量に対応付けることで、画像内人数および通過人数の計測を実現している。また技術の紹介に加え、実証実験等において明らかになった運用上の課題や、それに対する技術改良など、実用化の観点からも考察を行った。今後、本技術は、プライバシーの問題に配慮しながらクラウドサービスとしても展開していく予定である。

## 参考文献

- [1] 屋外広告調査フォーラム: *DEC: Daily Effective Circulation*, <<http://www.okugai-forum.jp/>>
- [2] Digital Place-based Advertising Association: *Audience Metrics Guidelines*, <<http://www.dp-aa.org/>>
- [3] デジタルサイネージコンソーシアム (指標部会), <<http://www.digital-signage.jp/>>
- [4] Sheng-Fuu Lin, Jaw-Yeh Chen and Hung-Xin Chao: *Estimation of Number of People in Crowded Scenes using Perspective Transformation*, IEEE Transactions on Man and Cybernetics, Vol.31(PartA), issue 6, pp.645-654 (2001).
- [5] 山内, 山下, 藤吉: 画像からの統計的学習法に基づく人検出, 信学論 D, Vol.j-96D, No.9, pp.2017-2040 (2013).
- [6] T. Zhao and R. Nevatia: *Tracking Multiple Humans in Crowded Environments*, Proc. CVPR2004, pp.406-413 (2004).
- [7] V. Rabaud and S. Belongie: *Counting Crowded Moving Objects*, Proc. CVPR2006, pp.705-711 (2006).
- [8] G. J. Brostow and R. Cipolla: *Unsupervised Bayesian Detection of Independent Motion in Crowds*, Proc. CVPR2006, pp.594-601 (2006).
- [9] M. Rodriguez, I. Laptev, J. Sivic, and J.-Y. Audibert: *Density-aware person detection and tracking in crowds*, ICCV2011, pp. 2423-2430 (2011).
- [10] A.N. Marana, L.F. Costa, R.A. Lotufo, S.A. Velastin: *On the Efficacy of Texture Analysis for Crowd Monitoring*, Proc. International Symposium on Computer Graphics, Image Processing, Vol.6, pp.3521-3524 (1998).
- [11] S.Y. Cho, T.W.S. Chow, C.T. Leung: *A Neural-Based Crowd Estimation by Hybrid Global Learning Algorithm*, IEEE Transactions on Systems, Man and Cybernetics, Part B, Vol.29, issue 4, pp.535-541 (1999).
- [12] S. Fujisawa, G. Hasegawa, Y. Taniguchi, H. Nakano: *Pedestrian Counting in Video Sequences using Optical Flow Clustering*, ACA'12 Proceedings of the 11th international conference on Applications of Electrical and Computer Engineering, pp. 51-56 (2012).
- [13] Z. Zhang, M. Li: *Crowd density estimation based on statistical analysis of local intra-crowd motions for public area surveillance*, Optical Engineering, 51(4), 047204, pp.1-13 (2012).
- [14] I. Miyagawa, H. Arai, and H. Koike: *Simple Camera Calibration From a Single Image Using Five Points on Two Orthogonal 1-D Objects*, IEEE Transactions on Image Processing, 19(6), pp.1528-1538 (2010).
- [15] H. Arai, I. Miyagawa, H. Koike, and M. Haseyama: *Estimating Number of People Using Calibrated Monocular Camera Based on Geometrical Analysis of Surface Area*, IEICE Trans. Fundamentals, Vol. E92-A, No. 8, pp. 1932-1938 (2009).
- [16] H. Arai, N. Ito, and Y. Taniguchi: *Surface Area Analysis for People Number Estimation*, Proc. of VISAPP2014 (accepted) (2014).
- [17] H. Arai, I. Igarashi, N. Ito, and Y. Taniguchi: *Estimating the Number of Pedestrians based on People Flow Analysis and Tracking*, IEEE Transactions on Image Electronics and Visual Computing, Vol.1, No.1, pp.19-27 (2013).