原著論文

映像からの人数推定における 遺伝的アルゴリズムを使ったカメラパラメータの調整†

宮川 勲*・新井 啓之*・小池 秀樹*

荷重値積算による人数推定は、処理コストが低くリアルタイムでの人数計測が可能であるため、画像中の人数を計測する方法として有用である。ただし、その荷重値はカメラパラメータに基づいて算出されるため、事前のカメラキャリブレーション作業が必要となっていた。本論文では、荷重値積算による人数推定の枠組みの中で、荷重値算出に必要なカメラパラメータを遺伝的アルゴリズムにより適応的に調整する方法を提案する。実験では、計算機シミュレーションを使って提案方法の性能を評価し、実画像実験では目視による人数との比較結果を示し、提案方法の有効性を確認する。

キーワード：画像計測，人数推定，定点観測カメラ，遺伝的アルゴリズム，背景差分，カメラキャリブレーション

1. はじめに

街頭や店頭などにおいて薄型ディスプレイ上に様々な情報を提示するデジタルサイネージが急速な勢いで普及しつつある。今後、デジタルサイネージが次世代の電子広告メディアとして幅広く活用されていくためには、テレビ視聴率のように広告効果[12]を客観的に示す指標が必要となり、ディスプレイ前にいる人の数はその基本的な指標の一つになると言われている。特定領域内の人数を把握するには様々なセンシング技術の応用が考えられるが、コスト面を考慮するとカメラを使った画像計測が現実的である。

定点観測カメラで得た映像から人数や群集の密集度を計測する先行技術として、人の顔や上半身などの特定部位を検出し人数を計測する方法[5]、時空間ボリューム内で画像中の特徴点をクラスタリングし、そのクラスタ数を人数とする方法[7]、テクスチャ解析によって得た画像特徴量と群集の密集度を関連付け、その事前学習を使って人数を算出する方法[3]、並びに、画素から人数へ換算する荷重値を用い、前景画素の荷重値を積算して人数を推定する方法[1]などが挙げられる。このような従来技術の中で、少人数から多人数までの人数を安定して計測できる点と、実環境では処理コストが低くリアルタイムでの人数推定が要求される点を考慮すると、荷重値積算による人数推定法

[1]が有用であると考えられる。ただし、その荷重値はカメラパラメータに基づいて算出されるため、このパラメータをカメラキャリブレーションによって事前に求めておく必要がある。

カメラキャリブレーションとは、校正対象のカメラで何らかの被写体を撮影し、その2次元画像から投影モデルに従って内部パラメータ（焦点距離、画像中心、レンズ歪など）、外部パラメータ（姿勢、位置）を推定する作業である。これまでに、多種多様なアプローチによるカメラキャリブレーション[13]が提案されており、実空間においてスケール倍の不定性が無いカメラパラメータを得るには既知の参照物体^{*1}を使ったカメラキャリブレーションが常套手段である。ところが、カメラ配置や設置条件によっては参照物体を使ったカメラキャリブレーションが行えるとは限らない。例えば、混雑する公共の場所によっては、キャリブレーション作業のために人の進入を禁止し、参照物体を撮影することが許されない場合がある。このため、カメラキャリブレーションに必要な撮影時間と作業時間が確保できないかもしれない。また、十分な精度のカメラパラメータを得るには、多くの参照点を観測することが不可欠であり、対象カメラに対してどのように参照物体を空間中に配置すべきかを決め、それに適した大きさの参照物体を設計・準備しなくてはならない。

† Camera Parameter Adjustment using Genetic Algorithm for Estimating the Number of People from Image Sequence
Isao MIYAGAWA, Hiroyuki ARAI and Hideki KOIKE

* 日本電信電話株式会社NTTサイバースペース研究所
NTT Cyber Space Laboratories, NTT Corporation

※1 本論文での参照物体とは、カメラにより観測可能な目印（参照点）を有する物体を指す。例として、2次元参照物体には白黒の市松模様を描いたボードが利用され、各格子点が参照点に使われる。

実運用において、参照物体を使ったカメラキャリブレーションが困難な場合、我々は別の手段で対処しなければならない。既存のカメラシステムとの整合性をとるには、観測した映像を利用することが望ましい。そこで、参照物体を用意するのではなく、カメラを通して“人”を撮影することができるので、荷重値積算による人数推定[1]の枠組みの中でカメラパラメータを獲得する。我々は、そのカメラパラメータを適応的に調整するアルゴリズムとして、生物が環境に合わせて進化する過程をモデル化した遺伝的アルゴリズム[6]に着目する。これまでに、遺伝的アルゴリズムを使ったカメラキャリブレーション[4]が提案されているが、カメラパラメータ獲得には空間中の3次元点(control point)が利用されている。また、カメラキャリブレーションにおいて、カメラパラメータの最適化にはバンドル調整(bundle adjustment)[9]が有効であると報告されている[13]。最適化アルゴリズムにLevenberg-Marquardt法などが利用されるが、パラメータ最適化計算を有効に機能させるには“良い初期値”が必要である。参照物体を使うことを前提としない場合、従来の方法でこの初期値を獲得することは容易ではない。これに対し、遺伝的アルゴリズムは“良い初期値”を必要とせず、与えられた環境に最も適応するパラメータを獲得するという特長がある。我々の提案する適応的なパラメータ調整とは、参照点を一切使わず、人数推定に十分なカメラパラメータを観測画像を利用して自動的に調整するものである。もし、それぞれの環境に応じて人数推定に使うパラメータを適応的に調整することが可能となれば、その結果として事前作業のカメラキャリブレーションは不要となるため、カメラを使った人数計測システムの運用面で大きな貢献ができる。

本論文は以下の構成としている。まず、2節では、カメラによって実空間の3次元点と画像上の投影点を結び付ける透視投影モデルを記述し、本研究で扱うカメラパラメータを定義する。次に、3節では、画素から人数へ換算するための荷重値を導出し、その荷重値の積算による人数推定[1]を説明する。続いて、4節では、その人数推定の枠組みにおいて、遺伝的アルゴリズムを使った適応的なカメラパラメータの調整方法を提案する。5節では、計算機シミュレーションを使って提案方法の性能を評価し、実画像実験では目視計測による人数との比較結果を示し、提案方法の有効性を確認する。最後に、6節において、本研究のまとめと展望を述べる。

2. 投影モデルとカメラパラメータ

荷重値積算による人数推定法[1]では、その荷重値はカメラパラメータに基づいて算出される。以降の説明を円滑にするため、本節において、カメラパラメータを定義しておく。

一般的に、定点観測カメラを使って人数を計測する場合、そのカメラは人の身長より高い位置から俯瞰する姿勢で設置される。図1はカメラと被写体(人物)の模式図を描いたものである。実空間の3次元座標を表現するため、任意に世界座標系 $X_w Y_w Z_w$ を設定する。カメラは固定されており、カメラを通して観測できる画像上の投影点はピンホールカメラと呼ばれる透視投影モデルに従うと仮定する。すなわち、世界座標系上の3次元点を $P_j = (X_j, Y_j, Z_j)$ 、カメラ視点の位置を $C = (Tx, Ty, Tz)$ とすると、画像上で観測される点 P_j の2次元座標 $p_j = (x_j, y_j)$ は、

$$x_j = f \frac{R_{11}(X_j - Tx) + R_{12}(Y_j - Ty) + R_{13}(Z_j - Tz)}{R_{31}(X_j - Tx) + R_{32}(Y_j - Ty) + R_{33}(Z_j - Tz)}, \quad (1)$$

$$y_j = f \frac{R_{21}(X_j - Tx) + R_{22}(Y_j - Ty) + R_{23}(Z_j - Tz)}{R_{31}(X_j - Tx) + R_{32}(Y_j - Ty) + R_{33}(Z_j - Tz)} \quad (2)$$

図1 カメラと被写体(人)の配置関係

の射影関係で結びつけることができる^{※2}。ここで、 f は焦点距離であり、 $R_{11}, R_{12}, \dots, R_{33}$ は 3×3 の回転行列の要素に対応する。 ϕ, ω, θ をそれぞれカメラ座標系 XYZ における X 軸、 Y 軸、 Z 軸周りの回転角とすると、各回転要素は、

$$\left\{ \begin{array}{l} R_{11} = \cos(\theta) \cos(\omega), \\ R_{12} = \cos(\theta) \sin(\omega) \sin(\phi) + \sin(\theta) \cos(\phi), \\ R_{13} = -\cos(\theta) \sin(\omega) \cos(\phi) + \sin(\theta) \sin(\phi), \\ R_{21} = -\sin(\theta) \cos(\omega), \\ R_{22} = -\sin(\theta) \sin(\omega) \sin(\phi) + \cos(\theta) \cos(\phi), \\ R_{23} = \sin(\theta) \sin(\omega) \cos(\phi) + \cos(\theta) \sin(\phi), \\ R_{31} = \sin(\omega), \\ R_{32} = -\cos(\omega) \sin(\phi), \\ R_{33} = \cos(\omega) \cos(\phi) \end{array} \right. \quad (3)$$

で与えられる。式(1), (2)で定義される投影モデルにおいて、本研究で扱うカメラパラメータは焦点距離 f , XYZ 軸周りの回転角 ϕ , ω , θ , 並びに視点位置 (T_x, T_y, T_z) である。

3. 荷重値積算による人数推定

本節では、前節で定義したカメラパラメータを使って画素 p_j の荷重値を導出し、荷重値積算による人数推定法を記述する。

カメラを通して被写体の像を構成する画素の集合 p_j , $j=1, 2, \dots, P$ が得られるとする。その画素の集合 p_j , $j=1, 2, \dots, P$ から人数を推定するには、各画素へ投影される実空間中の人の体積を推量することが直接的である。この発想から、画像座標 (x_j, y_j) が式 (1), (2) で表される投影モデルに従うと仮定した上で、各画素に投影される視体積 (visual hull) を考える。視体積とは視点 C から外界へ向かう光束 (錐体) の体積を指す。

観測された画像が正方格子の画素から構成されると仮定し，画素 $p_j = (x_j, y_j)$ の 4 近傍のサブ画像座標：

$$\left\{ \begin{array}{l} \mathbf{p}_j^{(n)} = (x_j^{(n)}, y_j^{(n)}), \quad (n = 1, 2, 3, 4), \\ x_j^{(1)} = x_j + 1/2, \quad y_j^{(1)} = y_j - 1/2, \\ x_j^{(2)} = x_j + 1/2, \quad y_j^{(2)} = y_j + 1/2, \\ x_j^{(3)} = x_j - 1/2, \quad y_j^{(3)} = y_j + 1/2, \\ x_j^{(4)} = x_j - 1/2, \quad y_j^{(4)} = y_j - 1/2 \end{array} \right. \quad (4)$$

※2 より詳細な投影モデルでは、画像中心、アスペクト比、スキュー値、並びに、放射型のレンズ歪などが含まれるが、本研究では比較的低解像度の画像を使用するため、そのような内部パラメータを考慮していない。

に着目する。これら 4 近傍のサブ画素で囲まれた微小領域へ射影される視体積を図 2 に図示する。ここで、人の高さを h (一定) とし、図示した視体積のうち、平面 $z_w = h$ と平面 $z_w = 0$ で囲まれた四角錐台を実質的な視体積 $V(x_i, y_i)$ と考える。

4近傍のサブ画素 $p_j^{(n)}$, $n=1\sim 4$ へ射影される平面 $Z_w=h$ 上の点をそれぞれ $A_j^{(n)}$, 平面 $Z_w=0$ 上の点をそれぞれ $B_j^{(n)}$, 並びに, 画素 p_j へ投影される平面 $Z_w=0$ 上の点を Q_j とする. 式(1), (2)において, $Z_j=h$ とした場合,

$$\begin{bmatrix} X_j \\ Y_j \end{bmatrix} = \begin{bmatrix} m_{j1} & m_{j2} \\ n_{j1} & n_{j2} \end{bmatrix}^{-1} \begin{bmatrix} m_{j1} & m_{j2} & m_{j3} \\ n_{j1} & n_{j2} & n_{j3} \end{bmatrix} t, \quad (5)$$

$$\mathbf{t} = [Tx, Ty, Tz - h]^T, \quad (6)$$

$$m_{jq} = \frac{x_j}{f} R_{3q} - R_{1q}, \quad (q = 1, 2, 3), \quad (7)$$

$$n_{jq} = \frac{y_j}{f} R_{3q} - R_{2q}, \quad (q = 1, 2, 3) \quad (8)$$

と記述されるので、カメラパラメータが与えられると対象画素の画像座標 p_j 、並びに、その4近傍のサブ画素の画像座標 $p_j^{(n)}$ から、 $Q_j = (X_j, Y_j, 0)$ 、 $A_j^{(n)} = (X_j^{(n)}, Y_j^{(n)}, h)$ がそれぞれ求められる ($B_j^{(n)}$ の算出には $h=0$ とした式(6)を式(5)に代入して求める)。 $S_a(x_j, y_j)$ 、 $S_b(x_j, y_j)$ をそれぞれ $A_j^{(n)}$ で形成される四角形の面積、 $B_j^{(n)}$ で形成される四角形の面積とすると、四角錐台の体積 $V(x_i, y_i)$ は、オベリスク公式：

$$V(x_j, y_j) = \frac{\hbar}{3} \times \left\{ S_a(x_j, y_j) + S_b(x_j, y_j) + \sqrt{S_a(x_j, y_j)S_b(x_j, y_j)} \right\} \quad (9)$$

図2 1画素の視体積

により算出することができる^{※3}。

式(9)で得られる視体積 $V(x_j, y_j)$ は画素 p_j の4近傍 $p_j^{(n)}$ で囲まれた微小な画像領域内へ射影されるが、 $V(x_j, y_j)$ は視点からの見かけ上の視体積であって、人の位置は考慮されていない。もし、人の位置を特定することができるならば、視体積 $V(x_j, y_j)$ から実際の人の部分だけを抜き取り、その体積を平均的な人の体積で割れば人数へ換算するための係数が得られるだろう。しかし、単眼のカメラであるが故に、人の位置を正確に推定することは容易ではない。

そこで、人の位置を特定するのではなく、以下の“簡易な見えモデル”を導入して、人数を推定するアプローチをとる。

- 人の表面積はどの方向からでも面積 S_0 とする。
(人の形状を円柱や楕円体で近似する)
- 人の位置は平面 $Z_w = 0$ 上で一様分布に従う。
(人はエリア内を偏らず均等に分散している)

この“簡易な見えモデル”の導入は、 $Z_w = 0$ 上を一様分布する面積 S_0 の板を単眼カメラで観測することを意味する。我々は、このモデルを使って、視体積 $V(x_j, y_j)$ から人数を推量する。

図3は、図2の Y' 軸(平面 $Z_w = 0$ 上での視線方向)に垂直な方向から見た射影図を示しており、点 $A_j^{(n)}$ 、 $n = 1 \sim 4$ のうち視点 C に最も近い点を A_j 、点 $B_j^{(n)}$ 、 $n = 1 \sim 4$ のうち視点 C から最も遠い点を B_j として描

図3 視線方向に沿った各断面

※3 $V(x_j, y_j) = \frac{S_0(x_j, y_j)}{3} \times \frac{T_z^3 - (T_z - h)^3}{T_z^2}$ を使って計算することもできる。

※4 図2に示すように視点 C を平面 $Z_w = 0$ 上へ下ろした垂線の足 $C' = (Tx, Ty, 0)$ から Q_j への単位ベクトルを $\mathbf{u}_j$ 、 A_j を平面 $Z_w = 0$ 上へ下ろした垂線の足 A_j' から B_j へのベクトルを $\mathbf{v}_j$ とすると、長さ $L(x_j, y_j)$ は $\mathbf{u}_j$ と $\mathbf{v}_j$ の内積で求められる。

かれている。読者は、視体積 $V(x_j, y_j)$ を視線方向に沿って面積 S_0 の板でスライスすることを想像してほしい。すなわち、視線方向に沿って点 A_j で交差し始め、体積 $V(x_j, y_j)$ を順番にスライスして、最後に点 B_j で交差を終える。 $\Delta S_j^{(k)}$ は添え字 k で示した位置での断面積である。

上記の仮定から、人を面積 S_0 の板モデルと考えると、面積比 $\Delta S_j^{(k)} / S_0 (< 1)$ は、添え字 k で示した位置での“人数”と解釈できる。面積比 $\Delta S_j^{(k)} / S_0$ が十分に小さければ、この面積比は視線方向に対してのみ変化する。面積 S_0 の板は一様に分布すると仮定しているので、視線方向に対する視体積 $V(x_j, y_j)$ の長さを $L(x_j, y_j)$ ^{※4} とすると、添え字 k の位置での“見えモデル”の存在確率 $p_j^{(k)}$ は全て $1 / L(x_j, y_j)$ となる。よって、面積比 $\Delta S_j^{(k)} / S_0$ の期待値：

$$\begin{aligned} \text{avg}\left(\frac{\Delta S_j^{(k)}}{S_0}\right) &= \int_k \frac{\Delta S_j^{(k)}}{S_0} \times p_j^{(k)} \\ &= \frac{1}{S_0 L(x_j, y_j)} \left(\int_k \Delta S_j^{(k)} \right) \\ &= \frac{V(x_j, y_j)}{S_0 L(x_j, y_j)} = w(x_j, y_j) \end{aligned} \quad (10)$$

が画素 p_j から推量できる人数になる。

人物像として P 個の画素 p_j 、 $j = 1, 2, \dots, P$ が抽出されているので、各画素 p_j に対応する $w(x_j, y_j)$ を積算すれば、見かけ上の人数：

$$N = \sum_{j=1}^P w(x_j, y_j) \quad (11)$$

が得られる。このように、 $w(x_j, y_j)$ を積算すれば人数が計測できるという意味から、文献[1]では $w(x_j, y_j)$ は画素 p_j の“荷重値”と呼ばれている。

$w(x_j, y_j)$ は、カメラの焦点距離 f 、カメラ座標軸での回転角： ϕ 、 ω 、 θ 、世界座標系上での視点位置 (Tx, Ty, Tz) 、並びに、“簡易な見えモデル”のパラメータ： h 、 S_0 によって計算できる物理量である。移動によって人の位置が変わったとしても、人の像として観測される画素 p_j の画像座標 (x_j, y_j) が変わるだけであり、荷重値 $w(x_j, y_j)$ は時間的に不変である。荷重値 $w(x_j, y_j)$ は画像座標 (x_j, y_j) と一対一で結び付けられているので、事前に全ての画像座標に対して $w(x_j, y_j)$ を求めておけば、各時間で観測された人物像の各画素に対応する $w(x_j, y_j)$ を積算するだけで、その画像中の人数を逐次計測することができる。

4. GAによるカメラパラメータの自動調整

前節では、カメラパラメータが既知であるという前提で、荷重値積算による人数推定方法を記述した。本節では、その枠組みの中で遺伝的アルゴリズムを使ってカメラパラメータを自動的に調整する方法を提案する。

提案方法の処理フローを図4に示す。カメラパラメータの調整に使用する標本画像を M 枚とし、それぞれの画像から人物像として抽出した画素を $p_{ij} = (x_{ij}, y_{ij})$, $i=1, 2, \dots, M$, $j=1, 2, \dots, P_i$ (各標本画像あたりに抽出される前景画素の総数は異なる), 並びに、各標本画像における人数(目標値)を N_i , $i=1, 2, \dots, M$ とする^{*5}。これらのデータが図4の処理フローにおける入力データとなる。

図5に集団モデルと個体内の染色体の構成を示す。個体数を X とし、各個体は、カメラ回転角: ϕ , ω , θ , カメラ視点位置(T_x , T_y , T_z), 並びに、焦点距離 f の7つの染色体から構成される(各染色体の長さを Y ビットすると、個体の遺伝子長は $7Y$ ビットになる)。最初の処理において、各個体の染色体を構成する遺伝子配列をランダムに決定し、初期集団を生成する。次に、各画像から抽出された画素 $p_{ij} = (x_{ij}, y_{ij})$ とその4近傍のサブ画素 $p_{ij}^{(n)}$, $n=1, 2, 3, 4$ に対応する $A_j^{(n)}$, $B_j^{(n)}$, Q_j を式(5)を使って算出し、式(9)により視体積 $V(x_{ij}, y_{ij})$ を求め、式(10)に従って荷重値 $w(x_{ij}, y_{ij})$ を得る。続いて、抽出された画素の荷重値を式(11)に従って積算し、各画像での推定人数 $\hat{N}_i$, $i=1, 2, \dots, M$ を得る。このとき、目標人数と推定人数の間の推定誤差:

$$\Delta N = \sum_{i=1}^M \|N_i - \hat{N}_i\| \quad (12)$$

を、環境にどれだけ適応しているかを評価する“適応度”として採用する。

この適応度に従って全ての個体から“自然選択(淘汰)”を行う。この選択手段にルーレットルールを利用する。集団内の各個体 n の適応度 ΔN の逆数を f_n とし、各個体について選択確率 p_n を

$$p_n = \frac{f_n}{F}, \quad F = \sum_{n=1}^X f_n \quad (13)$$

により得て、その累積確率 q_n を

^{*5} 3節では各画素の違いを表すために添え字 j を使用していたが、提案方法では1枚以上の標本画像を使うため、その添え字として i も使う。

図4 提案方法の処理フロー

図5 集団と個体の構成

$$q_n = \sum_{k=1}^n p_k, \quad (n = 1, 2, \dots, X) \quad (14)$$

により計算する。次の世代に残す個体を決めるために、各個体について乱数 r を発生させ、

$$q_{n-1} \leq r < q_n, \quad (n = 1, 2, \dots, X) \quad (15)$$

が満たされるとき、 n 番目の個体を選択する(ただし、 $q_0 = 0$)。このプロセスを X 回繰り返し、 X 個の個体を選択する。

上記のルーレットルールによって選ばれた各個体に対して乱数 r を発生させ $r < P_c$ (交叉確率) を満たすとき、その個体を交叉させる親として選ぶ。これを X 個の個体に対して繰り返す(選ばれた親の数が奇数のときは、もう一度選択して親の数を偶数にする)。次に、選ばれた順に交叉するペアとし、各ペアの個体を

構成する染色体を順番に取り出し、乱数を発生させて染色体長での交叉位置を決め各染色体について一点交叉を行う。

続いて、突然変異のプロセスでは、この新しい集団に含まれる個体に対し、各遺伝子の先頭ビットから数えた位置において順番に乱数 r を発生させ $r < P_m$ (突然変異確率) が満たされたとき、その遺伝子座のビットを反転させる。

以上の自然選択、交叉、突然変異の操作を世代数 G を超えるまで繰り返し、 G 世代後において式 (12) に最も適応した個体をカメラパラメータとして出力する。

5. 実験結果

本節では、シミュレーション画像を使った実験と実画像を使った実験により、提案方法の性能と有効性を検証する。5.1節では、我々のシミュレーション環境において標本画像を生成し、それぞれの標本画像を使って提案方法によるカメラパラメータの自動調整を試す。そのカメラパラメータに基づき荷重値を算出し、標本画像とは別に準備した評価画像から人数推定の実験を行い、提案方法の精度評価を行う。5.2節では、実画像において、平面射影変換を利用したカメラキャリブレーション法[10, 11] (以降、従来方法と称する) と提案方法によるカメラパラメータを比較する。さらに、各カメラパラメータを使って荷重値を算出し、時系列画像から推定した人数を比較する。以上の実験で得た結果を踏まえて、5.3節では提案方法に関する考察を行う。

本実験で実装した遺伝的アルゴリズムは、単純遺伝的アルゴリズム(SGA: Simple Genetic Algorithm)であり、主なパラメータを表1に示す。また、実環境では処理コストを優先としたパラメータ調整がシステム導入には有利であることを考慮し、以下の実験では3枚の標本画像を使った実験結果を示す。

表1 SGAの主要パラメータ

個体数 X	100
染色体の長さ Y	16 ビット
交叉確率 P_c	0.50
突然変異確率 P_m	0.01
最大世代数 G	1,000

5.1 シミュレーション実験

6 m×6 mの範囲内において N 人の矩形モデル ($h=1.7\text{m}$, $S_0=0.51\text{m}^2$) を一様乱数に従って配置し、配置した全てのモデルを表2に示す外部パラメータに基づき透視投影像を計算し、焦点距離 $f=240$ 画素、解像度 200×200 画素のシミュレーション画像を生成した。図6に、矩形モデルに対するカメラ C_1 とカメラ C_2 の俯瞰姿勢と配置関係を模式的に描いた。カメラ C_1 は5.2節で使用するカメラパラメータに近い値に設定されており、カメラ C_2 はカメラ C_1 より矩形モデルをやや高い位置から、やや急な傾斜角度で俯瞰する姿勢となっている。

図7、図9、図11はそれぞれ5人、10人、15人を配置してカメラ C_1 で得た標本画像である(グループA)。表3は各標本画像から提案方法によって得たカメラパラメータである。全サンプルを見ると、カメラパラメータに統一性を確認できないが、5人、10人、15人配置のサンプル別に比較すると、5人配置のサンプルにおいてカメラ回転角 ϕ は真値に近く、そのばらつきも小さい。カメラキャリブレーションの観点で評価すると、このシミュレーション実験では真値に近いカメラパラメータが得られていない。しかし、本研究の目的は、荷重値積算による人数推定をカメラキャリブレーションを行った場合と遜色無く提案方法により

表2 外部パラメータ

Camera	Rotation [deg]			Translation [m]		
	ϕ	ω	θ	Tx	Ty	Tz
C_1	-55.0	0	0	0	9.0	6.0
C_2	-40.0	0	0	0	7.0	8.0

図6 俯瞰姿勢と配置関係

表3 カメラパラメータの比較

No.	[deg]			[m]			[pixel]
	ϕ	ω	θ	Tx	Ty	Tz	f
A1	-52.3	-9.6	-9.7	3.6	11.8	10.2	393.1
A2	-54.4	-8.2	-8.0	-2.3	5.3	5.9	254.7
A3	-42.3	9.6	1.6	-2.8	14.7	9.2	338.5
A4	-51.2	-5.8	6.4	-0.7	13.8	6.4	267.8
A5	-54.2	-9.2	-4.8	4.4	5.9	5.1	208.9
A6	-59.6	0.6	-0.8	4.9	7.4	5.3	239.0
A7	-58.4	-3.7	4.0	0.9	8.3	5.8	261.6
A8	-61.2	10.0	-0.2	4.2	8.2	5.9	300.9
A9	-46.1	-8.9	-2.6	-4.6	15.0	6.7	247.4
A10	-66.8	-7.5	-7.9	-0.8	12.1	5.1	311.4
A11	-40.2	4.5	3.9	-4.0	6.5	11.3	387.8
A12	-41.9	9.2	4.8	2.9	10.3	9.7	342.6
A13	-55.2	5.7	-9.0	4.1	13.6	5.8	247.0
A14	-40.7	-6.3	-7.2	-1.9	10.6	10.9	386.2
A15	-49.6	5.2	6.4	3.4	8.5	5.4	213.2

実現することである。これを検証するため、提案方法に使った標本画像と分けて、別途、5、10、15、20、25人を一様乱数に従って配置したシミュレーション画像(矩形モデルの位置もそれぞれ異なる)をそれぞれ100枚ずつ“評価画像”として準備した。表3のカメラパラメータに基づき荷重値を生成し、各人数ごとに用意した評価画像から人数を計測した。それぞれ100枚の画像から推定した人数の平均値を図8、図10、図12に、評価画像における各人数の絶対平均誤差を表4にまとめた。“Camera Parameter”で示されている赤の曲線、並びに、表4のサンプルNo.A0は、表2に示したカメラ C_1 のパラメータに基づいた荷重値積算による推定人数である。推定誤差を四捨五入し大まかに表すと、サンプルNo.A0では約1～2人の推定誤差となっている。これに対して、サンプルNo.A1～No.A5(5人の標本画像)では約1～7人の推定誤差、サンプルNo.A6～No.A10(10人の標本画像)では約1～6人の推定誤差、サンプルNo.A11～No.A15(15人の標本画像)では約1～6人の推定誤差という結果がそれぞれ得られた。サンプルNo.A0では3人以上の人数誤差が無いので、推定誤差が3人以上になる観点で評価すると、5人配置より10人、15人配置の標本画像において推定人数の誤差がやや小さい。

次に、カメラ C_2 に対するシミュレーション実験を示す。図13、図15、図17はそれぞれ5人、10人、15人を配置してカメラ C_2 で得た標本画像(グループB)であり、表5は各標本画像から提案方法によって得たカメラパラメータである。サンプルNo.B0は、表2に示したカメラ C_2 のパラメータに基づいた荷重値積算による推定人数である。このシミュレーション実験でも評価画像を別途生成し、各人数ごとに用意した評価画像から人数を計測した。推定人数の平均値を図14、

表4 人数推定誤差の比較(単位:人)

No.	N=5	N=10	N=15	N=20	N=25
A0	0.72	0.86	0.79	0.92	2.15
A1	0.76	0.81	0.94	1.01	2.13
A2	0.34	1.09	2.44	3.86	6.03
A3	0.39	1.08	2.43	3.78	5.93
A4	0.64	1.73	3.34	5.02	7.40
A5	0.45	0.51	1.14	2.04	3.92
A6	0.74	0.96	0.83	0.96	2.11
A7	0.44	0.50	0.99	1.84	3.67
A8	0.30	0.59	1.64	2.91	4.92
A9	0.30	0.92	2.20	3.51	5.62
A10	0.50	0.67	1.30	2.30	4.23
A11	0.54	0.57	1.17	1.76	3.45
A12	0.38	0.66	1.70	2.77	4.74
A13	0.33	0.49	1.46	2.60	4.56
A14	0.39	0.71	1.82	2.90	4.89
A15	0.41	1.25	2.66	4.18	6.42

表5 カメラパラメータの比較

No.	[deg]			[m]			[pixel]
	ϕ	ω	θ	Tx	Ty	Tz	f
B1	-34.3	7.8	7.1	-3.7	9.5	9.5	299.1
B2	-58.7	5.8	2.8	0.2	8.4	9.2	384.6
B3	-59.0	-8.0	-9.2	4.0	9.9	9.2	384.2
B4	-32.0	-9.9	-8.8	5.0	6.2	8.0	246.5
B5	-44.2	-10.0	-6.9	4.7	9.3	9.8	333.7
B6	-59.5	10.0	9.0	-3.2	8.0	5.9	253.4
B7	-54.4	3.2	-6.5	1.2	6.7	5.3	208.8
B8	-59.0	8.3	5.5	-4.5	6.7	6.7	271.3
B9	-31.0	7.9	-4.3	1.7	7.7	7.5	224.7
B10	-67.8	6.8	1.0	0.3	9.2	6.3	379.4
B11	-63.6	7.7	3.4	0.8	9.8	7.6	390.3
B12	-34.1	-2.6	-2.1	1.4	8.0	9.7	298.8
B13	-52.1	10.0	5.0	-1.9	7.6	6.7	245.2
B14	-43.3	-8.0	-6.7	4.2	8.9	6.4	201.2
B15	-43.6	-0.4	-2.9	0.4	5.0	9.1	302.6

図16、図18に、評価画像における各人数の絶対平均誤差を表6にまとめた。サンプルNo.B0では各人数に対して約1人の推定誤差であった。これに対して、サンプルNo.B1～No.B5(5人の標本画像)では約1～4人の推定誤差、サンプルNo.B6～No.B10(10人の標本画像)では約1～3人の推定誤差、サンプルNo.B11～No.B15(15人の標本画像)では約1～4人の推定誤差という結果がそれぞれ得られた。グループAと比べると、グループBでは各サンプルの推定誤差が低下していることが確認できる。特に、サンプルNo.B1～No.B5での調整では10人、15人配置での調整とほぼ同

図7 シミュレーション画像(5人配置)

図8 推定人数の結果(5人配置)

図9 シミュレーション画像(10人配置)

図10 推定人数の結果(10人配置)

図11 シミュレーション画像(15人配置)

図12 推定人数の結果(15人配置)

図13 シミュレーション画像(5人配置)

図14 推定人数の結果(5人配置)

図15 シミュレーション画像(10人配置)

図16 推定人数の結果(10人配置)

図17 シミュレーション画像(15人配置)

図18 推定人数の結果(15人配置)

表6 人数推定誤差の比較(単位:人)

No.	N=5	N=10	N=15	N=20	N=25
B0	0.55	0.76	0.80	0.72	1.19
B1	0.23	0.54	1.26	2.21	3.63
B2	0.52	0.74	0.87	0.83	1.42
B3	0.70	0.95	1.19	1.02	1.12
B4	0.26	0.66	1.43	2.46	3.93
B5	0.22	0.66	1.39	2.40	3.87
B6	0.91	1.48	1.91	1.76	1.50
B7	0.62	0.75	1.23	1.71	2.91
B8	1.14	2.06	2.69	2.62	2.30
B9	0.27	0.45	1.01	1.84	3.20
B10	0.63	0.81	1.28	1.94	3.19
B11	0.44	0.73	1.24	2.25	3.68
B12	0.30	0.40	0.72	1.40	2.58
B13	0.51	0.71	0.87	0.99	1.76
B14	0.42	0.49	0.77	1.04	2.07
B15	0.20	0.57	1.27	2.26	3.71

程度の精度で人数が推定されている。

以上のシミュレーション実験によると20人までの人数推定では数人程度の誤差であり、提案方法により人数推定に十分なカメラパラメータが調整されていることを示した。よって、荷重値積算による人数推定において提案方法の有効性を確認した。

5.2 実画像を使った実験

実画像実験には121フレームの時系列画像(各画像の大きさは200×160画素)を使用した。標本画像中の人数の違いによって人数推定がどうなるかを調べるため、5人程度が映っている画像(サンプルC)と、10人程度が映っている画像(サンプルD)をそれぞれ標本画像として準備した。図19、図22はそれぞれサンプルC、サンプルDの標本画像である。実画像の場合、各画像から画像処理により前景画素を抽出する必要がある。屋外では緩急の照明変化が考えられるため、その状況に応じて適切な背景差分を選択する。例えば、照明状態が変化した場合では背景モデルを逐次更新する適応型の背景差分[8]などの応用が考えられる。我々の使用した時系列画像は全体にわたって照明状態がほぼ一定だったため、背景画像を一定シーンの平均で生成し、その背景画像に対する差分を各時系列画像から抽出した。さらに、この差分処理による人物像内部の画素欠損を補うため、膨張と収縮を組み合わせた適当なモルフォロジ演算を施し前景画素を得た(実装には

OpenCV[2]で用意されている関数を利用した)。図20、図23には、各標本画像から抽出された前景画素が白塗りにより示されている。

この前景画素を使って提案方法によりカメラパラメータを調整した。一方、人の映っていない画像においてタイルの格子点を参照点として見立て、複数の参照点の画像座標を測定するとともに、各格子点の空間位置をできるだけ正確に計測した。得られた参照点の空間座標と画像上での2次元座標を使って従来方法によりカメラキャリブレーションを行った。従来方法によって得たカメラパラメータと提案方法により調整されたカメラパラメータを表7に示す。サンプルCにおいて、提案方法はカメラ視点の位置(T_x , T_y , T_z)と焦点距離 f を従来方法で得たカメラパラメータに近い値まで調整しているのが分かる。

続いて、表7のカメラパラメータから荷重値を算出し、時系列画像に対する人数推定の実験を行った。121フレームの時系列画像から推定した人数のグラフを図21、図24に示す。赤線で示したグラフは目視によって計測した人数(ground truth)であり、破線で示したグラフは従来方法による推定人数であり、実線で示したグラフが提案方法による推定人数($h=1.7m$, $S_0=0.51m^2$ として推定)である。破線が示しているように、従来方法は目視による計測数に追従して人数を推定しており、3節で記述した荷重値積算による人数推定の妥当性が示されている。図21では、従来方法に基づく人数と比べると、提案方法に基づく人数はやや少なめである。ただし、40~60フレームでは、提案方法は従来方法よりも目視による計測数に近いことが確認できる。一方、図24では、提案方法に基づいて推定された人数は従来方法に基づく人数とほぼ一致している。この結果では、カメラキャリブレーションと遜色無く、時系列画像から人数を推定しており、提案方法の有効性が認められる。これを数量的に把握するため、全フレームにわたる推定人数の絶対平均誤差 N_e とその標準偏差 N_σ を表7にまとめた。この評価基準では、サンプルDからパラメータを調整すると従来方法による結果に近く、サンプルCから調整した場合では推定誤差のばらつきがやや大きいことが分かる。なお、図20、図23には人物像の領域に前景画素として抽出できていない箇所がある。これは、背景画像において、服装の一部がタイルの模様と黄色のラインに同化してしまったことに起因する。目視計測による人数と荷重値積算による推定人数の間の誤差は、この前景画素の欠損が影響しているものと考えられる。

図19 サンプルCの標本画像

図20 背景差分による前景画素の抽出

図21 時系列画像に対する人数推定の比較

図22 サンプルDの標本画像

図23 背景差分による前景画素の抽出

図24 時系列画像に対する人数推定の比較

表7 調整パラメータと計測誤差

	提案方法		
	従来方法	サンプル C	サンプル D
ϕ [deg]	-56.4	-47.3	-52.7
ω [deg]	13.4	7.2	4.5
θ [deg]	-15.4	3.4	5.3
T_x [m]	2.7	2.5	-2.8
T_y [m]	9.1	8.6	10.9
T_z [m]	6.1	6.6	5.4
f [pixel]	340.3	356.9	297.0
N_μ	0.61	0.65	0.62
N_σ	0.47	0.55	0.45

5.3 考察

上記の実験結果を踏まえて、提案方法に関する考察を行う。まず、標本画像中の人数と計測人数の信頼性について考察する。実画像実験では、目視による人数計測と比べて、荷重値積算による人数推定の誤差は約1人程度であり、提案方法でもほぼ同等の結果が得られた。グループAに関する計算機シミュレーションは実画像取得の状況に近い条件を模擬しており、表4に示されているように、全サンプルの $N=10$ に対する推定誤差は約1人という結果が得られている。これは実画像の推定誤差と符合する。この点において、我々は提案方法によって3枚の標本画像でも人数推定に十分なカメラパラメータが調整されることを示した。また、グループAにおいて、5人を配置したサンプルでは、10人、15人を配置したサンプルの場合と比べると推定誤差がやや大きいことが分かった。サンプルCを使った実画像実験でも、提案方法による人数はカメラキャリブレーションの場合と比べてやや少なかった。このように、標本画像の人数が多くなると人数推定の精度が向上する傾向が見られる。各画素の荷重値算出には、面積 S_0 の板モデルが一樣に分布するという仮定が使われている。この仮定から推察すると、多数の板の投影像が画像上を覆い尽くすことで、一樣分布に従った荷重値算出の信頼度が高くなったと考えられる。

次に、標本画像中の人数とオクルージョンの影響について考察する。グループBの標本画像を生成したカメラ C_2 は、カメラ C_1 に比べてやや高い位置からより急な傾斜角度で矩形モデルを俯瞰する。図7、図9、図11では矩形モデルの重複が多く、オクルージョン頻度が高いのに対し、図13、図15、図17ではその隠蔽が減少していることが確認できる。表4と比べると表6では各サンプルとも数人程度の推定誤差となってい

る。これについては、カメラ姿勢が変化したことでもオクルージョンの発生頻度が減少し、提案方法が有効に機能したものと考えられる。しかしながら、提案方法によるパラメータ調整は荷重値積算の品質に依存している。荷重値積算はあくまで“見かけの人数”を推量するため、大勢の人を映した標本画像で調整することはオクルージョンの影響が強くなり推定人数の精度劣化を招きかねない。さらに、標本画像において人数が多くなると前景画素が増加するため、パラメータ調整の計算コストは高くなる。よって、実運用においては、計算コストを抑え、かつ、パラメータ調整の信頼性を高める上で、試験的に10～15人程度の標本画像からパラメータを調整し、人数推定が安定しているかどうかを判定することが必要になる。

続いて、カメラキャリブレーションとカメラパラメータ調整の違いについて考察する。表3、表5ではサンプルによってはカメラパラメータにばらつきがあるにもかかわらず、表4、表6では推定人数の大幅な精度劣化に波及していなかった。これらの実験が示しているように、提案方法は人数推定に十分なカメラパラメータを調整しただけであって、カメラキャリブレーションと同等のカメラパラメータを獲得してはいなかった。そもそも、参照物体を使ったカメラキャリブレーションでは、カメラパラメータは参照物体に定義した座標系と関連付けられている。提案方法は参照点を用いず、前景画素の荷重値の積算による人数が適応性が評価される。提案方法にとって、世界座標系の原点は“見えモデル”群の重心に応じて変動するため視点位置が標本画像によって異なる。さらに、荷重値算出には“見えモデル”の方向が考慮されていないので ω 、 θ の回転角も標本画像に左右される。従って、提案方法によって調整されたカメラパラメータはカメラキャリブレーションで得たカメラパラメータに必ずしも一致しない。遺伝的アルゴリズムは、このような“緩和的な”推定システムにおいて、環境に応じて適応的なパラメータを獲得するという点で有効な手法であった。

6. おわりに

本論文では、荷重値積算による人数推定の枠組みの中で、荷重値算出に必要なカメラパラメータを遺伝的アルゴリズムによって適応的に調整する方法を提案した。実環境では、処理コストを優先としたパラメータ調整がシステム導入には有利であることを考慮し、シミュレーション実験と実画像実験では3枚の標本画像を使って提案方法によりカメラパラメータを調整した。シミュレーション実験では提案方法によって20人

までの人数を数人程度の誤差で推定することを示した。実画像実験では時系列画像から約1人の誤差で人数を推定しており、目視による計測結果と比べても遜色無い精度であることを確認した。これらの実験結果から、提案方法の有効性を示すことができた。提案方法は人数推定に十分なカメラパラメータを観測画像を利用して調整することができるため、ユーザの作業量を削減する点でも大いに価値がある。

最後に、本研究の展望について述べる。荷重値積算による人数推定法は、単眼カメラを観測手段としているためオクルージョンの影響による推定誤差を内在している。カメラの設置状況によっては今回のシミュレーション実験で得た推定誤差よりも大きくなることが予想される。このため、オクルージョンによる人数推定誤差の補正を荷重値積算による人数推定法に取り入れることは、提案方法の調整精度を向上させる上でも重要となる。また、実画像実験において背景差分による前景画素抽出が人数推定に影響を与えることに触れた。屋外環境で見られるように照明変動が著しいと、前景画素抽出の品質が低下することがある。照明変化に追従して前景画素を安定的に抽出する方法が実用化されれば、提案方法をよりロバストなカメラパラメータ自動調整法へ発展させることもできる。

参 考 文 献

- [1] H. Arai, I. Miyagawa, H. Koike, and M. Haseyama. Estimating the Number of People in Video Sequence via Geometrical Model. *Proc. IEEE International Conference on Pattern Recognition*, 2008.
- [2] G. R. Bradski and V. Pisarevsky. Intel's Computer Vision Library: Applications in calibration, stereo, segmentation, tracking, gesture, face and object recognition. *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Vol.2, pp.796-797, 2000.
- [3] S. Y. Cho, T. W. S. Chow, and C. T. Leung. A Neural-Based Crowd Estimation by Hybrid Global Learning Algorithm. *IEEE Trans. Systems, Man, and Cybernetics, Part B*, Vol.29, No.4, pp.535-541, 1999.
- [4] Q. Ji and Y. Zhang. Camera Calibration with Genetic Algorithms. *IEEE Trans. Systems, Man, and Cybernetics-Part A: Systems and Humans*, Vol.31, No.2, pp.120-130, 2001.
- [5] S. F. Lin, J. Y. Chen, and H. X. Chao. Estimation of Number of People in Crowded Scenes using Perspective Transformation. *IEEE Trans. Systems, Man, and Cybernetics, Part A*, Vol.31, No.6, pp.645-654, 2001.
- [6] M. Mitchell. 遺伝的アルゴリズムの方法. 東京電機大学出版局, 1997.
- [7] V. Rabaud and S. Belongie. Counting Crowded Moving Objects. *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Vol.1, pp.705-711, 2006.
- [8] C. Stauffer and W. E. L. Grimson. Adaptive Background Mixture Models for Real-Time Tracking. *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Vol.2, pp.246-252, 1999.
- [9] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon. Bundle Adjustment-A Modern Synthesis. *Vision Algorithms: Theory and Practice*, pp.298-375, 2000.
- [10] Z. Zhang. A Flexible New Technique for Camera Calibration. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.22, No.11, pp.1330-1334, 2000.
- [11] 金谷健一. 画像理解-コンピュータ数理の基礎-. 共立出版, 1994.
- [12] 熊倉広志. 広告効果の概念と測定・分析の方法. 日本経営工学会 経営システム (特集: データを活かすマーケティング), Vol.18, No.1, pp.19-25, 2008.
- [13] 植芝俊夫, 岡谷貴之, 佐藤智和. カメラキャリブレーション. 情報処理学会研究報告 2005-CVIM-148, Vol. 2005, No.18, pp.1-18, 2005.
(2008年6月15日 受付)
(2009年1月15日 採録)

[問い合わせ先]

〒239-0847 神奈川県横浜賀市光の丘1-1
日本電信電話株式会社 NTTサイバースペース研究所
宮川 勲
TEL: 046-859-2705
FAX: 046-859-2829
E-mail: miyagawa.isao@lab.ntt.co.jp

著者紹介

みやがわ いさお
宮川 勲 [非会員]

1991年福井大学工学部電子工学科卒業。同年、日本電信電話(株)入社。カラーファクシミリにおける複合符号化と入出力色信号処理の研究、並びに、ファクシミリ通信機器の開発を経て、コンピュータビジョンに関する研究開発に従事。2005年福井大学大学院工学研究科博士後期課程修了。博士(工学)。2005年12月から2007年3月までNHK放送技術研究所において多視点カメラシステムを使った3次元映像生成の研究に従事。現在、日本電信電話(株)NTTサイバースペース研究所勤務。IEEE Computer Society、電子情報通信学会各会員。

あらい ひろゆき
新井 啓之 [非会員]

1989年東京理科大学理工学部物理学科卒業。1991年北海道大学大学院理学研究科(物理)修了。同年、日本電信電話(株)入社。図面・文書の画像認識に従事。2001年から2005年までNTTデータ(株)勤務。現在、日本電信電話(株)NTTサイバースペース研究所画像メディア通信プロジェクト主任研究員。情報通信研究機構(NICT)ナチュラビジョンプロジェクト特別研究員(2001年～2005年)。電子情報通信学会会員。

こいけ ひでき
小池 秀樹 [非会員]

1985年東北大学理学研究科数学専攻博士課程前期2年の課程修了(理修)。同年、日本電信電話(株)入社。画像処理、画像認識に関する研究に従事。2001年から2007年までNTTコムウェア(株)に転籍し、RFIDの研究開発に従事。現在、日本電信電話(株)NTTサイバースペース研究所画像メディア通信プロジェクト主幹研究員。電子情報通信学会会員。

Camera Parameter Adjustment using Genetic Algorithm for Estimating the Number of People from Image Sequence

by

Isao MIYAGAWA, Hiroyuki ARAI and Hideki KOIKE

Abstract :

We use foreground pixels as indicators of the presence of people to estimate the total number of people in an image sequence; the proposed method has low processing cost and can output the number in real time. The number of people is computed from the visual hull for each pixel, based on the camera parameters of a stationary camera, and a simple geometric model for each person. The proposed method uses a genetic algorithm to automatically adjust the camera parameters, in the framework of calculating the total number of people. We evaluate the performance of the method with computer simulations and demonstrate its effectiveness on real images.

Keywords : image measurement, estimation of number of people, stationary camera, genetic algorithm, background subtraction, and camera calibration

Contact Address : **Isao MIYAGAWA**

NTT Cyber Space Laboratories, NTT Corporation
1-1, Hikarini-no-oka Yokosuka Kanagawa, 239-0847, JAPAN
TEL : 046-859-2705
FAX : 046-859-2829
E-mail : miyagawa.isao@lab.ntt.co.jp