2 台のカメラを用いた歩行者の検出・同定

太 田 直 哉[†] 木 山 真 伸[††] 吉 川 健 一[†††]

道路などを撮影した画像を用いて，そこを通過する歩行者を安定に検出，同定し，追跡する手法を提案する．提案手法は，画素単位の処理で進入物体の有無を判定する「進入物体検出」，その結果から 1 人 1 人の歩行者を同定する「歩行者同定」，同定された各歩行者を追跡する「歩行者追跡」の 3 段階の処理からなる．提案手法の 1 つの特長は「進入物体検出」において照明の変化に影響されず，影を検出しない利点を持つ 2 台のカメラを用いる方式を使用し，その方式の正しい動作のために必要な正確な幾何学的変換の決定とカメラの感度較正を行う方法を開発したことである．提案手法のもう 1 つの特長は,「歩行者同定」に正則化逆フィルタと呼ぶウイナーフィルタに似た処理を採用し，近接した歩行者に対して分離の良い同定を可能にしたことである．前記の 2 台のカメラを用いる手法では,「進入物体検出」において歩行者の上半身が二重映しのように広がって検出され，接近した歩行者が分離しにくいという欠点を持つが，この同定処理によってこの欠点を軽減した．屋外シーンを用いた実験では，照明の変化や影に影響されない追跡が可能であることが確認された．全体の追跡成功率は約 93%であった．

Pedestrian Detection and Identification Using Two Cameras

Naoya Ohta,[†] Masanobu Kiyama[††] and Kenichi Yoshikawa[†††]

We propose a pedestrian detection, identification, and tracking method using image information. The method consists of three stages. The first one is incoming object detection, which detects entering objects by pixel-wise operations. The second stage is pedestrian identification, which identifies individual pedestrians, and the third stage is pedestrian tracking, which tracks the position of the identified pedestrians. For the incoming object detection, we adopt a method using two cameras, which is insensitive to illumination change and shadows. One novel point of this paper is to develop a method for determining accurate geometrical transformation and calibrating camera sensitivities, which are needed for a proper operation of the incoming object detection process. The other novel point is to use a process similar to the Wiener filter, which we named *regularized inverse filter*, for the pedestrian identification. This improves separation of close-by pedestrians. Experiments using outdoor scenes proved that the method provided a stable detection and tracing result against the illumination change and the shadow problems. The ratio of successfully traced pedestrians was 93%.

1. はじめに

監視カメラにより得られる画像を用いて，そこに進入する物体を検知するシステムは従来より数多く報告されている[1)~4)]．対象となる物体も車輌や人物など多岐にわたり，また想定するカメラの設置場所も様々であるが，本論文では重要施設における不審者発見や交通量調査などの用途を想定して，道路を上方から見下ろした画像を用いて歩行者を検出，同定，追跡する処理について述べる．

一般にこのような処理は，(1) 入力画像の画素ごとに物体が存在するかどうかを判定する処理，(2) 物体が存在すると判定された画素の画像上での形状から検出対象を同定する処理，(3) 同定された物体を時系列上で追跡する処理，という 3 段階の処理で構成されることが多く，本論文でもこの構成としている．本論文では，これらの処理をそれぞれ，(1)「進入物体検出」，(2)「歩行者同定」，(3)「歩行者追跡」，と呼ぶことにする．以下これらの処理の概要と本論文で用いた方式の特徴を説明する．

まず，(1)「進入物体検出」では入力画像の各画素について独立に演算を行い，その画素に撮影されている物体の点が進入物体（歩行者）のものであるかどうかを判定する．この目的のために用いられる代表的な

† 群馬大学工学部情報工学科
Department of Computer Science, Gunma University
†† 三洋 LSI デザイン・システムソフト株式会社
Sanyo LSI Design Inc.
††† 富士写真フィルム株式会社
FUJIFILM Corporation

処理は，背景差分[1),2),4)]とフレーム間差分[1),3)]である．しかし背景差分においては，つねに正しい背景画像を保持しておくことが困難な問題である．背景画像を適応的に更新するアルゴリズム[5)~7)]も提案されているが，照明の様々な変動に完全には対応できないことが多い．また，影を検出してしまうという問題もある．一方フレーム間差分は基本的に移動物体の検出処理であるため，適当な速度で移動している物体に対しては有効に働くが，静止したり想定している速度よりも高速な物体に対しては有効でない．そこで提案手法では，カメラを 2 台用いることでこれらの問題を解決する．

2 台のカメラを用いる処理に関しては 2 つの手法が提案されており，その 1 つはステレオ画像処理によるものである[8),9)]．これはステレオによって対象物体の 3 次元形状を得て，その形状を解析することで進入物体を検出する手法であり，原理的に照明の変動や影の問題は生じない．しかしテクスチャがない部分での対応点決定が困難であることや，演算量が多いといった，ステレオ画像処理の持つ問題が生じる．もう 1 つの手法は，背景物体のみのときに左右の画像が一致するような幾何学的変換によって画像を変形し，差分をとる手法である[10),11)]．これは一度幾何学的変換を求めてしまえば対応点の決定に関する問題もなく，演算量も少ないので，本論文ではこの手法を用いる．しかしこの手法を適用して良い結果を得るためには，正確な幾何学的変換を求めると同時に左右のカメラの感度特性も一致させる必要がある．本論文ではこれを行う方法について述べる．またこの手法の欠点として，検出される物体領域が本来の形状に関して二重写しのようになる（たとえば図 5 (b) を参照のこと）点があげられるが，これについては次の「歩行者同定」を工夫することで対処することとする．

次に，(2)「歩行者同定」であるが，ここでは前記の処理で進入物体が存在すると判定された画素が構成する形状を解析して，歩行者 1 人に対応する領域を同定する．これは 2 値画像の形状解析であり，1 つの方針として膨張・収縮やラベリング処理を組み合わせる方法[12)]が考えられる．しかしこの方針は人手で調整すべきパラメータが多くなり，さらにそれらのパラメータと検出される形状との関係が明確でないという問題を生じがちである．そこで本論文では画像から形状情報を直接抽出できるテンプレートマッチングを基本とすることとした．入力画像とテンプレートとのマッチング演算に関してはハフ変換による方法[4)]も報告されているが，ここでは演算の高速化手法として FFT を利用できる相関演算を基礎とした．ただし前記「進入物体検出」での二重写しの問題をできるだけ緩和するため，正則化逆フィルタと名づけたウイナーフィルタ的な手法を導入し，歩行者の分離性能を向上させている．なお相関値の分離を良くする方法の 1 つとして位相相関法が知られているが，これとここで用いた手法との関係は付録 **A.2** にまとめた．

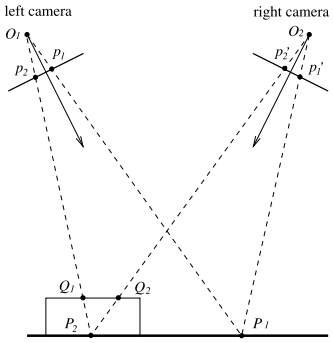


図 1 進入物体検出の原理
Fig. 1 Principle of object detection.

次の，(3)「歩行者追跡」は，1 人 1 人の歩行者に対応する点を時系列的に追跡する処理である．このための処理として，コンデンセーション（Condensation）やカルマンフィルタを用いて，人物の存在確率が最も大きいところをトレースすることが行われている[13)]．しかし対象の動きが複雑でなく，また歩行者同定の誤りもそう多くないことから，ここでは比較的単純なアルゴリズムを採用した．このアルゴリズムは過去の動きから次の時刻での位置を予測し，その予測位置に最も近い点を対応付ける．

以下，上記の各処理について順次詳細を説明した後，実験によって提案手法の性能を評価し，最後にまとめを行う．

2. 進入物体検出

本章では「進入物体検出」についてその原理を述べた後，その原理を実際に適用するために必要な処理，すなわち画像の幾何学的変換の決定やカメラの感度特性の較正について述べる．さらにこの処理に関する実験結果を示す．

2.1 原　　理

本論文で用いた進入物体検出の原理を**図 1** に示す．この図において下部の太線は背景物体（地面）であり，これを左右 2 台のカメラで撮影した様子が描かれている．背景物体上の点 P_1 は左右のカメラ上で p_1 および p_1' に対応するが，これらの画像点は物体上の同じ点を撮影しているので，物体表面が完全拡散面（Lambertian）であり，さらに 2 台のカメラの感度が

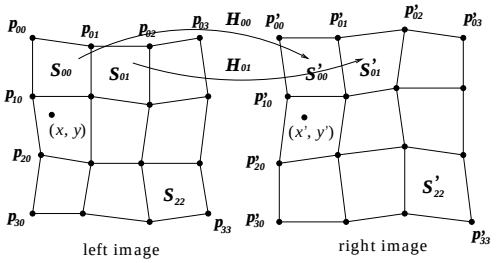


図 2 複数の矩形による監視領域の分割
Fig. 2 Surveying region divided by multiple rectangles.

同一であることを仮定すれば，その画素値は一致する．しかし P_2 で示された部分のように，進入物体が存在する場合には画像上の点 p_2 および p'_2 は異なる点 Q_1 および Q_2 に対応し，一般に物体上のこの 2 点の明るさは異なるので，画像上の点 p_2 および p'_2 での画素値も異なる．したがって，これらの対応する画像点の画素値の差分を計算して，大きな値を持つ場合に進入物体があると判定する．

2.2 幾何学的変換の決定

前述の原理を実際に適用するにあたって，左右の画像における対応点をどのように算出するかの問題がある．背景物体が平面である場合には，この対応は射影変換で関係付けられるが，一般には背景物体の表面は平面ではない．そこで背景物体を区分的な平面で近似する．すなわち画像上の進入物体を検出すべき領域（以降，監視領域と呼ぶ）を区分的な矩形に分割し，その矩形領域では背景物体は平面であることを仮定する．これにより各矩形内での対応は射影変換によって関係付けられるが，この射影変換は各矩形の 4 頂点の対応から決定できる．すなわち右画像の各画素を，それが属している矩形を判定し，対応する射影変換で変換する．

ここで射影変換を表現するために同次座標系を導入し，上記の処理を具体的に示すと**図 2** のようになる．この図において $\boldsymbol{p}_{ij}$ および $\boldsymbol{p}'_{ij}$ がそれぞれ左右の画像上での対応点を同次座標で表現したもの，S_{ij} および S'_{ij} が対応する矩形領域，H_{ij} がそれらの間の射影変換を表している．左画像上の座標 (x, y) の点がどの矩形領域に属しているかの判定は，まずこの点を同次座標 $\boldsymbol{p} = (x\ y\ 1)^\top$ で表現し，以下の条件を調査する．

$$(\boldsymbol{p}, \boldsymbol{p}_{i\,j} \times \boldsymbol{p}_{i\,j+1}) \geq 0 \tag{1}$$

$$(\boldsymbol{p}, \boldsymbol{p}_{i\,j+1} \times \boldsymbol{p}_{i+1\,j+1}) \geq 0 \tag{2}$$

$$(\boldsymbol{p}, \boldsymbol{p}_{i+1\,j+1} \times \boldsymbol{p}_{i+1\,j}) \geq 0 \tag{3}$$

$$(\boldsymbol{p}, \boldsymbol{p}_{i+1\,j} \times \boldsymbol{p}_{ij}) \geq 0 \tag{4}$$

ここで記号 $\times$ および $(\cdot, \cdot)$ はそれぞれベクトルの外積と内積を表す．これらの条件がすべて成立するとき注目点は S_{ij} に属しており，射影変換 H_{ij} によって変換される．射影変換 H_{ij} のパラメータを $a_{\cdot\cdot}$（ドットは添字を表す）とすれば，右画像上の対応点の座標 (x', y') は次式で計算される．

$$x' = \frac{a_{11}x + a_{12}y + a_{13}}{a_{31}x + a_{32}y + 1} \tag{5}$$

$$y' = \frac{a_{21}x + a_{22}y + a_{23}}{a_{31}x + a_{32}y + 1} \tag{6}$$

画像の幾何学的変形は，右画像上の位置 (x', y') の画素値を位置 (x, y) に持つ画像を生成することで行われる．なおこの際必要となる画素の補間は双一次内挿法（bi-linear interpolation）を用いた．

次に矩形の 4 つの頂点から射影変換 H_{ij} のパラメータを求める方法について述べる．添字の煩雑さを避けるため，ここでは左右の画像での各頂点の座標をそれぞれ (x_k, y_k) および (x'_k, y'_k) $(k = 1, \ldots, 4)$ とする．すると式 (5) および (6) より次の関係が各 k について成立する．

$$\begin{aligned} x'_k = {} & a_{11}x_k + a_{12}y_k + a_{13} \\ & - a_{31}x'_k x_k - a_{32}x'_k y_k \end{aligned} \tag{7}$$

$$\begin{aligned} y'_k = {} & a_{21}x_k + a_{22}y_k + a_{23} \\ & - a_{31}y'_k x_k - a_{32}y'_k y_k \end{aligned} \tag{8}$$

これらを連立させればパラメータ $a_{11}, \ldots, a_{32}$ が計算される．

2.3 幾何学的変換とカメラ感度の較正

上記の幾何学的変換は左右の画像の対応点から決定されるが，人手による対応点の決定はその精度が十分でない．また提案方式が適切に動作するためには，左右のカメラの感度が較正されていなければならない．ここではこれら 2 つの問題への対処法を述べる．

基本的な考え方は，背景のみの画像を撮影したときに，右の画像に幾何学的変換を施した画像と左の画像の差が最も小さくなるように各矩形の頂点の位置とカメラの感度パラメータを修正することである．これを行うには次のような評価関数 J を定義する．

$$J = \sum_{x,y} \left\{ (\alpha \hat{I}_r(x, y) + \beta) - I_l(x, y) \right\}^2 \tag{9}$$

ここで $I_l(x, y)$ は左画像，$\hat{I}_r(x, y)$ は右画像に幾何学的変換を施した画像，α および β はそれぞれ左右のカメラの感度とバイアスを補正する係数である．また画素の総和は監視領域全体でとる．なお画素値を線形式で補正することの妥当性は付録 **A.1** で議論した．監視領域を構成する頂点の数を N とすると変数の数は $2N + 2$ となるが，これらの変数に関して上式の J を最小化する．最小化のアルゴリズムは Powell 法[14]によった．

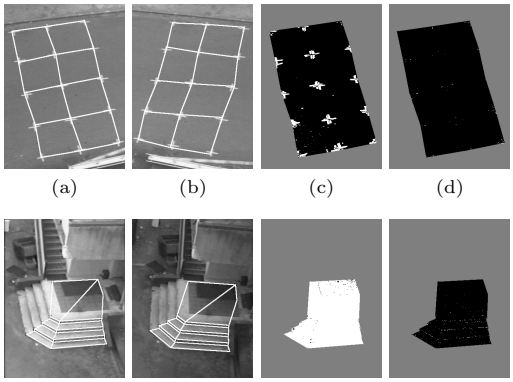


図 3 幾何学的変換とカメラ感度の較正の効果：(a) (e) 左画像，(b) (f) 右画像，(c) (g) 較正前，(d) (h) 較正後

Fig. 3 Effectiveness of calibration: (a) (e) left image, (b) (f) right image, (c) (g) uncalibrated, (d) (h) calibrated.

上記の較正の効果を**図 3** に示す．上段が比較的平面に近い背景物体での例，下段が凹凸のある物体での例である．それぞれ (a) (e) が左画像，(b) (f) が右画像，(c) (g) が較正前の差分画像，(d) (h) が較正後の差分画像である．左右の画像上には監視領域を示す白線が重ね描きされている．なお差分画像に関しては監視領域以外を灰色，監視領域内で差分の絶対値が 20 レベル以下の部分を黒，20 レベルを超える部分を白で表示している．画像の全階調数は 256 である．上段の道路の例では，対応点を決定するための目印が道路上に存在しないため白いマークを付加し，それらを特徴点として手動で対応付けを行い，初期値としている．較正を行う前と後の差分画像（図 3 (c) および (d)）を比較すると，較正により左右の画像の差がなくなっていることが分かる．監視領域内での差分の絶対値の平均は，較正前には 10.8 レベルであったものが較正後には 3.4 レベルに減少した．ここでは左右のカメラに同じものを用いたためそれらの感度差は少なく，校正後の α，β の値はそれぞれ 0.97，0.95 であった．図 3 下段の階段の例では，カメラの感度較正の効果を見るために左右のカメラの絞りを異なった値にした．そのため較正前の差分画像では大きな差を生じ，ほぼ全面が白くなっているが，較正後は差分がほとんど生じていないことが分かる．較正前に 35.6 レベルであった差分の絶対値の平均は，較正後には 4.2 レベルに減少した．また α，β の値はそれぞれ 1.298，17.29 であった．幾何学的変換の精度を評価するため，上記の 2 つの例について，右画像に変換を施した画像と左画像とを重ね合わせ，監視領域の各頂点として用いた特徴点の位置ずれを目視によって確認した．結果は十分良い一致を示し，330×450 画素の大きさの画像でほぼ 1 画素以下であった．

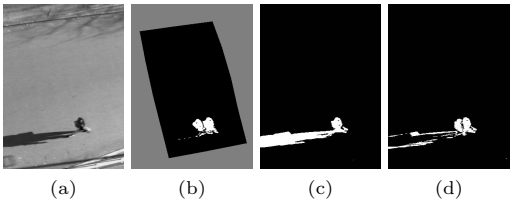


図 4 影のある画像での歩行者検出：(a) 左画像，(b) 提案手法，(c) 背景差分，(d) フレーム間差分

Fig. 4 Pedestrian detection from images with shadow: (a) left image, (b) proposed method, (c) background subtraction, (d) frame difference.

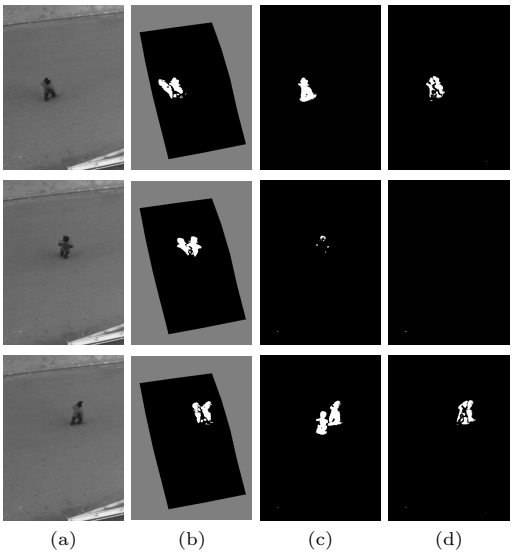


図 5 立ち止まる歩行者の検出：(a) 左画像，(b) 提案手法，(c) 背景差分，(d) フレーム間差分

Fig. 5 Detection of standing pedestrian: (a) left image, (b) proposed method, (c) background subtraction, (d) frame difference.

2.4 歩行者の検出結果

本論文で採用した進入物体検出手法の利点を確認するために，背景差分およびフレーム間差分との比較結果を**図 4** および**図 5** に示す．これらの図で (a) が左画像，(b) が提案手法による差分画像であり，(c) および (d) はそれぞれ比較のために示した背景差分☆とフレーム間差分の結果である．図 4 は影のある画像での検出結果であるが，背景差分およびフレーム間差分で

☆ ここでは基本的な性質を比較するという意味で，画像の累積による単純な背景生成法を用いた．改良された背景生成手法を用いれば，ある程度は改善されると考えられる．

(a) (b) (c)

図 6 凹凸のある背景物体での歩行者検出：(a) 左画像，(b) 右画像，(c) 検出結果

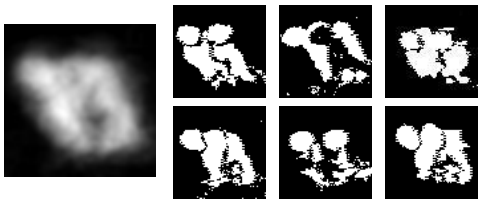
Fig. 6 Pedestrian detection under non-plane background: (a) left image, (b) right image, (c) detection result.

は影が検出されるが，提案手法では検出されないことが確認できる．図 5 は立ち止まる歩行者の検出結果であり，上段から下段に時間が経過している．フレーム間差分では歩行者が立ち止まると検出ができなくなっている．背景差分では立ち止まってからしばらく時間が経過すると検出されなくなり，再び歩き始めるとまた検出されるが，立ち止まっていた領域も検出されるという問題が生じている．これに対し提案手法ではこのような問題はない．なお直立した人物を撮影した場合，提案手法では V の字型の差分が検出されるはずである．しかし図 5 (b) 中段の画像で，足元の部分が検出されていないのは以下の理由による．地面に近い部分では，2 枚の画像上で対応する画素が，物体上の近い位置を写すことになる．この部分は歩行者のズボンの裾にあたるが，この部分は同一色をしているので差分が生じない．さらに図 5 (b) の中段以外の差分画像が V の字型をしていないのは，上記の理由に加えて足を広げているからである．

図 6 は凹凸のある背景物体への適応性を確認するために図 3 下段のシーンで歩行者を検出した結果である．この図の (a) が左画像，(b) が右画像，(c) が差分画像である．これにより凹凸のある背景物体に対しても問題なく歩行者が検出されることが分かる．

3. 歩行者同定

「歩行者同定」では，前章で述べた「進入物体検出」によって画素単位に判定された進入物体領域を歩行者 1 人 1 人に対応する領域として同定する．処理の基本的方針は以下のとおりである．まず歩行者の形状を平均化したテンプレートを作成する．そして 2 値化後の差分画像の各部分でそのテンプレートとの類似性を計算し，極大となる点を 1 人の歩行者に対応する点として検出する．テンプレートを用いた対象の検出において，原理的には対象の形状は変化しないことが仮定されるが，歩行者は歩行によって主に手と足の部分の形状が変化する．しかし頭と胴体の形状の変化は少なく，



(a) (b)

図 7 歩行者同定用のテンプレート：(a) テンプレート画像，(b) 歩行者領域の例

Fig. 7 Template for pedestrian identification: (a) template, (b) samples of pedestrian region.

歩行者を上方から見おろした画像を用いていることからこの部分が相対的に大きな面積を占める．そこで形状を平均化したテンプレートを用いれば，手と足の部分の変形に対処可能と考え，この方式を採用した．

実際に作成したテンプレートを**図 7** (a) に示す．これは 1 人の歩行者に対応する差分画像の領域を人手で 50 枚切りだし，これらを平均したものである．これに使用した歩行者領域の例として 6 枚を図 7 (b) に示した．

次にテンプレートと差分画像との類似性の計算であるが，相互相関演算が一般的である．しかし単なる相関演算では近接した歩行者の分離が良くないので，ここでは相関のピークを鋭くする方法を考える．基本的な考え方は次のとおりである．歩行者を検出するために理想的な相関画像とは，各歩行者領域の中心に δ 関数があり，それ以外は 0 となるような画像である．ここで，実際に得られる差分画像はそのような理想的な画像がテンプレートを PSF（Point Spread Function）としてぼけていると考える．したがってそのぼけを修復する処理を施せば，より理想に近い相関画像が得られると期待される．もちろんこの処理は不良設定問題（ill-posed problem）であるので，ウイナーフィルタのように正則化の効果を持つ処理を適用する．ここではこれを正則化逆フィルタと呼ぶ．

具体的な計算は以下のとおりである．差分画像およびテンプレートのフーリエスペクトルをそれぞれ $F(u,v)$ および $T(u,v)$ とすると，相関画像のフーリエスペクトル $C(u,v)$ を以下のように得る．

$$C(u,v) = \frac{T^*(u,v)}{|T(u,v)|^2 + \eta} F(u,v) \qquad (10)$$

ここで u および v は空間周波数であり，肩の $*$ は複素共役を表す．そして $C(u,v)$ を逆フーリエ変換すれば相関画像が得られる．上式の η は安定化項で，小さくすれば逆フィルタに近づき，大きくすれば通常の相関演算に近づく．**図 8** にこの変化を示した．この図の

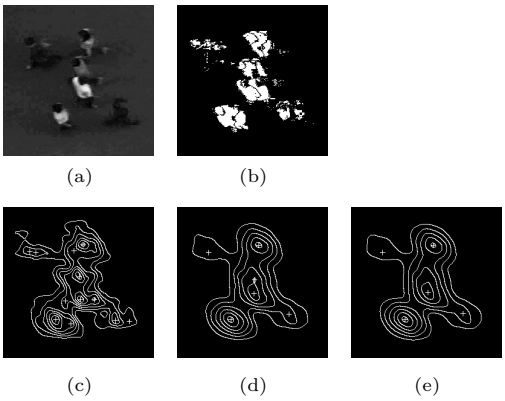


図 8 相関画像の η による変化：(a) 原画像，(b) 差分画像，(c) $\eta = 0.1$，(d) $\eta = 0.5$，(e) $\eta = 1.0$

Fig. 8 Correlation images in different ηs: (a) original image, (b) difference, (c) $\eta = 0.1$, (d) $\eta = 0.5$, (e) $\eta = 1.0$.

(a) は原画像，(b) は差分画像，(c)，(d)，(e) はそれぞれ η を 0.1，0.5，1.0 としたときの相関画像を等高線表示したもので，各画像の大きさは 250×250 画素である．相関画像中の十字は極大点を示している．なお，分母のもう 1 つの項である $|T(u,v)|^2$ の最大値は 0.997 であった．このシーンには 6 人の歩行者がおり，画像上側の 3 人が右から左，下側の 3 人が左から右に歩いている．画像中央の 2 人に注目すると，η が 1.0 の相関画像ではそれらに対応するピークが 1 つで分離されていない．一方 $\eta = 0.1$ では擬似のピークが多く生じている．η が 0.5 の場合には擬似ピークなく，中央の歩行者が分離されているのが分かる．最終的に η の値は 0.5 に決定したが，これは以下の手順によった．接近した歩行者のシーンから $\eta = 1.0$ では分離できない 23 例を選び，η を 0.1 ずつ小さくしながらピークの状況を観察した．いずれかの例で擬似ピークを生じたのが $\eta = 0.4$ の場合であったので $\eta = 0.5$ を採用した．この値で分離できない例は 23 例中 8 例あったが，この問題は歩行者を追跡する段階で考慮することとした．なお，通常の相関演算に較べ鋭いピークを得る 1 つの手法として知られている位相相関（phase correlation）とここで用いた演算との関係は付録 **A.2** に述べた．

次の処理は相関画像のピーク（極大点）の検出である．まず画素値がしきい値（具体的には 40）以上で，かつその 8 近傍より大きい値を持つ画素を歩行者位置の候補として検出する．これら候補位置と半径 2 画素程度の近傍☆にある画素を用いて，最小 2 乗法で 2 次関数をあてはめる．そしてのヘシアンが負値で，極値を与える点が候補位置の 1 画素以内にあるとき，歩行者の位置として採用する．以上のように 2 次関数をあてはめる理由は，8 近傍より大きいスケールでピークの形状を解析し，安定な結果を得るためである．

☆ 正確には標準偏差 1.5 画素のガウス分布の重みを用いている．

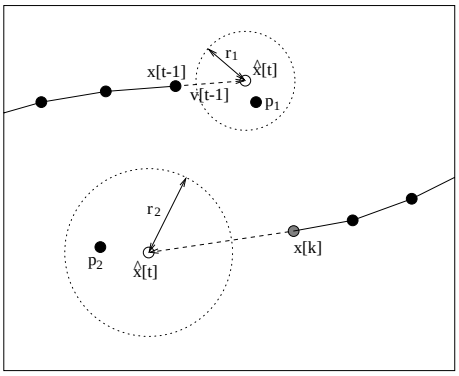


図 9 歩行者の追跡アルゴリズム

Fig. 9 Algorithm for pedestrian tracking.

4. 歩行者追跡

「歩行者追跡」では，動画像の各フレームで検出される歩行者位置を時系列的に接続し追跡する．処理の基本は，過去の移動情報から次のフレームでの位置を予測し，その予測位置に最も近い歩行者位置を対応付けることであるが，歩行者同定の失敗によって生じる歩行者位置の欠落にも対処する必要がある．最初にこのような欠落がない場合に有効な処理について述べ，次に欠落がある場合に有効な処理を述べる．なお，後に述べる実験に使用した画像の大きさは 330×450 画素である．

歩行者位置の欠落がない場合の対応付けを説明した図が**図 9** の上部の図である．追跡中のある歩行者の時刻 t での位置を $\boldsymbol{x}[t]$，速度を $\boldsymbol{v}[t]$ とする．時刻 $t-1$ までの追跡が終了しているとき，時刻 t での予測位置 $\hat{\boldsymbol{x}}[t]$ を次式で計算する．

$$\hat{\boldsymbol{x}}[t] = \boldsymbol{x}[t-1] + \boldsymbol{v}[t-1] \tag{11}$$

そして $\hat{\boldsymbol{x}}[t]$ から一定の半径 r_1（実験では 20 画素）以内にある検出点 p_1 をその歩行者のものとして対応付け，時刻 t での位置 $\boldsymbol{x}[t]$ とする．半径 r_1 以内に複数の検出点が存在した場合には $\hat{\boldsymbol{x}}[t]$ に最も近いものを対応点とする．ここで歩行者の速度 $\boldsymbol{v}[t]$ は，歩行者の出現時には

$$\boldsymbol{v}[t] = \boldsymbol{x}[t] - \boldsymbol{x}[t-1] \tag{12}$$

で計算し，その後は

$$\boldsymbol{v}[t] = c\boldsymbol{v}[t-1] + (1-c)(\boldsymbol{x}[t] - \boldsymbol{x}[t-1]) \tag{13}$$

で計算する．ここで定数 c は，実験では 0.2 とした．

以上のアルゴリズムで歩行者位置 $\boldsymbol{x}[t]$ に対応付ける検出点がない場合，その歩行者位置を消失点として登録する．しかしここで消失点として登録する情報は実際に消失した時点での情報ではなく，d フレーム（実験では 2 フレーム）前の位置 $\boldsymbol{x}[t-d]$ と速度 $\boldsymbol{v}[t-d]$ とする．これは消失する直前の検出結果は不安定になる場合が多いので，そのデータを削除するためである．

次に歩行者位置の検出に欠落がある場合に有効となる処理を図 9 の下部の図を用いて説明する．前記の処理で対応がつかない検出点 p_2 がある場合，一定期間（実験では 12 フレーム）以内に消失点がないか，過去の記録を検索する．ここで検索する期間を制限するのは，あまり時間が経った消失点は同一歩行者である可能性が少ないからである．消失点がある場合には，消失した時刻 k の位置 $\boldsymbol{x}[k]$ と速度 $\boldsymbol{v}[k]$ から次式を用いて現在時刻 t での予測位置 $\hat{\boldsymbol{x}}[t]$ を計算する．

$$\hat{\boldsymbol{x}}[t] = \boldsymbol{x}[k] + (t-k)\boldsymbol{v}[k] \tag{14}$$

そしてその位置から半径 r_2 に出現点 p_2 がある場合にはこれを接続する．ここで半径 r_2 は次式で計算する．

$$r_2 = \max(r_m, \mu|\hat{\boldsymbol{x}}[t] - \boldsymbol{x}[k]|) \tag{15}$$

これは，予測位置が消失点から離れるに従って予測の精度が減少するのに対応するためである．なお上式の定数は，実験では r_m を 20 画素，μ を 0.8 とした．

5. 歩行者追跡実験

提案手法の性能を評価するために 2 つのシーンを用いて歩行者の追跡実験を行った．シーン 1 は約 30 分の長さで雲の移動による照明の変化が激しい．シーン 2 は約 1 時間の長さで，接近した歩行者が多数出現する．なお出現した歩行者の数はシーン 1 では 42 人，シーン 2 では 137 人であった．

シーン 1 から，明瞭な影が生じている状況での追跡結果を**図 10** に示す．図において左の画像から右の画像に向かって時刻が経過しているが，影にまったく影響されずに歩行者が追跡されていることが分かる．**図 11** は同じくシーン 1 からの画像で，立ち止まる歩行者の追跡結果である．歩行者は左から監視領域に進入し，中央の画像に示した位置で約 3 秒立ち止まるが，本手法では問題を生じることなく歩行者追跡が完了する．この歩行者の速度変化を示したグラフが**図 12** である．このグラフの縦軸が速度（pixel/frame 単位），横軸がフレーム番号であるが，本実験では NTSC の画像を 1 枚おきに使用しているため，フレームの時間間隔は 1/15 秒である．フレーム 30 から 75 付近にかけて歩行者は静止している．なお平均的な歩行者の速度を知るために，通常の歩行をしている歩行者 20 人の速度の統計をとったところ，平均が 7.34，標準偏差が 2.49（ともに pixel/frame 単位）であった．照明の変化に関しては，背景差分では画像全体にわたって差分を生じるような照明の変化がシーン 1 全体で 3 回あったが，本手法ではまったく影響がなかった．

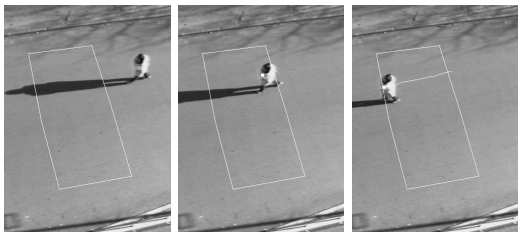


図 10　歩行者の追跡結果（影のあるシーン）
Fig. 10　Tracking of a pedestrian with shadow.

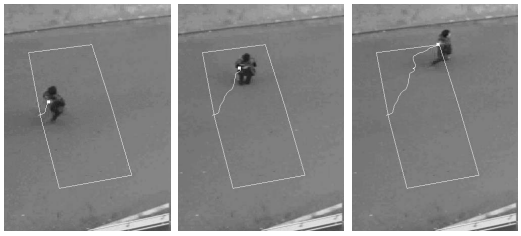


図 11　歩行者の追跡結果（立ち止まる歩行者）
Fig. 11　Tracking of a stopped pedestrian.

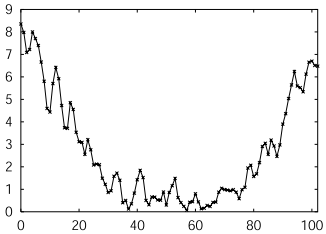


図 12　歩行者の速度（pixel/frame）
Fig. 12　Velocity of a pedestrian (pixel/frame).

次にシーン 2 から，接近した歩行者の追跡例を**図 13** に示す．この図において時刻は左から右，上から下に経過しており，画像は 10 フレーム（0.67 秒）間隔で示されている．左右から 3 人ずつの歩行者が接近し，監視領域内ですれ違うが，右から進入する最も下側の歩行者と左から来る最も上側の歩行者がフレーム 30（図 13 の左下の画像）付近で非常に接近し，このフレームを含む 4 フレームで右から進入する歩行者の同定に失敗する．すなわち，この 4 フレームで検出すべき歩行者数は 5 であるが，同定数は 4 となる．ただし，これは正則化逆フィルタの効果が現れた結果であり，通常の相関ではフレーム 30 付近で前述の歩行者が同

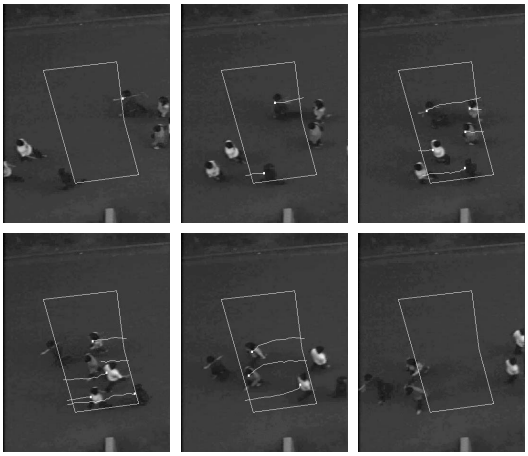


図 13 歩行者の追跡結果（接近した歩行者）
Fig. 13 Tracking of pedestrians walking closely.

表 1 歩行者の追跡成功数
Table 1 Numbers of successfully tracked pedestrians.

状 況	A	B	C	計
人 数	132	41	6	179
成功数	125	36	6	167
成功率	95%	88%	100%	93%

定されないフレーム数は 10 であった．なお図 13 の画像で監視領域内に歩行者が存在するフレームは 48 であるが，本手法および通常の相関とも上記のフレーム以外では正しい数の歩行者が同定された．ここで述べた 4 フレームの歩行者同定の欠落は，それを許す追跡アルゴリズムが有効に働き，再び検出され始めた時点で軌跡が正しく接続されて追跡が成功する．

シーン 1 および 2 全体での追跡結果を**表 1** に示す．ここで単独もしくは複数だが離れて移動する歩行者が状況 A，接近して並走する歩行者が状況 B，反対方向から接近し交差する歩行者が状況 C として，別個に集計してある．ここでは，監視領域への進入から退出まで歩行者を完全に追跡に成功した場合のみを成功と数えている．全体での追跡成功率は 93%であった．追跡失敗の原因は，雲により日が陰り暗くなった状況で，道路表面とほとんど同じ反射率の服を着た歩行者が出現した場合がほとんどである．この場合でも歩行者が検出されるフレームは多くあるが，検出されないフレームが増えるのでその部分の接続ができず，全体としての追跡には失敗している．

6. おわりに

道路を上方からカメラで撮影し，その画像を用いて，そこを通過する歩行者を追跡する処理について報告した．この手法の 1 つの特長は「進入物体検出」において照明の変化に影響されず，影を検出しない利点を持つ 2 台のカメラを用いる方式を使用し，その方式の正しい動作のために必要な正確な幾何学的変換の決定とカメラの感度較正を行う方法を開発したことである．提案手法のもう 1 つの特長は,「歩行者同定」に正則化逆フィルタを採用し，近接した歩行者に対して分離の良い同定を可能にしたことである．実験により照明の変化や影に影響されない追跡が可能であることを確認した．合計約 1 時間 30 分のシーンでの実験では 179 人の歩行者のうち 167 人の追跡に成功し，提案手法は 9 割程度の精度を持っていることを示した．

歩行者追跡の失敗の主な原因は，照明が暗くなった状況で背景に近い反射率の服を着た歩行者が出現し，歩行者が検出されないフレームが続いた場合がほとんどである．本論文では「歩行者追跡」に単純なアルゴリズムを使用したが，これを高度化することで追跡の成功率が高くなると予想される．この点の改善は将来の課題としたい．また本手法を一般の物体検出・追跡処理として見た場合，検出対象を歩行者に限っていることが制限としてあげられる．検出対象をたとえば自動車やバイクにまで広げることも考えていきたい．

謝辞 本研究の一部は文部科学省科学研究費基盤研究 C（2）（No.14580404）および大川情報通信基金によった．

参 考 文 献

1) 高橋政雄，北村忠明，小林芳樹：空間微分および差分処理を用いた車両抽出法，電子通信学会論文誌，Vol.J80-D-II, No.11, pp.2976–2985 (1997).
2) 天本直弘，藤井明弘：画像処理技術による障害物検出と移動物体追跡法，電子通信学会論文誌，Vol.J81-A, No.4, pp.527–535 (1998).
3) 内海 章，大谷 淳，中津良平：画素値分布の推定による時系列画像からの動物体領域抽出，電子通信学会論文誌，Vol.J81-D-II, No.8, pp.1769–1775 (1998).
4) 藤原秀人ほか：テクスチャ背景差分とシルエットモデルを用いた車両追跡手法，第 6 回画像センシングシンポジウム講演論文集 B-1，pp.17–22 (2000).
5) 伊藤 渡ほか：映像監視システムにおける差分用背景画像の照度変化追従特性向上の検討，情報処理学会研究報告，CVIM-112-1, pp.1–7 (1998).
6) Stauffer, C., and Grimson, W.E.L.: Adaptive Background Mixture Models for Real-Time Tracking, *Proc. CVPR'99*, pp.246–252 (1999).
7) 島井博行ほか：移動物体検出のためのロバスト統計を用いた適応的な背景推定法，画像の理解・

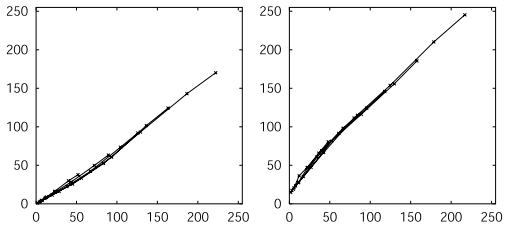
認識シンポジウム 2000 予稿集 II，pp.391–396 (2000).
8) Eveland, C., Konolige, K. and Bolles, R.C.: Background Modeling for Segmentation of Video-Rate Stereo Sequence, *Proc. CVPR'98*, pp.266–271 (1998).
9) 恩田勝政ほか：ステレオ画像処理による侵入者検出技術の開発，第5回画像センシングシンポジウム講演論文集，pp.55–60 (1999).
10) Onoguchi, K.: Shadow Elimination Method for Moving Object Detection, *Proc. ICPR '98*, pp.583–587 (1998).
11) 木山真伸，太田直哉，金谷健一：2台のカメラと射影変換を用いた進入者検出，情報処理学会研究報告，CVIM-118-8, pp.53–58 (1999).
12) 緒方敏博，園田頼信：複数対象物の運動軌跡の抽出，画像の理解・認識シンポジウム 1998 予稿集 I，pp.431–436 (1998).
13) Koller-Meier, E.B. and Gool, L.V.: Modeling and Recognition of Human Actions using a Stochastic Approach, *Proc. 2nd European Workshop on Advanced Video-based Surveillance Systems*, pp.17–28 (2001).
14) Press, H.P., et al.: *Numerical Recipes in C*, Cambridge University Press, Cambridge (1988).
15) 天野 晃，浅田尚紀，馬場雅志：ズームレンズの光学的ひずみの解析と補正，電子通信学会論文誌，Vol.J80-D-II, No.6, pp.1458–1465 (1997).
16) 竹村裕夫：CCD カメラ技術入門，コロナ社 (1997).

付　　録

A.1 画素値の線形補正の妥当性

本論文では2台のカメラ間の画素値の補正に線形式 (9) を用い，これを画像全体の画素に対して適用したが，ここではその妥当性を考察する．

画像の画素値に影響を及ぼす光学的な要因として，口径蝕と $\cos^4\theta$ 現象がある．これらは画像の周辺部が暗くなる効果を持つので，この現象が顕著ならば，本論文で行ったように画像全体に対して同一の変換を施すのは問題がある．しかし一般に口径蝕は望遠レンズ，$\cos^4\theta$ 現象は広角レンズで顕著になるので，標準的な画角のレンズでは問題が少ないこと，また必要ならば補正技術[15]が開発されていることを理由に，本論文ではこれらの要因を無視した．一方 CCD や電子回路の特性は入射光量と画素値の関係を直接的に規定する．本来 CCD は線形な光電変換特性を持っているが，ガンマ補正回路により非線形特性が付加される．ガンマ補正回路はガンマ値が 0.45 になるように設計される[16]ので，2台のカメラで同一のガンマ値が実現されていれば画素値の関係は再び線形になる．各々のカメラのゲインを a_1，a_2，バイアスを b_1，b_2，入射光量を E_1，E_2，共通のガンマ値を γ とすれば，各々の画素値 I_1，I_2 は次式で表される．



(a) Camera 1 vs. 2　(b) Camera 1 vs. 3

図 14 異なるカメラの画素値の関係
Fig. 14 Pixel value relation between different cameras.

$$I_1 = a_1 E_1^{\gamma} + b_1 \tag{16}$$

$$I_2 = a_2 E_2^{\gamma} + b_2 \tag{17}$$

一方，両画素が物体上の同一点に対応していれば E_1 と E_2 は線形に変化するので，ある比例定数を s として

$$E_2 = sE_1 \tag{18}$$

と書ける．したがって I_1 と I_2 の関係は

$$I_2 = \frac{a_2}{a_1} s^{\gamma} I_1 + \left(b_2 - \frac{a_2}{a_1} s^{\gamma} b_1\right) \tag{19}$$

となり，これは線形式 (9) で補正できる．しかし，特に異種のカメラが同一のガンマ値であるかは明らかでなく，さらにガンマ補正回路以外の回路や A/D コンバータの非線形性も検討を要する．そこで異なる種類のカメラ間で画素値の関係を調べる実験を行った．

実験には 3 台のカメラを使用した．カメラ 1 は SONY XC-711，カメラ 2 は SONY XC-75，カメラ 3 は SONY CCD-V200 である．なお，カメラ 1 は本文の実験で使用したものである．カメラ 1 および 2 は産業用カメラで，カメラ 1 はカラー，カメラ 2 は白黒である．カメラ 3 はコンシューマ用のビデオレコーダ一体型のカラーカメラで，オートアイリスはオフに設定した．なお，カラーカメラについては G チャンネルの出力を用いている．実験の手順は以下のとおりである．濃度の異なる矩形をタイル状にならべたパターンを紙に印刷し，照明の強度を変えながら 2 台のカメラで同時に撮影する．そして両者の画像間で，対応する矩形の画素値を比較する．**図 14** の (a) がカメラ 1 対カメラ 2，(b) がカメラ 1 対カメラ 3 の結果で，どちらも横軸がカメラ 1 の画素値，縦軸がもう一方のカメラの画素値である．これらのグラフでは，画像上で同一位置のデータを線でつないで描画してある．これら

表 2 相関演算の周波数空間での表現
Table 2 Frequency domain expressions of correlations.

演 算	周波数空間での表現
相 関	FT^*
位相相関	$\frac{FT^*}{\lvert F\rvert\lvert T\rvert}$
逆フィルタ	$\frac{FT^*}{\lvert T\rvert^2}$ $(=\frac{F}{T})$
正則化位相相関	$\frac{FT^*}{\lvert F\rvert\lvert T\rvert+\eta}$
正則化逆フィルタ	$\frac{FT^*}{\lvert T\rvert^2+\eta}$

の結果には完全ではないが線形性が認められる．さらに本文で行ったのと同様に，これらの結果に最小二乗法で直線をあてはめ，結果の一次式で一方の画素値を変換した後，両者の平方二乗平均誤差を計算した．カメラ 1 対カメラ 2 に関しては変換前の平均誤差が 38.2 レベルであったものが 4.6 レベル，カメラ 1 対カメラ 3 に関しては 24.6 レベルが 5.3 レベルに減少した．これらの結果がすべてのカメラにあてはまるとはいえないが，実験を行った範囲内では異なるカメラ間の画素値に線形性が認められ，画素値の線形補正は有効であると考えられる．

A.2 位相相関と正則化逆フィルタの関係

相関演算のピークを鋭くするための手法として知られているものの 1 つに位相相関（phase correlation）がある．ここではこの手法と本論文で用いた正則化逆フィルタとの関係を述べる．

照合すべき 2 つの画像 f および t のフーリエスペクトルを F および T とする．これら画像の相関演算を周波数空間で表すと FT^* になるが，位相相関とは各スペクトルをその振幅で割り，位相成分だけにして相関演算を行うものである（**表 2**）．これは f が t を (x_0, y_0) だけ平行移動しただけのパターンの場合，

$$\frac{FT^*}{|F||T|} = e^{-2\pi i(ux_0+vy_0)} \tag{20}$$

となって，実空間では位置 (x_0, y_0) の δ 関数なることから，その効果が正当化される．

3 章で行ったように画像照合を画像復元の視点から見れば，位相相関の考え方は逆フィルタに似ている．しかし位相相関が各々の画像の振幅 $|F|$ および $|T|$ で正規化するのに対し，逆フィルタでは一方の画像の振幅 $|T|$ のみで正規化するところが異なる（表 2）．上と同様に，1 つの画像パターンが平行移動した場合のみを考えれば両者は同一の結果を与える．各々の画像パターンが異なる場合には両者の結果は異なり，どちらが好ましいかは一概に結論付けられない．しかし本論文での用途のように，特定の画像パターン t をテンプレートとし，画像 f の中からそのパターンを見つけ出す場合には，テンプレート t を基準としている逆フィルタの方が自然であると考えられる．たとえば入力画像 f の画素がすべて 0 の場合，位相相関では演算が破綻する．

本論文で用いた正則化逆フィルタは，ウイナーフィルタのように逆フィルタに正則化の効果を付け加えたものである．これと同様に位相相関に正則化項を付け加えた正則化位相相関とも呼ぶべき演算も考えることができる（表 2）．しかし本論文での用途の場合には，上記と同様の理由でテンプレート t を基準とする正則化逆フィルタの方が自然であると考えられる．

(平成 14 年 11 月 5 日受付)
(平成 15 年 9 月 5 日採録)

太田 直哉（正会員）
1985 年東京工業大学総合理工学研究科物理情報工学専攻修士課程修了．1987 年日本電気（株）入社．C&C 情報研究所に在籍．1994 年より群馬大学工学部助手．1998 年より同助教授．コンピュータビジョン，パターン認識の研究に従事．博士（工学）．1991 年〜1992 年米国 MIT メディア研究所客員研究員．1998 年豪州アデレード大学客員研究員．電子情報通信学会，IEEE コンピュータソサエティ各会員．

木山 真伸
1999 年群馬大学工学部情報工学科卒業．2001 年同大学院博士前期課程修了．同年三洋 LSI デザイン・システムソフト（株）入社．群馬大学に在学中，コンピュータビジョンの研究に従事．

吉川 健一
2000 年群馬大学工学部情報工学科卒業．2002 年同大学院博士前期課程修了．同年富士写真フィルム（株）入社．群馬大学に在学中コンピュータビジョンの研究に従事．