

# 断片的な動画像の対応付けを利用した歩行者認識

## Walking Person Recognition by Matching Video Fragments

西山 正志† 湯浅 真由美† 若杉 智和† 柴田 智行† 山口 修†

Masashi Nishiyama†, Mayumi Yuasa†, Tomokazu Wakasugi†,

Tomoyuki Shibata†, Osamu Yamaguchi†

†(株) 東芝 研究開発センター, 川崎市

†Corporate Research and Development Center, TOSHIBA Corporation

E-mail: masashi.nishiyama@toshiba.co.jp

### Abstract

本稿では、動画像を用いた識別手法による歩行者認識に対して、複数のカメラから得られた複数人の顔画像を段階的に人物毎に対応付けし、統合された一つの動画像を用いて個人を識別する手法を提案する。一枚の顔画像で歩行者を識別する場合、照明、顔の向き、人数の影響で識別性能が低下する。そこで、最初に各カメラにおいて顔画像を人物毎に対応付けし断片的な動画像を生成する。次に、カメラ間で断片的な動画像を対応付ける。入力段階で顔の見え方の変動を人物毎に獲得できるため、動画像を用いた識別手法の効果が発揮され、高い識別性能を得ることができる。提案手法の有効性を、349人の歩行者データベースを用いて複数人物が同時に歩行する状況を模擬した識別実験で確認した。

### 1 はじめに

顔画像による個人識別は、生体情報をシステムに非接触で入力できるため利便性が高く、ユーザの心理的な負担が少ない [1]。我々は従来から動画像を用いた個人識別の手法 [2] を提案している。この手法は、刻々に変化する顔の見え方を動画像として入力し、辞書と比較をする。これにより、入力を一枚の顔画像とする場合と比べて、顔向きや表情の変化に頑健な識別を行うことができる。この手法を応用した顔照合システム FacePass [3] は入退室管理を目的としていたため、図 1(a) 左のように、カメラの前に歩行者が立ち止まるという動作が必要であった。また、歩行者顔照合システム FacePassenger [4] では、図 1(a) 右のように、歩きながらカメラへ顔を意識的に向けるという動作が必要であった。このように、従来のシステムはゲートの開閉が目的であったが、本研究では、図 1(b) のように、特定のエリア内にいつ誰

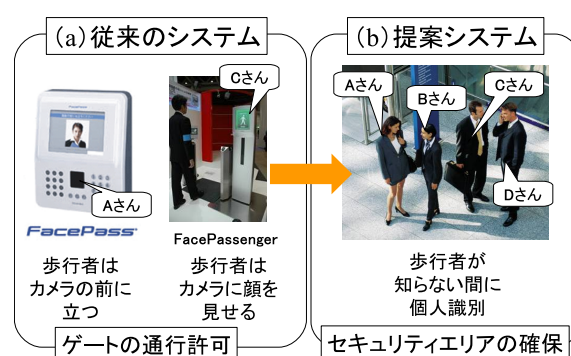


図 1 本研究で目指すシステムの概念図

が現れたかといったように、セキュリティエリアを確保するためのシステム構成を目指す。この場合、歩行者の協力を仰ぐことなくカメラの撮像範囲内に現れるだけで識別することができる方法が必要となる。

本研究が目指すシステムでは、例えば図 2 のように、通路に設置された複数のカメラで複数の歩行者の顔画像を同時に捉え照合を行う。各カメラで獲得された一枚の顔画像による照合は、入力と辞書の間で照明、顔の向きが大きく異なるため難しい。また、辞書として登録されている人数が増えるにつれ誤識別が発生する可能性が高くなる。このような状況でも高い識別性能を得るためには、一枚の顔画像だけを入力に用いるのではなく、同一人物の顔画像を一つの動画像として統合し、顔の見え方の変動を獲得した上で照合を行うことが有効である。歩行者の顔画像を統合することは、同時刻に同じ場所を歩く人物間での対応付け問題に帰着できる。

そこで、本稿では、図 3 のように、複数のカメラから得られる複数人の顔画像を段階的に動画像へ統合し個人を識別する手法を提案する。提案手法では、最初に各カメラにおいて人物毎に対応付けることで断片的な動画像を生成する。次に、カメラ間において断片的

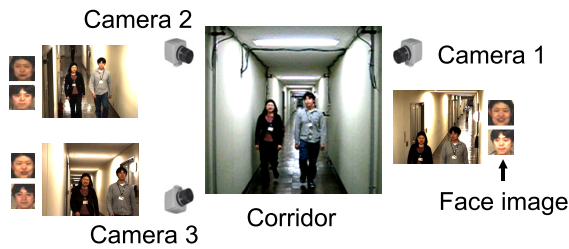


図2 複数カメラによる複数の歩行者の個人識別

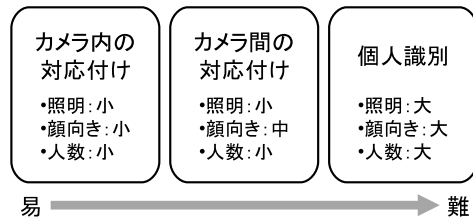


図4 動画像の統合における変動要因の影響

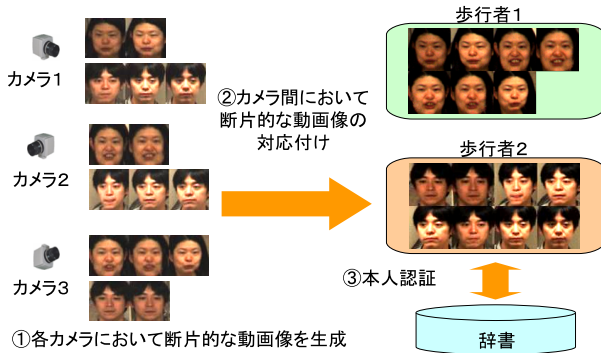


図3 段階的な対応付けによる動画像の生成

な動画像を対応付ける。この段階的な対応付けは、一枚の顔画像で照合を行う場合に比べて、図4のように、照明、顔の向き、特に人数の影響を抑えた上で処理することができる。提案手法の新規性は、動画像を用いた識別手法を適用するために、複数の歩行者の個人識別という全体問題を、比較的解くことが容易な対応付けの部分問題へ段階的に分け、最終的に統合する枠組みにある。

提案手法では段階的な対応付けを、画像間のパターンマッチングに基づく顔認識の手法のみで行う。これは、厳密なカメラキャリブレーションを行い、人物の3次元位置を追跡し対応付けを行う手法[12]と比べ、システムを導入するための時間や費用を抑えることができる。また、検出と追跡の状態遷移を考えるとなく、顔の隠れの問題にも容易に対応できる。さらに、カメラのフレームレートが低い場合でも対応できるという利点がある。

以下、2で従来手法について述べ、3で提案手法の段階的な対応付けについて述べる。次に、4で対応付けに用いる正規化された顔画像の生成について述べる。最後に、5で大規模人数が登録された歩行者データベースを用いた識別実験で提案手法の有効性を確認する。

## 2 従来手法

本章では、従来手法に関して、人物が歩行するか静止するかという動作条件、対象人物の数、そしてカメラの台数といった観点で整理する。

1台のカメラを用いて1人の歩行者の個人識別を行う手法が提案されている[5, 6, 7]。[5, 6]では、動画像から検出された顔画像の顔向きを推定し、正面向きの顔画像を選択し識別する。[7]では、動画像同士を比較することを目的とし追跡処理の中で識別処理も同時に行う。これらの手法では、歩行者が自由に通過する場合、カメラと顔が正対していない場合が発生し、識別に有効な正面向きの顔画像が得られるとは限らない。

1台の魚眼レンズ付きカメラを用いて、着席している複数人物の個人識別を行う手法が提案されている[8]。この手法では、中央を向く着席した人物の顔毎に動画像を生成し個人識別を行う。同様に、1台の魚眼レンズ付きカメラを用いて1人の歩行者の個人識別を行う手法が提案されている[9]。魚眼レンズ付きカメラは広範囲を撮像できるが、識別に有効な正面向きの顔画像が得られるとは限らない。

複数のカメラを用いて、静止した1人の人物の個人識別を行う手法が提案されている[10, 11]。[10]では、推定された顔向きに対応した辞書を用いて個人識別を行う。[11]では、獲得された全ての顔画像を一つの辞書に統合し識別する。また、複数のカメラを用いて1人の歩行者を識別する手法が提案されている[13, 14]。[13]では、ステレオカメラを用いて歩行者の顔領域の位置と姿勢を推定し識別する。[14]では、様々な顔向きの顔画像を標準顔モデルを用いて仮想的に生成し初期辞書に登録する。カメラから獲得された顔画像の顔向きを推定し辞書を更新する。ただし、[13, 14]ではカメラキャリブレーションが必要となる。[10, 11, 13, 14]では、複数のカメラを用いることで、ある1台のカメラでは顔画像を検出できなくとも、他のカメラで補うことができる。しかし、これらの手法は、識別対象をカメラの前の1人としており、複数人の顔画像が同時に獲得できる場合、どのように識別するかを検討していない。

本手法と同様に、複数のカメラを用いて複数の歩行者の個人識別を行う手法も提案されている[12]。この

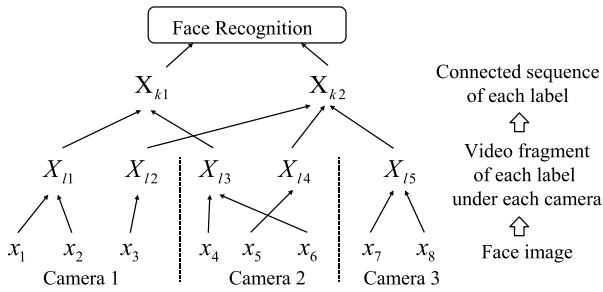


図5 段階的な対応付けの流れ

手法では、複数の歩行者の顔の様々な見え方を登録するために、検出と追跡のタスクをそれぞれのカメラに動的に割り当て顔画像の集合を生成する。各カメラから得られる人物毎の顔画像を対応付けるために、3次元的な追跡処理を用いる。3次元的な追跡を精度よく行うために厳密なカメラキャリブレーションが要求される。また、運用中に何らかの原因でカメラの位置がずれると追跡処理が破綻し識別性能が低下する。

### 3 段階的な対応付け

複数カメラを用いて複数の歩行者を、動画像を用いた識別手法で個人識別するために、カメラキャリブレーションを行うことなく顔画像を段階的に対応付け、人物毎の動画像を生成する方法について述べる。

#### 3.1 段階的な対応付けの枠組み

最初に各カメラにおいて顔画像を対応付けし断片的な動画像を生成する。断片的な動画像を式(1)で定義する。

$$X_l \equiv \{x_i \mid M_1(x_i) = l, i = 1, \dots, N\} \quad (1)$$

ここで、 $x$  は1枚の顔画像、 $M_1$  は顔画像に対してラベルを返す関数、 $l$  は断片的な動画像に付けられたラベル、 $N$  は獲得された顔画像の枚数を表す。関数  $M_1$  については3.2節で述べる。次に、カメラ間で断片的な動画像を対応付けし、個人識別で用いる統合された動画像  $X$  を生成する。 $X$  は式(2)で定義される。

$$X_k \equiv \{X_j \mid M_2(X_j) = k, j = 1, \dots, M\} \quad (2)$$

ここで、 $M_2$  は断片的な動画像に対してラベルを返す関数、 $k$  は統合された動画像に付けられたラベル、 $M$  は獲得された断片的な動画像の個数を表す。関数  $M_2$  については3.3節で述べる。図5に、三台のカメラの下で、二人の人物が歩行したときに段階的に対応付けされる流れを示す。

実システム上では、顔画像は時間の経過と共に順に獲得される。各カメラにおいて断片的な動画像を生成するために、顔画像  $x$  が獲得される毎に関数  $M_1$  でラ

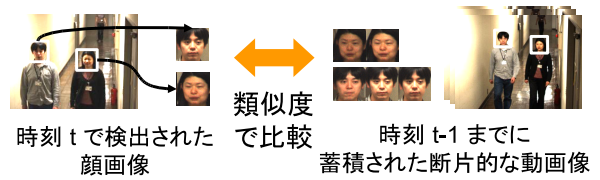


図6 カメラ内における断片的な動画像の生成

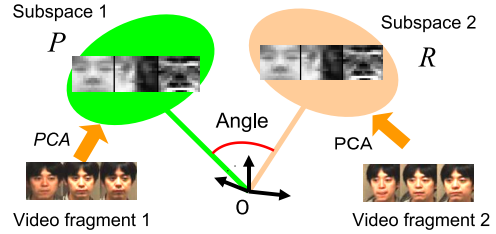


図7 相互部分空間法による断片的な動画像同士の比較

ベルを判定する。同じラベルをもつ  $x$  を断片的な動画像  $X$  に加える。一定の時間  $T_1$  以上新たな顔画像が追加されなかった断片的な動画像  $X$  は通過した人物と判定し、カメラ間の断片的な動画像の対応付けへ進む。 $X$  のラベルを関数  $M_2$  で判定し、同じラベルをもつ断片的な動画像  $X, X'$  を統合する。一定の時間  $T_2$  を経過した断片的な動画像は対応付けが終了したと判断し、統合された動画像  $X$  とする。この  $X$  を用いて個人識別を行う。

#### 3.2 断片的な動画像を生成するためのラベル付け

各カメラで獲得された顔画像  $x$  は、関数  $M_1$  により、図6のように、同じカメラにおいて蓄積された断片的な動画像と対応付けられる。対応付ける際には、断片的な動画像に属する最新の顔画像  $\bar{x} \in \bar{X}$  と  $x$  との間で式(3)の類似度  $S$  を算出する。

$$S = \frac{S_{simple}}{1 + \alpha(t - \bar{t})} \quad (3)$$

ここで、 $S_{simple}$  は  $x, \bar{x}$  間の単純類似度、 $\alpha$  は定数、 $t, \bar{t}$  は  $x, \bar{x}$  が獲得された時間を表す。単純類似度は、 $S_{simple} = \cos^2 \theta$  で定義される。 $\theta$  は、顔画像をラスタースキャンすることで変換されたベクトル同士のなす角度を表す。

関数  $M_1$  は、閾値  $S_1$  を越え最も高い類似度が算出された断片的な動画像のラベルを返す。また、算出された全ての類似度が  $S_1$  未満の場合、新たな人物が表れたと判定し、新たなラベルを返す。対応付ける断片的な動画像が1個も蓄積されていない場合も新たなラベルを返す。

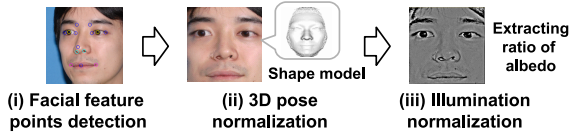


図 8 正規化された顔画像の生成

### 3.3 断片的な動画画像を対応付けるためのラベル付け

関数  $M_2$  により、カメラ間で断片的な動画画像同士を対応付ける．断片的な動画画像間の類似度  $S'$  を算出するために、動画画像同士を比較できる図 7 の相互部分空間法を発展させた直交相互部分空間法 (OMSM: Orthogonal Mutual Subspace Method)[15] を用いる．OMSM は、相互部分空間法の前処理として、人物間における顔の見え方変動の差を強調する線形変換を用いる．OMSM の計算方法は次節で述べる．関数  $M_2$  は、閾値  $S_2$  を越え最も高い類似度が算出された断片的な動画画像のラベルを返す．また、算出された全ての類似度が  $S_2$  未満の場合は新たなラベルを返す．

### 3.4 動画画像間の類似度の計算方法

OMSM を適用するために、顔画像の集合  $X$  に対して主成分分析を適用し部分空間を生成する．主成分分析を適用する際は自己相関行列 [16] を用いる．直交化行列  $O$  で線形変換された二つの部分空間を  $P, Q$  とすると、 $P$  と  $Q$  との間の類似度  $S'$  は、正準角と呼ばれる二つの部分空間がなす角度  $\theta$  により式 (4) で決定される．

$$S' = \cos^2 \theta \quad (4)$$

部分空間同士に共通するベクトルが存在すれば  $\theta = 0$  である． $\cos^2 \theta$  は、以下の行列  $R$  の最大固有値となる．

$$Ra = \lambda a \quad (5)$$

$$R = (r_{mn}) \quad (m, n = 1 \dots D_P) \quad (6)$$

$$r_{mn} = \sum_{l=1}^{D_Q} (\psi_m, \phi_l)(\phi_l, \psi_n) \quad (7)$$

ここで、 $\psi_m, \phi_l$  は部分空間  $P, Q$  の  $m, l$  番目の基底ベクトル、 $(\psi_m, \phi_l)$  は  $\psi_m$  と  $\phi_l$  の内積、 $D_P, D_Q$  は部分空間  $P, Q$  の基底ベクトルの本数を表す．ただし、 $D_P \leq D_Q$  とする．

統合された動画画像と登録されている辞書動画画像のマッチングにおいても、同様の手法を適用する．これにより、識別するための動画画像の生成と個人照合とをシームレスにつなげる．

## 4 顔画像の正規化方法

### 4.1 顔向きと照明変動の補正

これまで述べたように、段階的な対応付けにパターンマッチングの方法を用いる．この際問題になるのが、

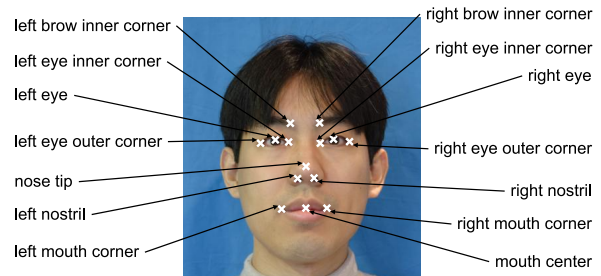


図 9 14個の顔特徴点

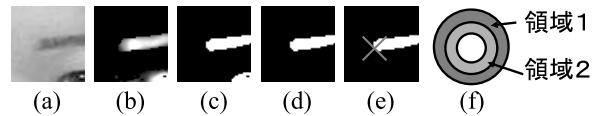


図 10 眉内端検出の処理の例．(a) 探索範囲、(b) 分離度マップ、(c) 2値化後、(d) 領域判定後、(e) 端点検出、(f) 分離度フィルタ

識別に有効な顔画像の解像度、カメラ位置の違いによる相対的な顔向きの変化、歩行による相対的な照明条件の変化である．本稿では、これを緩和するために顔画像  $x$  を生成する際、図 8 のように、(i) 一定以上の解像度をもつ顔領域に対して顔の特徴点を検出し、(ii) 3次元形状モデルを用いて顔向き正規化 [17] を適用し、(iii) 照明条件に影響されない拡散反射率の比を抽出する照明正規化 [18] を適用する．

### 4.2 顔特徴点の検出

正規化した顔画像  $x$  を生成するために、カメラで獲得された画像から顔領域と顔特徴点を検出する．3次元形状モデルを用いた顔向き補正を行うには、多数の顔特徴点が必要であり、図 9 のように、瞳、鼻孔、目尻、目頭、口端、眉内端、鼻頂点、口中点の計 14 点を検出する．最初に画像全体から Joint Haar-like 特徴と AdaBoost を用いた方法 [19] により複数の顔領域を検出する．次に、それぞれの顔領域から分離度フィルタとパターン認識による方法 [20] により、瞳と鼻孔を検出する．検出された瞳位置を基準として、目尻、目頭、口端をコーナー検出とパターン認識による方法 [21] により検出する．また、顔の左右中心と唇のエッジから口中点を検出する．最後に眉内端と鼻頂点を検出する．

### 4.3 修正分離度を用いた眉内端の検出

眉は人によってその境界が明確でなく色が薄いといった場合があり、単純な処理では検出が難しい．また、形状の変動も大きいため、パターン照合などを用いる方法も難しい．そこで、円形分離度フィルタ [20] が円形の形状のみならず、一定の幅をもつような形状も抽出が可能であることを利用して、眉内端を検出する．

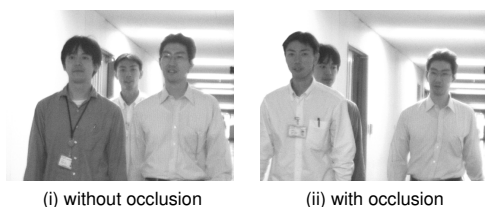


図 11 断片的な動画の生成するための撮影画像の例

まず、検出された瞳位置を基準として、図 10(a) のように、探索範囲を設定する．探索範囲内で円形の修正分離度フィルタに基づく分離度マップ (b) を作成する．修正分離度フィルタは、図 10(f) の円形分離度フィルタ [20] と形状は同じであるが、2 領域の平均値の差を分離度値に加える処理を行うことにより、画素値が低い内側領域だけを取り出すことができる．修正分離度  $\eta'$  は式 (8) で定義される．

$$\eta' = \eta + \beta(\overline{P_1} - \overline{P_2}) \quad (8)$$

ここで、 $\eta$  は通常分離度値、 $\overline{P_1}$ 、 $\overline{P_2}$  は領域領域 1, 2 における輝度平均値、 $\beta$  は定数とする．

分離度マップを 2 値化し (c)、2 値画像のラベリング処理により領域を分割し、領域毎に眉内端の判定を行う (d)．例えば左眉内端の場合の判定条件は、各分割領域が探索領域の右端に接し、かつ、上端、下端、左端に接しないこととする．この条件を満たした候補のうち、領域の左端の点が推定点に最も近いものを選択する (e)．推定点はあらかじめ取得した眉内端位置の統計的な分布を元に決定しておく．

#### 4.4 鼻頂点の検出

本稿では、鼻頂点は光の反射により周辺と比較して輝度が大きくなると仮定し、この仮定を満たす点を鼻頂点として検出する．まず、鼻孔の位置を基準とした探索範囲内で通常分離度フィルタ [20] のピーク位置と、前節で述べた円形の修正分離度フィルタによる分離度のピーク位置を検出しする．その後、前者のピーク位置のうち、後者のピーク位置に近いものを除去する．これにより、鼻孔などの輝度の低い領域を除外することができる．残ったピーク位置のうち最も分離度値の高いものを鼻頂点位置として検出する．

## 5 実験

### 5.1 1 台のカメラにおける断片的な動画の生成

各カメラにおいて、どのような断片的な動画が生成されるかについて実験した．ここではカメラ 1 台を用い、図 11(i),(ii) のように、廊下を 3 人が同時に歩く

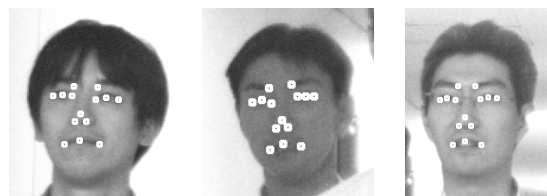


図 12 検出された顔特徴点の例

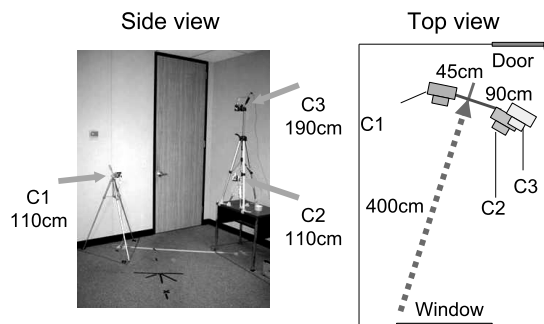


図 13 評価データベースの撮影環境

場合について実験を行った．(i) では 3 人の顔が全てのフレームで映るよう歩くことで隠れを発生させず、(ii) ではお互いが前面に回り込むことで隠れを意図的に発生させた． $1024 \times 768$  pixels の解像度で秒間 7.5 フレームで撮影した．1 つの動画の撮影時間は約 7 秒間であった．各動画に対して顔画像を自動で検出した．検出された 14 の顔特徴点の例を図 12 に示す．大きさが  $64 \times 64$  の顔画像を生成するために、顔領域を検出し顔の特徴点を用いて顔向き正規化 [17] を適用し、照明正規化 [18] を行った後、ダウンサンプリングすることで 1024 次元のベクトルへ変換した．

それぞれの撮影画像から顔画像は (i) で 76 枚、(ii) で 59 枚検出された．他人と誤って対応付かないように閾値  $S1$  を設定したため、一人につき複数の断片的な動画が生成された．このうち、断片的な動画を構成する顔画像が最大枚数のものを挙げると、(i) で人物 A が 19 枚、人物 B が 5 枚、人物 C が 4 枚、(ii) で人物 A が 7 枚、人物 B が 8 枚、人物 C が 11 枚であった．低解像度の顔画像は、本人同士であっても類似度が低いため対応付きにくいという傾向が見られた．隠れが発生する (ii) でも少数枚の顔画像からなる断片的な動画が生成できることが確認できた．

### 5.2 断片的な動画の対応付けの評価

次に、カメラ間において断片的な動画が対応付けできるかを確認するために実験を行った．複数人が同時に歩く場合は、人の組み合わせなど様々な状況が考えられるため、歩行者が 1 人のみで撮影されたデータベースを用いて模擬実験を行った．以下では、各カメラにおいて断片的な動画は生成できていると仮定し

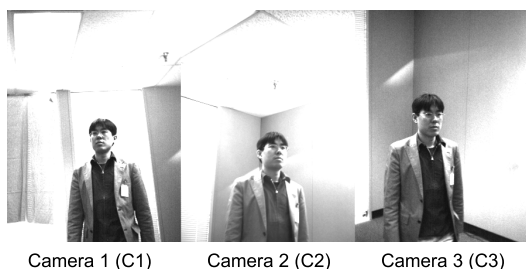


図 14 評価データベースのサンプル画像

表 1 (i) 各動画像において検出された顔画像の平均枚数と (ii) 349 × 2 個の動画像の中で検出された顔画像の枚数が 7 枚未満の動画像が含まれる割合 (%) .

Camera	(i)	(ii)
C1	14.8	12.0
C2	20.4	5.2
C3	19.3	5.4
All	54.5	1.6

て評価を行う．まず，複数のカメラによる歩行者認識の有効性を示すために歩行者が 1 人の場合の実験について述べ，次に，提案手法の有効性を示すために歩行者が複数人の場合の模擬実験について述べる．

評価データベースと評価基準．人物が歩く様子を撮影した動画像を 349 人について収集した．3 つのカメラ (C1, C2, C3) を図 13 のように配置した．窓付近をスタート位置とし，扉に向かって歩く動画像を各人物につき 2 回撮影した．識別実験を行うために，一方の動画像を辞書とし，もう一方の動画像を入力とした．各人物には図中の破線矢印上を，各カメラを見ることなく顔を進行方向へ向けて歩いてもらった．各カメラの動画像は，768 × 1024 pixels の解像度，秒間 15 フレームで撮影した．1 つの動画像の撮影時間は約 4 秒間であった．図 14 に撮影された動画像の一部を示す．

各カメラの動画像において，検出された顔画像の平均枚数を表 1(i) に示す．それぞれの動画像には 1 人しか映っていないため，検出された全ての顔画像から一つの断片的な動画像を構成した．5.1 節の実験と比べて断片的な動画像を構成する枚数が多いのは，このデータベースではフレームレートが倍になっていることが一つの要因であると考えられる．表中の All では，カメラ C1, C2, C3 から得られた全ての顔画像を用いて，統合された動画像を構成した．

この実験では，動画像を用いて個人識別をするために直交相互部分空間法を用いる．その際の部分空間の

表 2 1 人のみが歩く場合の識別性能

Camera	CMR(%)	EER(%)
C1	81.4	16.0
C2	92.6	7.2
C3	91.7	7.4
All	97.7	2.0

基底ベクトルの本数は 7 とし，基底ベクトルの次元数は 1024 とした．これらのパラメータは実験的に決定した．直交化行列は，それぞれの人物の辞書部分空間から生成した．動画像を構成する枚数が 7 枚未満の場合は，生成される部分空間の次元数が 7 に満たない．この場合は，識別実験において本人類似度と他人類似度を全て 0 とした．表 1(ii) に，349 × 2 個の動画像の中で検出された顔画像の枚数が 7 枚未満の動画像が含まれる割合を示す．C2, C3 に比べて C1 の平均枚数が低い理由として，歩行者が通過する場所の近くにカメラを設置したため，相対的な顔向きの変動が大きく，顔領域検出と顔特徴点検出が難しいことが考えられる．識別性能の評価には以下の 2 つの基準を用いた．

1. 一位正解率 (CMR:Correct Match Rate)  
辞書に登録された人物の中で本人との類似度が最も高くなる割合を表す．
2. 等価エラー率 (EER:Equal Error Rate)  
FAR(他人受理誤り率) と FRR(本人排除誤り率) が等しい時の割合を表す．FAR 以下の式で求まる．

$$FAR = \frac{\text{他人類似度がしきい値以上の試行数}}{\text{全試行数} - \text{本人の試行数}} \quad (9)$$

一方，FRR は以下の式で求まる．

$$FRR = \frac{\text{本人類似度がしきい値以下の試行数}}{\text{本人の試行数}} \quad (10)$$

実験結果 (1 人のみが歩行する場合)．ここでは同時に 1 人のみが歩く場合について，1 台のカメラで個別に個人識別を行った場合と，全てのカメラを使って個人識別を行った場合の性能を比較する．1 人のみ場合は，カメラ間とカメラ内の顔画像の対応付けにおいて誤対応が発生せず，理想的な状況での識別性能を推定することができる．表 2 に結果を示す．C1, C2, C3 では各カメラから得られた断片的な動画像からそれぞれ辞書部分空間を生成し，All では統合された動画像から辞書部分空間を作成した．C2, C3 に比べて C1 の識別性能が低い理由は，表 1(ii) の検出された顔画像が 7 枚未満の平均人物数が多いためである．実験結果より，全てのカメラを使った場合の All が，一台のカメラのみを

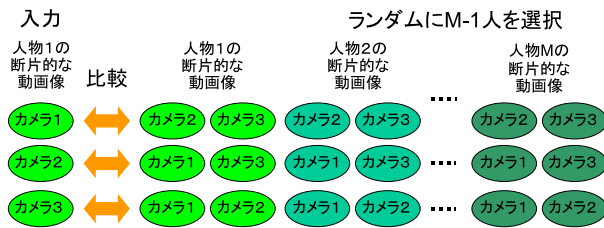


図 15 同時に複数人が歩く場合の実験条件

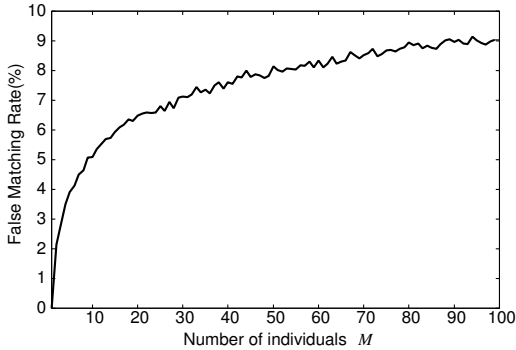


図 16 同時に歩く人数が変化した時の断片的な動画の誤対応率

使った場合 C1, C2, C3 に比べて大きく識別性能が改善されていることが確認できる。

実験結果 (同時に複数人が歩行する場合) . 同時に複数の人物が歩行する場合について, 先程と同じデータベースを用いて模擬実験を行った.  $M$  人が同時に歩行する状況を模擬するために, 1 人の人物に対してランダムに  $M - 1$  人を選択した. 図 15 のように, あるカメラの断片的な動画を入力とし, 残りの断片的な動画と比較することで対応付け実験を行った. この対応付け実験を 349 人全てに対して行った. 人の選び方による実験結果の変動を軽減するため,  $M - 1$  人の選択を 10 回繰り返した. 入力された断片的な動画を,  $M \times 2$  個の断片的な動画の中で, 最も類似度が高くなったものと対応付けた. この実験では, 各人物について 2 台以上のカメラで断片的な動画が生成できていると仮定し,  $S2$  は 0 とした. 誤って他人と対応した割合を誤対応率とし,  $M$  を増やしたときの変化を図 16 に示す. この結果より, 対応付ける人数が少ない場合には誤対応率を低く抑えることができるといえる.

次に, この対応付けられた動画を入力部分空間とし, 識別実験を行った. 辞書部分空間生成するためには, 表 2 の All と同じ 349 人の統合された動画を用いた.  $M$  を増やしたときの CMR の推移を図 17, EER の推移を図 18 に示す. (i) は断片的な動画の対応付けを行わなかった場合を想定し, 表 2 の C1, C2, C3 の入力部分空間を用いた. (ii) は提案手法により自動的に

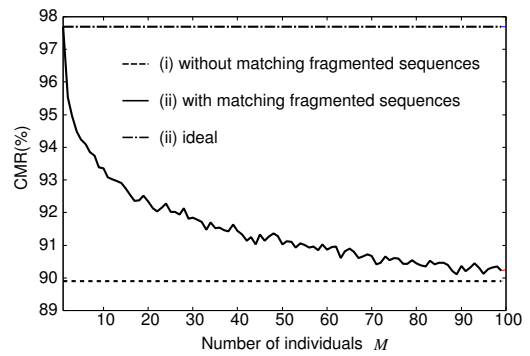


図 17 同時に歩く人数が変化した時の一位正解率

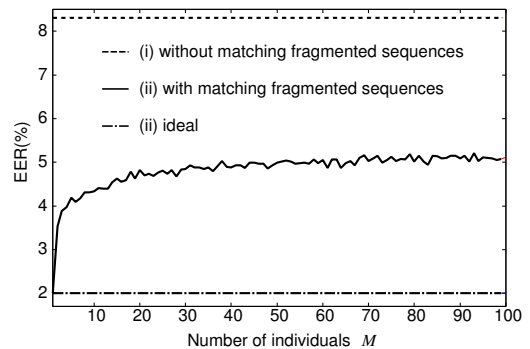


図 18 同時に歩く人数が変化した時の等価エラー率

対応付けを行った場合である. (iii) は手動で対応付けを行った場合で, 表 2 の All と同じである. 図 16 の誤対応率と連動して CMR と EER が変化していることが分かる. EER が (i) と (ii) で大きく違う原因として, 7 枚未満で構成される動画の数が違うことが考えられる. また, 誤対応であっても似たものが統合されたため結果として本人類似度が高くなったということも考えられる. 理想的な結果 (iii) と (ii) を比べると差があるが, 同時に歩く人数が 10 人未満であれば, (i) と比べて CMR と EER とともに大きく改善されている. 以上, 実験結果により, 提案手法の有効性を確認できた.

## 6 おわりに

本稿では, 複数のカメラが設置された環境において, カメラキャリブレーションや追跡処理を行わずに, 顔画像同士の対応付けのみで複数の歩行者を識別する手法を提案した. 複数のカメラから獲得される顔画像を, 断片的な動画として段階的に対応付ける方法を述べた. また, 正しく対応付けるために顔画像を正規化する方法について述べた. 提案手法の有効性を 349 人の歩行者データベースを用いた模擬実験により確認した.

例えば5人が同時に歩いているシーンを模擬した場合、一位正解率が89.9%から94.2%、等価エラー率が8.3%から4.2%に改善された。

今後の課題として、断片的な動画像の生成を大規模人数による模擬実験により評価すること、一人につき複数の断片的な動画像が生成される場合について評価すること、実運用するために各種パラメータをどのように設定するかを考察することが挙げられる。

## 参考文献

- [1] 赤松 茂, “コンピュータによる顔の認識 –サーベイ–,” 信学論 D-II Vol. J80-D-II, No. 8, pp. 2031-2046, 1997
- [2] 西山 正志, 山口 修, 福井 和広, “多重制約相互部分空間法を用いた顔画像認識,” 電子情報通信学会論文誌 D-II Vol. J88-D-II, No. 8, pp. 1339 - 1348, 2005.
- [3] 佐藤 俊雄, 助川 寛, 横井 謙太郎, 土橋 浩慶, 緒方 淳, 岡崎 彰夫, “立ち位置変動を考慮した顔照合セキュリティシステム「FacePass」の開発,” 映像情報メディア学会誌 Vol. 56, No.7, pp.1111-1117, 2002
- [4] 滝沢 圭, 長谷部 光威, 助川 寛, 佐藤 俊雄, 榎本 暢芳, 入江 文平, 岡崎 彰夫, “歩行者顔照合システム「FacePassenger」の開発,” FIT2005 I-010 pp.27-28, 2005.
- [5] 鹿毛 裕史, 羽島 一夫, 三輪 祥太郎, 橋本 学, M. Jones, J. Thornton, “ロバスト顔追跡によるベストショット顔画像記録システム,” 第10回画像センシングシンポジウム講演論文集, pp. 541 - 546, 2004.
- [6] Z. Yang, H. Ai, B. Wu, S. Lao, and L. Cai, “Face Pose Estimation and its Application in Video Shot Selection,” International Conference on Pattern Recognition 2004, pp. 322 - 325, 2004.
- [7] R. Chellappa, V. Kruger, and S. Zhou, “Probabilistic Recognition of Human Faces from Video,” The IEEE International Conference on Image Processing, Vol. I, pp. 41 - 44, 2002.
- [8] K. S. Huang, and M. M. Trivedi, “Streaming Face Recognition using Multicamera Video Arrays,” International Conference on Pattern Recognition 2002, pp. 213 - 216, 2002.
- [9] 小原 ゆう, 八木 康史, 横山 太郎, 谷内田 正彦, “全方位画像列からの個人識別,” 情報処理学会論文誌:コンピュータビジョンとイメージメディア, Vol. 43, No. SIG 4(CVIM 4) , pp.95 - 104, 2002.
- [10] 安本 護, 本郷 仁志, 渡辺 博己, 山本 和彦, 奥水 大和, “マルチカメラ統合を用いた人物識別と顔方向推定,” 電子情報通信学会論文誌 D-II, Vol.J84-D-II, No.8, pp.1772-1780, 2001.
- [11] 小坂谷 達夫, 山口 修, 福井 和広, “マルチカメラ動画像を用いた顔画像認識,” 第8回画像センシングシンポジウム講演論文集, pp. 319 - 324, 2002.
- [12] 加藤 丈和, 向川 康博, 尺長 健, “安定な顔認識のための分散協調登録,” 電子情報通信学会論文誌 D-II, Vol.J84-D-II, No.3, pp.500-508, 2001.
- [13] J. G. Wang, R. Venkateswarlu, and E. T. Lim, “Face tracking and recognition from stereo sequence,” 4th International Conference on Audio- and Video-based Biometric Person Authentication, pp. 145 - 153, 2003.
- [14] 田中 秀典, 北原 格, 斎藤 英雄, 村瀬 洋, 小暮 潔, 萩田 紀博, “複数視点映像における被写体の姿勢変動を考慮した見え方学習法,” 電子情報通信学会 信学技報 PRMU2005-268, pp. 61 - 68, 2006.
- [15] 河原 智一, 西山 正志, 山口 修, “直交相互部分空間法を用いた顔認識,” 情報処理学会 コンピュータビジョンとイメージメディア研究会 2005-CVIM-151 (3), pp. 17 - 24, 2005.
- [16] E. Oja, “Subspace Methods of Pattern Recognition,” Research Studies Press, England, 1983
- [17] T. Kozakaya, and O. Yamaguchi, “Face Recognition by Projection-based 3D Normalization and Shading Subspace Orthogonalization,” 7th International Conference Automatic Face and Gesture Recognition, 2006.
- [18] M. Nishiyama, and O. Yamaguchi, “Face Recognition Using the Classified Appearance-based Quotient Image,” 7th International Conference Automatic Face and Gesture Recognition, 2006.
- [19] T. Mita, T. Kaneko, and O. Hori, “Joint Haar-like Features for Face Detection,” Tenth IEEE International Conference on Computer Vision 2005, pp.1619 - 1626, 2005.
- [20] 福井和広, 山口修, “形状抽出とパターン照合の組合せによる顔特徴点抽出,” 信学論 (D-II), Vol. J80-D-II, No. 8, pp. 2170 - 2177, Aug. 1997.
- [21] 武口智行, 湯浅真由美, 山口修, “角点を持つ顔特徴点の検出,” 第6回計測自動制御学会システムインテグレーション部門講演会 (SI2005) 講演論文集, pp.1103 - 1104, 2005.