

顔画像の階層的な対応付けを用いた複数歩行者の認識

西山 正志^{†a)} 湯浅真由美[†] 柴田 智行[†] 若杉 智和[†]
山口 修[†]

Recognition of Moving People using Hierarchical Matching of Face Images

Masashi NISHIYAMA^{†a)}, Mayumi YUASA[†], Tomoyuki SHIBATA[†], Tomokazu WAKASUGI[†], and Osamu YAMAGUCHI[†]

あらまし 本論文では、歩行する複数人物の顔を識別する手法について述べる。歩行する人物の顔の見え方は、照明や顔の向きによる変動で大きく変わり、高い識別性能を得ることが難しい。また、複数人物が同時に歩行するため、一つのカメラだけでは隠れの問題が発生する。そこで、複数カメラから獲得された同一人物の顔画像を一つの画像列へ統合し、歩行者毎の画像列を用いて個人識別を行う。複数カメラにより隠れに対処でき、画像列を用いた識別方法で変動に頑健になる。歩行者毎の画像列を生成するために、変動の影響が少ないと考えられる各カメラにおいて、パターンマッチングを用いて同一人物の顔画像を対応付け、断片画像列を生成する。次に、断片画像列同士の類似性を用いてカメラ間の対応付けを行い、統合された画像列を生成する。最後に、統合された画像列を用いて個人を識別する。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、複数人が同時に歩行する状況を模擬した識別実験で確認した。

キーワード 歩行者認識, 顔画像, 階層的な対応付け, 相互部分空間法

1. まえがき

顔による個人識別は、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い[1]。本研究が目指すシステムは、歩行中の人物を識別対象とし、歩行者の顔画像から個人識別を行う。実用的なシステムでは、同時に歩く複数の人物を識別すること、歩行により相対的に変化する顔の向きや照明条件に影響を受けないことが求められる。

本論文では、図1のように、複数カメラを用いて、通路を歩く複数の人物から正面向きに近い顔画像を獲得する。歩行者の動線は、カメラが設置されたゲートへ向かって歩いて来る状況を仮定する。カメラを複数用いることで、ある1台のカメラでは隠れや顔向きにより正面向きに近い顔を検出できなくとも、他のカメラで捉えることを狙う。ここで、各カメラで獲得された1枚の顔画像毎に個人識別を行う場合を考える。個

人識別は、図2(i)のように、特徴空間において、入力ベクトルと参照データがなす分布との比較になる。参照データは人物毎に予め登録された画像列であり、以下では参照画像列と呼ぶ。入力データが1枚の顔画像のみでは、照明や顔の向きの影響で参照画像列と類似せず誤識別が発生する場合がある。これらの変動を抑えるために、顔画像に含まれる顔の向きや照明の影響を補正する手法[2]~[5]が提案されているが、1枚の顔画像のみでは変動を完全に除去することは難しい。そこで本論文では、画像列を用いた個人識別の手法[6]~[11]の枠組みを適用する。これらの手法では、刻々と変化する顔の見え方を画像列で表し、その画像列を入力データとし参照画像列と比較する。この識別方法は、図2(ii)のように特徴空間において、入力された画像列がなす分布と参照画像列の分布との比較になる。入力も分布で表されるため、変動の影響が緩和され、本人が正しく識別される可能性が高まる。画像列を入力として用いる識別手法を適用するためには、同時に歩く複数人物の顔画像から、同一人物の顔画像を対応付け、人物毎に統合された画像列を生成する必要がある。

そこで、本論文では、図3のように、複数のカメラ

[†] (株) 東芝 研究開発センター 〒 212-8582 川崎市幸区小向東芝町1
Corporate Research & Development, Toshiba Corporation,
1, Komukai-toshiba-cho, Saiwai-ku, Kawasaki, 212-8582,
JAPAN

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

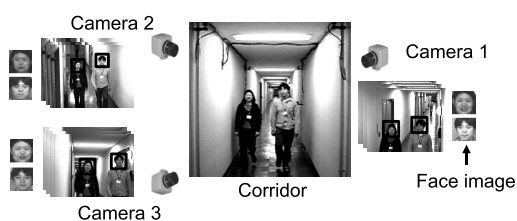


図 1 複数カメラによる複数の歩行者の個人識別
Fig.1 Face identification for moving people using multiple cameras.

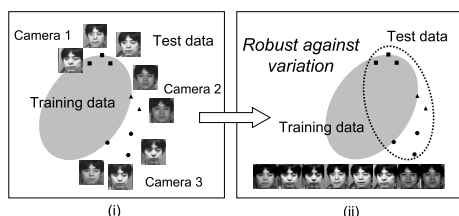


図 2 (i) 顔画像 1 枚毎による個人識別と (ii) 複数カメラから統合された画像列を用いた個人識別の比較
Fig. 2 Comparison two identification methods; (i) using a single face image from a single camera, (ii) using a set of face images integrated from multiple cameras.

から得られる複数人の顔画像を階層的に対応付け、統合された画像列を用いて個人を識別する手法について述べる。この対応付けを精度よく行うために、顔画像に対して変動を補正する手法についても述べる。提案手法では、最初に階層 1 で、各カメラ内での顔画像の対応付けを行う。次に階層 2 で、カメラ間で顔画像を対応付け、統合画像列を生成する。最後に階層 3 で、統合画像列を用いて個人識別を行う。このように対応付けの階層を設定した理由は、対応付けにおける変動要因の影響に注目し、照明、顔の向き、人数の影響が少ない順から処理を行うためである。図 4 に変動要因の影響をまとめる。階層 1 では、歩行で照明条件が相対的に変化するが、単一の視点なので顔の向きの影響が少ない上で対応付けを行うことができる。階層 1,2 では、同時刻に歩行する人数は限られるため、対応付けすべき対象の数を絞ることができる。階層 3 では、画像列を用いた識別手法の効果が発揮され、安定した個人識別を行うことができる。

以下、2. で従来手法との比較を行い、3. で階層的な対応付けについて述べる。次に、4. で変動が補正された顔画像の生成手法について述べる。最後に、5. で 349 人について撮影された歩行者データベースを用い

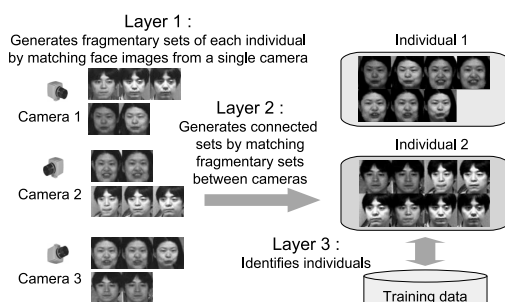


図 3 階層的な対応付けによる個人毎の画像列生成
Fig. 3 Generation of a set of face images for each individual by hierarchical matching.

Layer	1	2	3
Variation	Matching within a camera	Matching between cameras	Face identification
•Illumination	Medium	Medium	Large
•Pose	Small	Medium	Large
•Num. of individuals	Small	Small	Large

Easy → Difficult

図 4 顔画像の対応付けにおける変動要因の影響
Fig. 4 Influence of variation in matching of face images

た識別実験で提案手法の有効性を確認する。

2. 従来手法との比較

本章では、顔画像を用いた歩行者認識の従来手法と、提案手法の利点について述べる。[12] ~ [16] では、歩行による顔の向き変動を取り扱っている。カメラと正対している顔画像を選択する手法 [12], [13] , 特別な装置を用いる手法 [14], [15] , 様々な顔の向き顔画像を標準顔モデルを用いて仮想的に生成する手法 [16] が提案されている。これらの手法は顔向き変動を考慮しているものの、識別対象を 1 人のみとしており、複数の人物が同時に歩行する場合、どのように識別するかが検討されていない。

一方、複数の歩行者を対象とした個人識別の手法 [17] が提案されている。この手法では、同時に複数人が歩行する状況で、顔の様々な見え方を登録するために、複数のカメラから獲得された顔画像を、3 次元的な追跡処理を用いて人物毎に対応付ける。3 次元的な追跡処理の手法は [18], [19] でも提案されている。これらの手法は、複数カメラの撮像範囲内であれば自由に歩行することができる利点があるものの、3 次元的な追跡を精度よく行うためにカメラキャリブレーションが要求される。しかし、設置の時間や費用などを考えると、

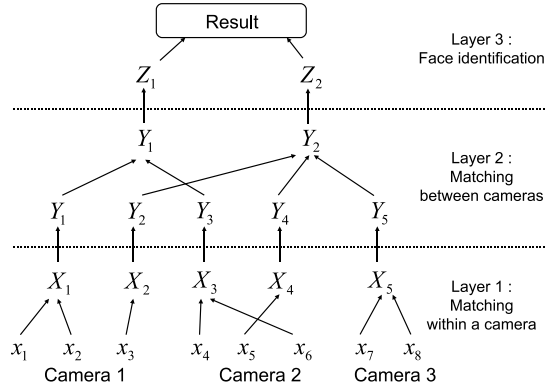


図 5 階層的な対応付けの構造の例

Fig. 5 Example of a structure of hierarchical matching.

カメラキャリブレーションを行わない方が有利である。

提案手法は [12] ~ [16] では取り扱っていなかった複数の歩行者に対して、画像列を用いた識別手法を適用することで、顔の向きや照明による変動に頑健な個人識別を行うことができる。また、フレーム間における幾何学的な位置の追跡 [17] ~ [19] を全く用いずに、画像間の類似性を計るパターンマッチングに基づく手法のみを用いて、顔画像の対応付けを行う。これにより、フレーム間における位置の連続性を維持するための追跡処理の必要がなくなり、カメラのフレームレートが低い場合や、フレームレートが異なるカメラを同時に用いる場合でも対応付けを行うことができる。また、隠れが生じた場合でも、検出と追跡の状態遷移を考慮することなく処理できる。さらに、厳密なカメラキャリブレーションを行うことなくシステムを構築することができる。

3. 顔画像の階層的な対応付け

本章では、各カメラから獲得された顔画像を階層的に対応付け、人物毎の画像列を生成し個人を識別する方法について述べる。

3.1 階層的な対応付けの枠組み

階層 1 で、各カメラにおいて同一人物の顔画像に対応付け、少数枚の顔画像から構成される断片画像列を生成する。顔画像はカメラ毎に時間の経過と共に獲得される。1 枚の顔画像 x を式 (1) で定義する。

$$x \equiv \{v, c, t\} \quad (1)$$

ここで、 v は顔画像を表す特徴ベクトル、 c はカメラ

番号、 t は獲得された時刻を表す。次に、階層 1 における断片画像列 X_i を式 (2) で定義する。

$$X_i \equiv \{x_1, \dots, x_m\} \quad (2)$$

ここで、 i は断片画像列へ付けられたラベル、 m は断片画像列を構成する顔画像の枚数を表す。階層 1 の X_i を構成する m 枚の顔画像は、同一カメラから獲得されたものとする。カメラ c で新たに顔画像 x_{new} が入力されると、これまでに獲得されている X_i との類似度 S_{1i} を求め、対応付ける断片画像列のラベル l を次式で決める。

$$l = \begin{cases} \underset{i}{\operatorname{argmax}} \{S_{1i} | i = 1, \dots, n_c\} & S_{1\max} \geq Th_{1s} \\ n_c + 1 & \text{otherwise} \end{cases}$$

ここで、 n_c はカメラ c において現時刻までに獲得された断片画像列の数、 Th_{1s} は閾値、 $S_{1\max}$ は $\max\{S_{1i} | i = 1, \dots, n_c\}$ を表す。 $S_{1\max}$ が Th_{1s} を下回った場合には、新たな人物が表れたと判断しラベル $n_c + 1$ を割り当てる。階層 1 の類似度は次式から算出される。

$$S_{1i} = \operatorname{Sim}1(x_{\text{new}}, X_i) \quad (3)$$

関数 $\operatorname{Sim}1$ については 3.2 節で述べる。ラベル l が決まると、 x_{new} は断片画像列へ式 (4) で追加される。

$$X_l \leftarrow \{x_{\text{new}}\} \cup X_l \quad (4)$$

一定時間が経過した X_i は、歩行者が撮像範囲から外れたと判定し、式 (5) により、階層 2 へ入力する断片画像列 Y_{new} へ更新する。

$$Y_{\text{new}} \leftarrow X_i \quad \text{if } t - t_i > Th_{1t} \quad (5)$$

ここで、 t は現在時刻、 t_i は X_i へ最後に追加された x が獲得された時刻、 Th_{1t} は閾値を表す。

階層 2 では、カメラ間で断片画像列に対応付け、人物毎の統合画像列を生成する。現時刻までに階層 2 で蓄積されている断片画像列を Y_i で表す。 Y_{new} が入力されると、 Y_i との類似度 S_{2i} を求め、対応付ける断片画像列のラベル k を次式で決める。

$$k = \begin{cases} \underset{i}{\operatorname{argmax}} \{S_{2i} | i = 1, \dots, n\} & S_{2\max} \geq Th_{2s} \\ n + 1 & \text{otherwise} \end{cases}$$

ここで、 n は現時刻までに階層 2 で蓄積された断片画像列の個数、 Th_{2s} は閾値、 $S_{2\max}$ は $\max\{S_{2i} | i =$

$1, \dots, n\}$ を表す．階層 2 の類似度は次式から決定される．

$$S_{2i} = Sim2(Y_{new}, Y_i) \quad (6)$$

関数 $Sim2$ については 3.3 節で述べる．ラベル k が決まると， Y_{new} は断片画像列へ式 (7) で追加される．

$$Y_k \Leftarrow \{Y_{new}\} \cup Y_k \quad (7)$$

一定の時間が経過した Y_i は，式 (8) により，階層 3 へ入力する統合画像列 Z_{new} へ更新する．

$$Z_{new} \Leftarrow Y_i \quad \text{if } t - t_i > Th_{2t} \quad (8)$$

ここで， t は現在時刻， t_i は Y_i へ最後に追加された x が獲得された時刻， Th_{2t} は閾値を表す．

最後に階層 3 において， Z_{new} と人物 i の参照画像列 Z_i を対応付けることで，次式により個人識別を行う．

$$ID = \begin{cases} \underset{i}{\operatorname{argmax}} \{S_{3i} | i = 1, \dots, h\} & S_{3\max} \geq Th_{3s} \\ impostor & \text{otherwise} \end{cases}$$

ここで， h は参照画像列が登録されている人数， Th_{3s} は閾値， $S_{3\max}$ は $\max\{S_{3i} | i = 1, \dots, h\}$ を表す．階層 3 の類似度 S_{3i} は，階層 2 と同じ関数を用いて $Sim2(Z_{new}, Z_i)$ で算出される．

図 5 に，3 台のカメラの下で，2 人の人物が歩行したときの階層的な対応付けの構造の例を示す．図中の x_i は各カメラで時間順に獲得された顔画像を表す．まず，階層 1 で x_i はカメラ毎に X_i へ対応付けられる．時間が経過すると X_i は Y_i へ更新され，階層 1 から 2 へ移る．階層 2 で Y_i はカメラ間で対応付けられ，同様に Y_i は Z_i となる．最終的に Z_i を用いて個人が識別される．

3.2 各カメラにおける断片画像列を生成するための類似度

階層 1 において， x_{new} と X_i との間の類似度を算出する関数 $Sim1$ は，式 (9) で定義される．

$$Sim1(x_{new}, X_i) = \frac{(\mathbf{v}_{new}^T \mathbf{v}_j)^2}{1 + \alpha(t_{new} - t_j)} \quad (9)$$

ここで， \mathbf{v}_{new} は x_{new} を表す式 (1) の特徴ベクトル， \mathbf{v}_j は最新の時刻 t_j をもつ $x_j \in X_i$ を表す特徴ベクトル， t_{new} は x_{new} が獲得された時刻， α は定数である． x_{new} と X_i が獲得された時間が離れている場合には類似度が小さくなるように，式 (9) の分母を設定した．これは，ある歩行者が撮像範囲から離れたが，その人

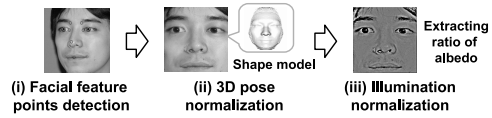


図 6 変動が補正された顔画像の生成
Fig. 6 Generation of a normalized face image.

物の断片画像列が階層 1 に残っている場合に，新たに出現した別の人物との誤対応を防ぐ効果をもつ．

他人の顔画像が誤って対応付けられると，正しく個人を識別できない問題が発生する．この問題を抑えるためには閾値 Th_{1s} の設定が重要であるが，本論文ではラベルが既知のデータセットを事前に用意し閾値を実験的に決める．

3.3 カメラ間における断片画像列を対応付けるための類似度

階層 2 において， Y_{new} と Y_i との間の類似度を算出するために，画像列同士を比較できる直交相互部分空間法 (OMSM: Orthogonal Mutual Subspace Method) [7] を用いる．OMSM では直交化行列を用いて，人物間における顔の見え方変動の差を強調する．これにより，異なる人物から生成された断片画像列間での類似度が低くなり，他人と誤って対応付けられる割合を抑えることができる．

OMSM を適用するために，それぞれの断片画像列に対して主成分分析を適用し，部分空間を生成する．断片画像列 Y_{new} から生成された部分空間に対して直交化行列を適用した部分空間を \mathcal{Y}_{new} ，同様に Y_i に対する部分空間を \mathcal{Y}_i とする．関数 $Sim2$ の出力は，二つの部分空間 \mathcal{Y}_{new} と \mathcal{Y}_i がなす正準角 θ により式 (10) で定義される．

$$Sim2(Y_{new}, Y_i) = \cos^2 \theta \quad (10)$$

なお，階層 1 において，1 人の人物から同じカメラの断片画像列が複数生成される場合がある．これらに対応付けるために， Y_{new} が階層 2 に入力されると，まず，同一カメラ内の断片画像列について OMSM を用いた対応付けを試み，その後，カメラ間の対応付けを行う．

4. 顔画像の生成方法

4.1 顔向きと照明による変動の補正

これまで述べたように階層的な対応付けには，パターンマッチングの手法を用いる．対応付けに用いる特徴ベクトル \mathbf{v} は，顔の向きや照明による変動の影響

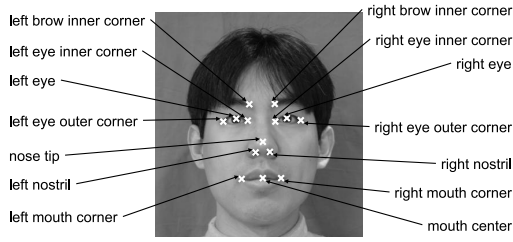


図 7 顔向き正規化に用いる顔特徴点

Fig. 7 Facial feature points for facial pose normalization.

が少ない方が、精度よく対応付けることができる。本論文では、変動を緩和するために、(i) 顔領域に対して顔特徴点の検出を行い、(ii) 3次元形状モデルを用いて顔向き正規化 [4] 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 [5] を適用する。この補正処理を図 6 に表す。v は、図 6 (iii) の顔画像をラスタースキャンすることで求める。次節より、顔特徴点の検出手法について述べる。

4.2 顔特徴点の検出

カメラ映像から顔画像 x を生成するために、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の正規化を行うには、文献 [4] で議論したように、顔の広範囲に存在する特徴点が多数必要である。このため本論文では、図 7 に示す、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。これらの顔特徴点は、文献 [4] の実験結果より、顔特徴点の数と識別性能との関係から決定した。顔特徴点を検出する手法は、例えば Elastic Bunch Graph Matching [20] など多数提案されているが、本論文では従来より我々が提案してきた手法 [22], [23] を拡張し、14 個の顔特徴点を個々に検出するアプローチをとる。顔向きの正規化 [4] には 14 個の顔特徴点全てが必要であり、顔の一部が隠蔽され顔特徴点が 1 個でも欠けた場合には顔画像の生成は行わない。

以下、顔特徴点検出の流れを述べる。最初に画像全体から Joint Haar-like 特徴と AdaBoost を用いた方法 [21] 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照合による方法 [22] 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 [23] で検出する。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。以下 4.3 節では眉内端、

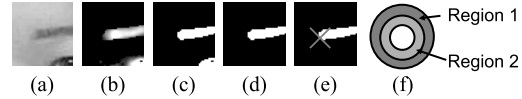


図 8 眉内端検出の例。(a) 探索範囲、(b) 分離度マップ、(c) 2 値化、(d) 眉領域判定、(e) 端点検出、(f) 分離度フィルタ

Fig. 8 Examples of detecting an inner corner points of an eyebrow. (a) Searching area, (b) Separability map, (c) Binarization, (d) Detection of eyebrow region, (e) Detection of corner points, (f) Separability filter.

4.4 節では鼻頂点を検出する手法を述べる。

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輪郭が明確でなく、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 [22] 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。

まず、検出された瞳位置を基準として、図 8(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。従来の分離度 η [22] は式 (11) で定義される。

$$\eta = \frac{\sigma_b^2}{\sigma_T^2} \quad (11)$$

$$\sigma_b^2 = u_1(\bar{I}_1 - \bar{I}_m)^2 + u_2(\bar{I}_2 - \bar{I}_m)^2$$

$$\sigma_T^2 = \sum_{i=1}^u (I_i - \bar{I}_m)^2$$

ここで、 u_1, u_2 は領域領域 1, 2 内の画素数、 I_i は画素値、 \bar{I}_1, \bar{I}_2 は領域領域 1, 2 における画素値の平均、 u は領域領域 1 と 2 合わせた領域内の画素数、 \bar{I}_m は領域領域 1 と 2 合わせた領域の画素値の平均とする。本論文では、修正分離度 η' を式 (12) で定義する。

$$\eta' = \eta + \beta(\bar{I}_1 - \bar{I}_2) \quad (12)$$

ここで、 β は定数とする。修正分離度フィルタは、図 8(f) の円形分離度フィルタ [22] 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。これにより、領域 1 に比べて画素値が小さい領域 2 だけを検出することができる。

次に、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

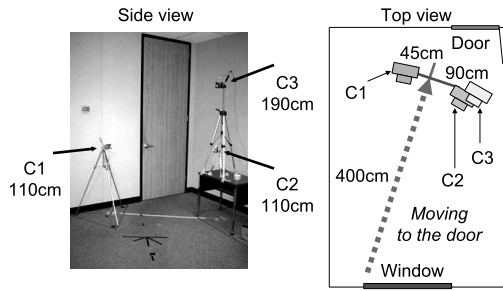


図 9 評価データベースの撮影環境

Fig. 9 Setting for collecting a database for the evaluation.



図 10 評価データベースのサンプル画像

Fig. 10 Sample images of a database for the evaluation.

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

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の円形分離度フィルタ [22] の出力がピークをとる位置と、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 歩行者データベースによる性能評価

提案手法の有効性を確認するために、歩行する人物を撮影したデータベースを用いて評価を行った。屋内

の通路に設置された複数のカメラを用いる場合、照明条件、歩行位置、人物など様々な変動要因が考えられる。本論文では、複数の人物が同時に歩く場合を想定し、その人数を変えた場合について評価を行った。まず 5.1.2 節で、同時に歩く人物は 1 人のみとし、複数カメラから獲得された統合画像列を用いて個人識別を行うことの実験的有効性を示す。次に 5.1.3 節で、複数の人物が同時に歩行する状況を模擬した実験で、階層的な対応付けの有効性を示す。

5.1.1 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シーケンスを 349 人について収集した。3 つのカメラ (C1, C2, C3) を図 9 のように配置した。窓付近を歩行開始位置とし、扉に向かって歩くシーケンスを各人物につき 2 回撮影した。識別実験を行うために、一方のシーケンスを参照データとし、もう一方のシーケンスを入力データとした。各人物は、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た。各カメラのシーケンスは、 768×1024 pixels の解像度、フレームレートは 15fps で撮影した。1 つのシーケンスの撮影時間は約 4 秒間であった。図 10 に撮影されたシーケンスの一部を示す。1 回の歩行から検出された顔画像の平均枚数は、1 台のカメラにつき 17.9 枚であった。補正処理を適用した後の顔画像の大きさは 64×64 pixels とした。

識別性能の評価には以下の 2 つの基準を用いた。以下では、本人の参照画像列と比較して算出された類似度を本人類似度、他人の参照画像列と比較して算出された類似度を他人類似度と呼ぶ。

(1) 一位正解率 (CMR : Correct Match Rate)

他人類似度と比べて本人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)

他人受誤り率 (FAR : False Acceptance Rate) と本人排除誤り率 (FRR : False Rejection Rate) が等しい時の割合を表す。FAR は以下の式で求まる。

$$FAR = \frac{\text{他人類似度が } Th_{3s} \text{ 以上の試行数}}{\text{全試行数} - \text{本人の試行数}} \quad (13)$$

一方、FRR は以下の式で求まる。

$$FRR = \frac{\text{本人類似度が } Th_{3s} \text{ より小さい試行数}}{\text{本人の試行数}} \quad (14)$$

表 1 顔画像の対応付けが正しい理想的な場合の識別性能。
(A) 対応付けなし, (B) カメラ内の対応付けのみ,
(C) 階層的な対応付け

Table 1 Recognition performance in the case that face images are matched correctly. (A) Without matchings, (B) Matching face images within a camera, (C) Hierarchical matching.

Condition	CMR(%)	EER(%)
(A)	80.3	8.7
(B)	93.7	5.6
(C)	98.6	2.0

5.1.2 複数カメラによる個人識別の有効性の検証
複数カメラから獲得された顔画像を統合し、個人識別を行うことの有効性を確認する。ここでは、顔画像の対応付けが完全に成功したとして評価する。このため、同時に歩行する人数は 1 人とした。対応付けを自動的に行った場合の個人識別の評価については後で述べる。識別性能を比較するために、次の 3 つの実験条件を設定した。

- (A) 1 枚の顔画像毎による個人識別
- (B) カメラ毎の断片画像列による個人識別
- (C) 統合画像列による個人識別

個人識別を行う回数は、(A) が $349 \times 3 \times 17.9 = 18742$ 回、(B) が $349 \times 3 = 1047$ 回、(C) が 349 回とした。用いる顔画像の枚数は、(A),(B),(C) とともに同じである。階層 1 では、大きさ 64×64 の顔画像をダウンサンプリングし、 $32 \times 32 = 1024$ 次元のベクトルを用いて対応付けを行った。階層 2 ではさらにダウンサンプリングし、 $16 \times 16 = 256$ 次元のベクトルを用いた。これらのベクトルの次元数は、各階層における対応付けの性能と計算量から決定した。ベクトルの次元数が小さい方が、類似度を算出する際の計算量が少なくなる。階層 1 の対応付けの性能は、 $64 \times 64 = 4096$ 次元と 1024 次元とでほぼ同じであったが、256 次元では低下した。階層 2 の対応付けの性能は、1024 次元と 256 次元とでほぼ同じであった。(B), (C) で個人識別を行う際、部分空間の基底ベクトルの本数は 4 として、OMSM を適用した。直交化行列は、各人物の参照部分空間から生成した。(A) で個人識別を行う際は、1 枚の顔画像を入力ベクトルとし部分空間法を用いた。ただし、参照部分空間、入力ベクトルへは直交化行列を適用した。個人識別のための参照部分空間は、手入力に対応付けられた統合画像列から作成した。(B), (C) において、生成された画像列の枚数が 4 枚以

下の時は、個人識別は行わず棄却した。つまり、通過したが入力部分空間が生成できず個人識別できなかった人物とし、本人類似度を 0 として FRR, CMR を計算した。以下では、これを画像列の生成失敗と呼ぶ。

それぞれの実験条件に対する識別性能を表 1 に示す。階層的な対応付けを行った場合の (C) が、行わなかった場合の (A),(B) と比べて識別性能が高いことが確認された。(B) から (C) への性能向上の理由として、以下の二つが挙げられる。

- 画像列の生成失敗の減少

評価データベースには帽子を深く被った人物が複数含まれている。高い位置に設置されたカメラ C3 では、帽子により顔が遮蔽され、顔領域や顔特徴点が検出されず、画像列の生成に失敗した。一方、低い位置に存在するカメラ C1, C2 では、顔領域や顔特徴点が検出され画像列の生成に成功した。(B) は各カメラで個別に画像列の生成を行うため、3 回の個人識別のうち 1 回が誤りとなった。一方、(C) は複数カメラが連携するため、画像列の生成に成功し個人識別が正解することが多かった。

- 顔向きと照明による変動への耐性向上

顔画像を生成する際に 4.1 節で述べた変動の補正が失敗した場合、変動の影響が大きく残り識別性能の低下がみられた。補正失敗の原因には、検出された顔特徴点の位置ずれにより顔向き正規化ができなかったこと、照明条件の推定を誤り照明正規化ができなかったことが考えられる。補正失敗の顔画像が画像列に多く含まれる場合、本人類似度が小さくなり誤識別となっていた。逆に、補正失敗の顔画像が画像列に含まれていても、その枚数が少なれば識別に成功することがあった。これは、画像列から部分空間を生成する際に用いる主成分分析の効果に関係すると考えられる。補正失敗の発生回数は身長、歩き方などの個人差の影響で、人物毎にカメラ間で偏りがでていた。このため、(B) では個人識別が失敗していた人物に対しても、(C) では複数カメラを用いるため、画像列において補正失敗の顔画像の割合が少なくなり個人識別が成功する場面があった。

以上から、各カメラから獲得された顔画像を画像列へ統合することは有効であるといえる。

5.1.3 階層的な対応付けを用いた個人識別の有効性の検証

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、識別実験を行っ

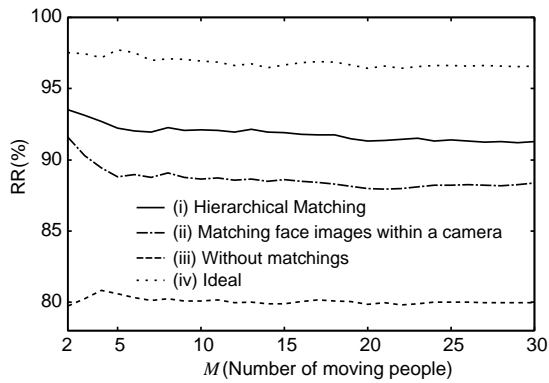


図 11 同時に歩く人数が変化した時の CMR(%)

Fig. 11 CMR(%) with increasing the number of individuals moving simultaneously.

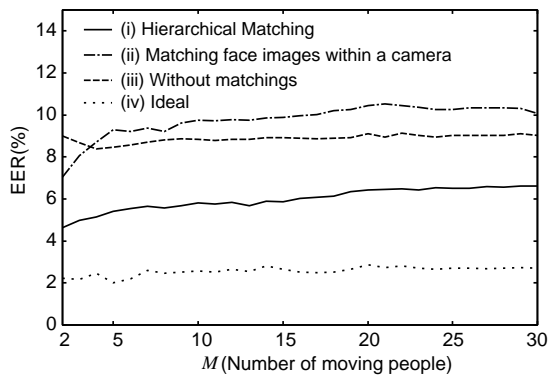


図 12 同時に歩く人数が変化した時の EER(%)

Fig. 12 EER(%) with increasing the number of individuals moving simultaneously.

表 2 本人類似度と他人類似度の平均

Table 2 Average of self and other similarities.

Condition	Similarity	
	Self	Other
(i)	0.60	0.20
(ii)	0.56	0.20
(iii)	0.37	0.13
(iv)	0.70	0.22

た．ここでは，複数の人物が横に並んで同時に歩行する状況を模擬した．まず，349 人の中からランダムに M 人を選択した．選択された人物から 1 枚の顔画像を初期時刻で階層 1 へ入力し，66 ミリ秒後に，次の顔画像を入力した． M 人の選択を 200 回繰り返し CMR, EER を算出した．階層 1, 2 の変数は $Th_{1s} = 0.4, Th_{2s} = 0.4, Th_{1t} = 1\text{sec.}, Th_{2t} = 3\text{sec.}$ とした．これらの変数は， $M = 10$ の時に，他人の顔

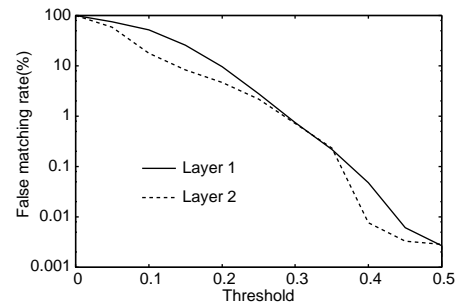


図 13 閾値 Th_s が変化した時の誤対応率 (%)

Fig. 13 False matching rate with changing threshold.

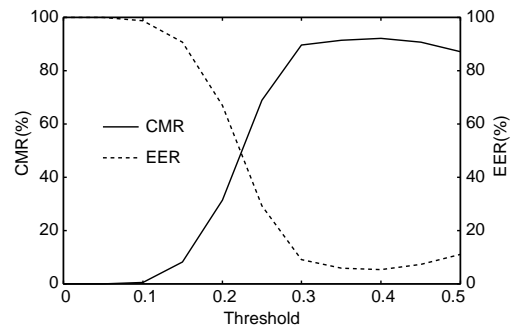


図 14 閾値 Th_s が変化した時の CMR(%) と EER(%)

Fig. 14 CMR(%) and EER(%) with changing threshold.

画像が誤って統合される確率が 0.1% 以下となるように実験により決定した．誤って他人の顔画像が統合画像列へ混入した場合，個人識別において正解が定まらないため，他人類似度として FAR のみを計算した．

識別性能を比較するために，次の 4 つの実験条件を設定した．

- (i) 階層的な対応付けを用いた個人識別
- (ii) 階層 1 のみを用いた個人識別
- (iii) 対応付けを用いない個人識別
- (iv) 対応付けが完全に成功した場合の個人識別

(i) と (ii) では対応付けを自動で行い，(iv) では対応付けを手作業で行う．また，(i) と (iv) では統合画像列，(ii) では断片画像列，(iii) では 1 枚の顔画像を個人識別に用いる．(iii) は表 1(A) と，(iv) は表 1(C) の理想的な性能と対応する．

M が変化したときの CMR の推移を図 11，EER の推移を図 12 に示す．例えば $M = 5$ 人が同時に歩いているシーンを模擬した場合，提案手法 (i) は (iii) と比べて，CMR が 80.3% から 92.2%，EER が 8.7% から 5.4% に改善されている．(i) の CMR と EER は，

理想的な場合 (iv) には劣るものの、(ii) や (iii) に比べて大きく改善されている。(ii) の EER が (iii) に比べて高くなる原因として、個人識別に用いる断片画像列の生成に失敗し、入力部分空間が生成できなかったケースが多いためと考えられる。この現象は (i) では起きていないことから、階層 2 の対応付けには効果があるといえる。

次に、対応付けにより生成された画像列を構成する顔画像の平均枚数について述べる。(i) では 15.9 枚、(ii) では 8.7 枚、(iv) では 53.6 枚であった。(i) は (iv) に比べて少ないものの、(ii) に比べて枚数は増加している。枚数の増加は、識別性能の向上に寄与していると考えられる。また、さらなる識別性能の向上のためには、対応付けの性能向上が必要であることが分かる。

次に、本人類似度と他人類似度の平均値を表 2 に示す。実験条件は図 11 と同じである。(i),(ii) では (iii) に比べて、他人類似度が少し増えているものの、本人類似度が理想的な条件 (iv) に近づいている。また、(i) は (ii) と比べても本人類似度が向上していることが分かる。

最後に、各階層における対応付け誤りと識別性能の関係について述べる。対応付け誤りの発生は、閾値 Th_{1s} , Th_{2s} に依存する。図 13 に、閾値と誤対応率の関係を示す。ここでは、 $Th_{1s} = Th_{2s} = Th_s$, $M = 5$ とした。誤対応率は、対応付け誤りの発生数を対応付けを行った総数で割ったものである。閾値が高くなるにつれて誤対応率が下ることが確認できる。図 14 に、閾値と識別性能の関係を示す。閾値を高く設定することで識別性能は向上するが、0.45 付近より性能低下が見られた。これは、本人同士の顔画像が対応付けられず画像列の生成失敗が発生したことが原因と考えられる。この結果より、提案手法の識別性能は閾値に依存するため、適切な閾値の設定が重要である。以上、実験結果により提案手法の階層的な対応付けによる個人識別の有効性を確認できた。

5.2 隠れが生じる場合の統合画像列生成

この節では、追跡処理で問題とされる隠れに対して、階層的な対応付けの有効性について述べる。提案手法を用いたシステムを、通路に設置し歩行者を撮影した。図 15 に、このシステムを用いて撮影した画像の例を示す。この例では、廊下を同時に歩く 3 人の歩行者が、お互いに相手の前面へ回りこむことで、隠れを意図的に発生させた。横幅が約 160cm の廊下の両

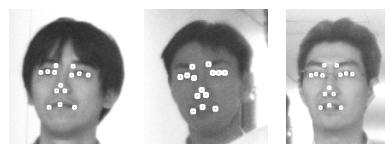


図 16 検出された顔特徴点の例
Fig. 16 Examples of detected facial feature points.

端へカメラをそれぞれ 1 台配置し、床からカメラの高さは 110cm とした。カメラ 1 とカメラ 2 ではレンズの焦点距離が異なる。顔画像を生成するための計算機を 1 台、階層的な対応付けと個人識別を行うために同じ構成の計算機を 1 台用いた。各カメラにおいて 1024×768 pixels の解像度の画像を、フレームレートを 7.5fps で撮影した。68 枚の顔画像を自動で検出した。検出された顔特徴点の例を図 16 に示す。次に、生成された統合画像列を図 17 に示す。この例では、統合画像列を構成する枚数は、人物 1 で 20 枚、人物 2 で 19 枚、人物 3 で 24 枚であった。一度隠れて再び出現した人物でも対応付けが成功している。この実験では、一方のカメラで顔の一部が隠蔽された場合、他方のカメラで隠れが生じていない顔画像が獲得され、その顔画像が画像列へ統合されていた。以上の結果より、フレーム間の追跡処理では失敗するような隠れの問題にも、提案手により本人同士で対応付けが成功し、人物毎の統合画像列が生成されることが認できた。

6. む す び

本論文では、複数のカメラが設置された環境において、複数の歩行者から獲得された顔画像を対応付けることで統合画像列を生成し、個人を識別する手法について述べた。提案手法では、画像列を用いた識別手法を適用するために、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問題へと階層的に分け、最終的に統合する枠組みを用いた。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実験により確認した。

今後の課題として、模擬実験ではなく複数歩行者による隠れが実際に生じているデータベースでの性能評価、提案手法とは異なる対応付けの手法を用いた場合との性能比較、階層的な対応付けで誤って他人の顔画像が統合される割合を減らすために閾値設定以外の方法を検討すること、カメラの配置方法の検討が挙げられる。

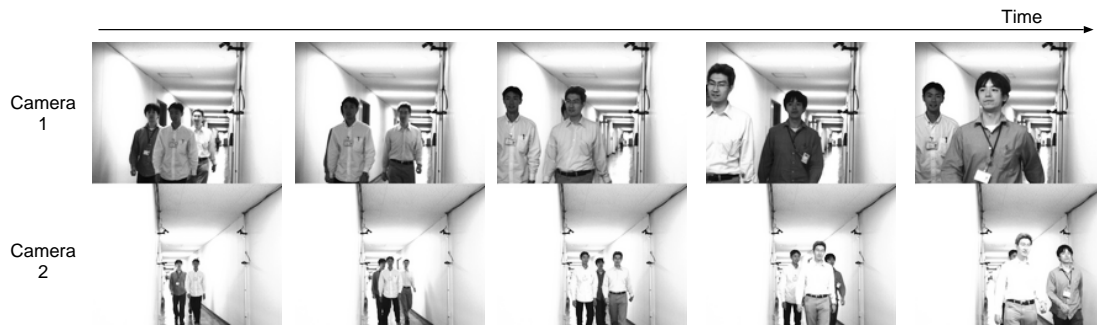


図 15 隠れを伴う歩行者を撮影したシーケンスの例 (上下の組は同時刻に獲得したフレームを表す)

Fig. 15 Example of a video sequence captured from moving people with occlusion. The pair of upper and lower images represent the same frame acquired simultaneously.



図 17 生成された統合画像列の例

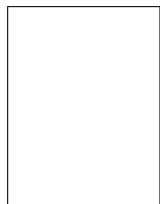
Fig. 17 Examples of connected sets.

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] V. Blanz, and T. Vetter, “Face Recognition Based on Fitting a 3D Morphable Model,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 25, no. 9, pp. 1063-1074, 2003.
- [3] H. Wang, S. Z. Li, and Y. Wang, “Generalized Quotient Image,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 498 - 505, 2004.
- [4] T. Kozakaya, and O. Yamaguchi, “Face Recognition by Projection-based 3D Normalization and Shading Subspace Orthogonalization,” 7th International Conference Automatic Face and Gesture Recognition, pp. 163 - 168, 2006.
- [5] M. Nishiyama, and O. Yamaguchi, “Face Recognition Using the Classified Appearance-based Quotient Image,” 7th International Conference Automatic Face and Gesture Recognition, pp. 49 - 54, 2006.
- [6] 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.
- [7] 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.
- [8] G. Shakhnarovich, J. W. Fisher, and T. Darrell, “Face Recognition from Long-Term Observations,” 7th European Conference on Computer Vision, pp. 851 - 868, 2002.
- [9] 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.
- [10] L. Wolf, and A. Shashua, “Learning over Sets using Kernel Principal Angles,” Journal of Machine Learning Research, 4:913-931, pp. 913-931, 2003.
- [11] O. Arandjelovic, and R. Cipolla, “Face Recognition from Image Sets using Robust Kernel Resistor-Average Distance,” The First IEEE Workshop on Face Processing in Video, 2004.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] H. Tanaka, I. Kitahara, H. Saito, H. Murase, K. Kogure, and N. Hagita, "Dynamically Visual Learning for People Identification with Sparsely Distributed Cameras" 14th Scandinavian Conference on Image Analysis, pp. 130 - 140, 2005.
- [17] 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.
- [18] A. Nakazawa, H. Kato, S. Hiura, and S. Inokuchi, "Tracking Multiple People using Distributed Vision Systems," International conference on Robotics & Automation 2002, pp. 2974 - 2981, 2002.
- [19] N. Ukita, and T. Matsuyama, "Real-time Cooperative Multi-target Tracking by Communicating Active Vision Agents," Computer Vision and Image Understanding, Vol. 97, No. 2, pp. 137 - 179, 2005.
- [20] L. Wiskott, J. Fellous, N. Kruger, and C. Malsburg, "Face Recognition by Elastic Bunch Graph Matching," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 775 - 779, 1997.
- [21] 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.
- [22] 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.
- [23] 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.

(平成 xx 年 xx 月 xx 日受付)



西山 正志 (正員)

2000 岡山大・工・情報卒 . 2002 同大学院博士前期課程了 . 同年 (株) 東芝入社 . 現在, 同社研究開発センターマルチメディアラボラトリー勤務 . 顔画像認識の研究に従事 . 2006 本会パターン認識・メディア理解研究会研究奨励賞受賞, 2007 第 12 回画像センシングシンポジウム優秀論文賞受賞 . 計測自動制御学会会員



湯浅真由美 (正員)

1990 阪大・理・物理卒 . 1992 同大学院博士前期課程了 . 同年 (株) 東芝入社 . 現在, 同社研究開発センターマルチメディアラボラトリー研究主務 . コンピュータビジョンに関する研究開発に従事 .



柴田 智行

2003 和歌山大・システム工・情報通信システム卒 . 2005 同大学院博士前期課程了 . 同年 (株) 東芝入社 . 現在, 同社研究開発センターマルチメディアラボラトリー勤務 . コンピュータビジョン, パターン認識等の研究に従事 .



若杉 智和 (正員)

1996 阪大・基礎工・物理卒 . 1998 同大学院博士前期課程了 . 同年 (株) 東芝入社 . 現在, 同社研究開発センターマルチメディアラボラトリー勤務 . 画像処理によるヒューマンインタフェースの研究開発に従事 .



山口 修 (正員)

1992 岡山大・工・情報卒 . 1994 同大学院工学研究科了 . 同年 (株) 東芝入社 . 現在, 同社研究開発センターマルチメディアラボラトリー研究主務 . コンピュータビジョン・顔画像処理の研究に従事 . 1996 情報処理学会全国大会優秀賞, 2002 本会論文賞, 2003 情報処理学会山下記念研究賞受賞 . 情報処理学会会員