

顔の大きさの変化にロバストな個人識別システム

平山高嗣[†] 岩井儀雄[†] 谷内田正彦[†]

顔画像を利用したヒューマンインタフェースを構成する場合、システムがユーザに拘束条件を与えないことが理想の形態となる。顔はその姿勢や表情により、その見え方が様々に変化するため、この変化にロバストな認識技術が必要になる。本研究では、入力画像中の顔領域と照合に用いるモデルの大きさが異なる場合でも、顔認識が可能となる照手法を提案する。これはフレキシブル特徴照手法を拡張した手法で、照合に用いる特徴量が顔の大きさの変化に対して正規化されるため、顔認識の精度を保つことができる。特徴量の正規化には、顔の大きさの確率的な推定手法と効率的な処理を行うために改良したガボールウェーブレット変換を用いる。我々は提案手法に基づいた個人識別システムを構築し、個人識別実験によりその有効性を検証した。

Person Identification System Robust for Scale Variations

TAKATSUGU HIRAYAMA,[†] YOSHIO IWAI[†] and MASAAHIKO YACHIDA[†]

Facial recognition technology needs to be robust for arbitrary facial appearances because a face changes according to facial expressions and facial poses. In this paper, we propose a method which automatically performs face recognition for variously scaled facial images. The method performs flexible feature matching using features normalized for facial scale. For normalization, the facial scale is probabilistically estimated and is used as a scale factor of an improved Gabor wavelet transformation. We implement a face recognition system based on the proposed method and demonstrate the advantages of the system through facial recognition experiments.

1. はじめに

近年、急速な普及と高性能化が進む情報機器は人間の生活に必要不可欠な存在となりつつあるが、一方で高機能化により情報機器を使いこなすことができない情報弱者が発生しつつある。そこで、人にやさしいインタフェースとして、人間のコミュニケーションに対応できる情報機器が求められている。現在、我々が利用できるコンピュータコミュニケーションとしては、文字を用いるものが主となっているが、より自然なやりとりのできるものが必要とされる。人間がコミュニケーションで伝達するメッセージは、そのわずか7%が言語によるものである一方、55%は顔の表情によるという報告¹⁾がある。つまり、人間のコミュニケーションにおいて、顔は非常に重要な役割を果たすといえ、顔画像を媒体としたヒューマンインタフェースが有効であると考えられている。こういった背景のもとで、画像に映し出された顔を認識する研究がさかんに行わ

れている。

顔はその姿勢や照明に代表される撮影条件の変化、表情や時間経過ともなう容貌の変化といったそれ自体の変形などによって、その見え方が様々に変化する。顔認識に関する研究がさかんになり始めた当初、テンプレート照合を使う手法²⁾、主成分分析(PCA)により抽出した固有顔(Eigenface)を利用する手法³⁾、線形判別分析(LDA)を利用する手法⁴⁾などが提案された。これらの手法は、ほぼ一様な照明条件の下で位置と大きさが正規化された無表情な正面顔を識別する手法として、その有効性がおおむね確認されている。しかし、従来研究における技術では、撮影のためにシステムがユーザに拘束を強いるため、自然なコミュニケーションを行うためのヒューマンインタフェースを構成することが困難である。

近年の顔認識の議論の中心は多様な見え方の顔のロバストな認識に移り、様々な手法が提案されている。それらの手法の中で現在、最も高い認識性能を示すことができると思われるのが、ガボールウェーブレット特徴量をノードに持つグラフを顔の表現に利用する手法である⁵⁾。ガボールウェーブレット特徴量は画像に対

[†] 大阪大学大学院基礎工学研究科
Graduate School of Engineering Science, Osaka University

するガボールウェーブレット変換^{6),7)}によって求められる。ガボールウェーブレットは哺乳類の1次視覚野におけるニューロンのフィルタ応答に似た特性を持つカーネルである。カーネルとして、いくつかの異なった大きさ(周波数)と方向の形状のものを用意することで、画像の特徴を十分に解析することができる。また、照明変化にロバストな特徴量を抽出できるという特性がある。ただし、ガボールウェーブレット変換は計算コストが非常に高く、カーネルの数に比例する。このため、実時間処理が求められるヒューマンインタフェースを構築するためには、顔認識手法の計算効率をできるだけ向上させる必要がある。基準となる顔モデルグラフと入力画像の照合をとる手法としては、弾性的マッチング^{8),9)}や我々の提案したフレキシブル特徴照合法^{10),11)}がある。これらの手法は、グラフを変形させて照合を行うため、基準となる顔モデルに対する入力画像の顔の変化にある程度対応することができる。しかし、この変化が大きい場合、照合の精度が著しく低下する。そこで、我々はまず顔の大きさに注目し、これらの手法を利用して顔の大きさの変化に対応できる認識手法を提案する。画像中の顔の大きさはカメラと被写人物の顔の距離に依存する。被写人物がこの距離を合わせることは一般に難しいタスクとされている¹²⁾ため、顔の大きさの変化にロバストな認識手法の開発が求められている。

顔の大きさの変化に対応するため、弾性的マッチングでは、いくつかのモデルグラフを用意し、それらの中から顔の大きさの変化に対し最適なものを選ぶという一般的な方法が取り入れられている^{9),13)}。この方法は、精度良い処理を行うために多くのモデルグラフを用意する必要があり、その数が増えるほど、計算効率が低下する。Krügerら¹³⁾はこの計算効率が弾性的マッチング自体にかかるものより非常に低くなるとして問題視している。また、弾性的マッチングの前処理として、顔を楕円で近似することで顔の大きさを推定する手法¹⁴⁾が提案されている。この手法は、反復処理によって楕円を変形させ適切な大きさに収束させるため、計算効率は一般的な方法と同程度となる。このように顔の大きさの変化に対応するためには、顔の大きさの推定精度と計算効率のトレードオフを解決しなければならない。これが1つ目の課題である。

さらに、顔の大きさの変化に合わせて画像の空間周波数も変動するため、それに対応した特徴量の抽出も必要である。これが2つ目の課題である。従来研究^{8)~11),13),15)~17)}では、あらかじめ多段階の大きさのガボールカーネルで特徴量を抽出することにより、

顔の大きさの変化に対応させることが主であった。しかし、この対応は最適な形状のガボールカーネルを利用していないため、認識精度の低下を招いている¹⁴⁾。これに対し、Lades¹⁸⁾は推定した顔の大きさを利用して、多段階の特徴量からの補間によって、顔の大きさに最適な特徴量を求める手法を提案している。しかし、補間の精度が高くないので、これに代わる手法として、画像の大きさを正規化する手法が提案されている¹⁹⁾。この手法の問題点としては、正規化にともなう画像の情報落ちが認識精度の維持を妨げることである。また、画像を正規化するための処理時間が必要になる。

我々が提案する手法は、上述の手法に比べ計算効率が良く、なおかつ高い認識精度を維持できる。まず、1つ目の課題に対しては、顔画像と1つの大きさのモデルグラフだけの照合で、確率的に顔の大きさの推定とグラフの大きさの伸縮とを行うスケール変換という高効率な手法を提案する。そして、2つ目の課題に対しては、スケール変換で推定した顔の大きさの情報を利用してガボールカーネルの形状を変更し、顔の大きさの変化に対して正規化した特徴量を抽出する手法を提案する。この手法は、画像を正規化するよりも効率良く正規化した特徴量を抽出できる。さらに、この手法により、ガボールカーネルの大きさの段階を削減しても認識精度が維持されると考えられる。この削減が顔認識手法の計算効率の面で非常に有効と思われる。また、我々はこの正規化の処理を効率良く行うために、一般的なガボールカーネルの生成式を改良する方法を提案する。

本稿で提案する手法は、弾性的マッチングより効率の良い手法であるフレキシブル特徴照合法に、スケール変換とガボール特徴量の顔の大きさに対する正規化を加えたものである。我々は本手法を顔位置検出、顔特徴点抽出、個人識別を一連の処理で行うシステムに導入し、本手法の有効性を顔認識実験によって検証する。

2. 顔の表現

2.1 特徴点と特徴量

本研究では、目、唇、鼻といった顔の30個の特徴点上に配置されたノードを結ぶグラフによって顔を表現する。図1に顔グラフの例を示す。各ノードは対応する特徴点の位置におけるガボール特徴量を保持する。この特徴量はガボールウェーブレット変換によって求められる。この表現により、本手法は顔画像を情報圧縮でき、モデルデータベースにも多数の顔モデルをグ

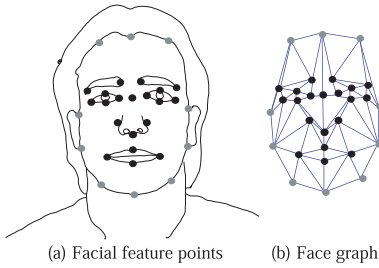


図 1 顔の表現
Fig. 1 Face representation.

ラフの形でコンパクトに蓄積できる．また，各ノードは隣接するノードとの距離も保持する．

2.2 ガボールウェーブレット変換

一般的なガボールウェーブレット変換は式 (1) で定義されるガボールカーネルと画像の畳み込みによって求められる．画像を周波数領域での表現に変換するものである．ガボールカーネルは，その特性が大きさ a と中心周波数のファクタ j と方向ファクタ θ によって制御されるため，画像の局所的な濃度変化を検出するフィルタとしてよく用いられる．画像の特徴を十分に解析するためには，異なった j と θ を持ったカーネルを用意すればよい．ガボールカーネルの特性としては，式 (1) の第 2 項中の減算により，関数の直流成分を補正しているため，照明条件に対してロバストになることもあげられる．また，この変換による応答は複素数となるが，その振幅成分は畳み込みの位置の変化に沿ってゆっくりと変化し，位相成分は激しく変化するため，一般的には振幅成分のみを用いることが多い．図 2 にカーネルの形状の例を示す．

$$\psi_{j,\theta} = \frac{1}{4\pi a^2 j \sigma} \exp\left(-\frac{x^2 + y^2}{(2a^j \sigma)^2}\right) \cdot \left[\exp\left(\frac{i u (x \cos \theta + y \sin \theta)}{a^j}\right) - \exp(-(u\sigma)^2) \right],$$

$$j = 1, \dots, N_j,$$

$$\theta = \frac{\kappa\pi}{N_\theta} \text{ with } \kappa = 0, \dots, N_\theta - 1. \quad (1)$$

本研究では，顔の大きさの変化に対応した特徴量を容易に抽出するために，上述のガボールウェーブレット変換を式 (2) のように改良する．式 (1) と式 (2) における a と j の関係に注目すると，式 (1) で定義されるガボールカーネルは j によって大きさと中心周波数が非線形に制御される．このため，顔の大きさの変化に対して自在にガボールカーネルの大きさと中心周波数を操作するには， j の設定が複雑となる．一方，我々が

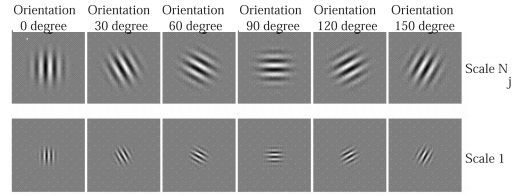


図 2 ガボールカーネルの形状
Fig. 2 Gabor kernels.

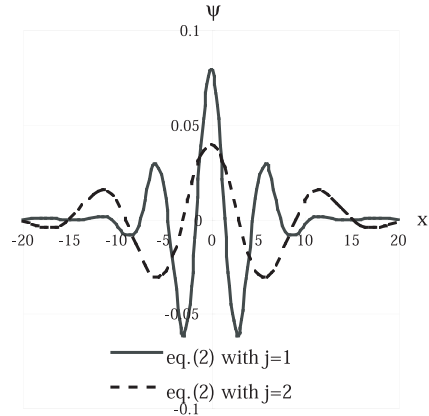


図 3 式 (2) の形状
Fig. 3 Shapes of Equation (2).

提案するものは，その関係が線形である．ゆえに，モデルグラフに対する入力画像の顔の大きさの比率 s_{opt} を j にかけて ($j \leftarrow j_{s_{opt}}$)，改良したカーネルを拡大縮小することで，顔の大きさの変化による空間周波数の変動を吸収したガボール特徴量が容易に求められる．つまり，顔の大きさの変化と正比例してカーネルの大きさを変えることが容易にできる． a と j の関係が非線形である式 (1) においても， $j \leftarrow \log(a j s_{opt}) / \log a$ とすることで， s_{opt} の変化と正比例してカーネルの大きさを変えることができる．しかし，これは式 (2) に比べ，約 20 倍の計算コスト (計算機による時間計測の結果) を必要とする． $j = 1, 2$ に対応する式 (2) を 2 次元表示したものを図 3 に示す．この図から， $j = 2$ のカーネルのピークは， $j = 1$ に対して x 軸方向に 2 倍された位置に現れていることが確認できる．つまり，改良したガボールカーネルは j に正比例してカーネルの大きさが変化することが確認できる．

$$\psi_{j,\theta} = \frac{1}{4\pi a^2 j^2 \sigma} \exp\left(-\frac{x^2 + y^2}{(2a j \sigma)^2}\right) \cdot \left[\exp\left(\frac{i u (x \cos \theta + y \sin \theta)}{a j}\right) - \exp(-(u\sigma)^2) \right]. \quad (2)$$

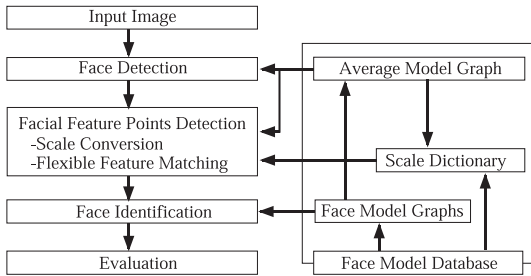


図 4 個人認識システムの概略
Fig. 4 Overview of the system.

3. 顔認識システム

3.1 概要

システムの概略を図 4 に示す．本システムは，顔位置検出，顔特徴点抽出，顔同定からなる．顔特徴点抽出は，スケール変換とフレキシブル特徴照合の 2 つのステージから構成される．また，顔モデルデータベースには，顔の大きさがほぼ等しい人物のモデル画像が登録されており，それらからモデルグラフと平均モデルグラフ，そしてスケール変換に用いられるスケール辞書がシステム構築時にあらかじめ作成される．なお，モデルグラフの各ノード位置は，鼻の頭（ハイライトとなる部位）に位置するノードを基準とした相対位置へと変換される．次節より，主要なモジュールについて詳細に述べる．

3.2 顔位置検出

顔位置の検出は，参照グラフと呼ばれる照合用のモデルグラフを用いる．参照グラフには平均顔グラフを用いる．平均顔グラフは，モデルグラフの各ノードが保持するガボール特徴量の平均と，ノード間距離の平均から作成される．顔位置の検出はまず，入力画像中を粗くラスタスキャンし，各位置で参照グラフをマッピングしてシャドウグラフを生成する．シャドウグラフは，画像中における特徴点の位置仮説に相当する．その仮説の妥当性は式 (3) によって計算され，その式の応答（ネットワーク応答）が小さいほど，シャドウグラフは参照グラフとの全体的な調和を高める．本手法は，このネットワーク応答が最小となるシャドウグラフの位置を顔の位置と見なす．我々はこの照合を，参照グラフのノードの空間位置関係を固定して順次行うことから，固定照合と呼ぶ．顔位置検出により，顔特徴点の大まかな位置も得ることができる．これらの位置は，正確な特徴点の位置を求めるために，顔特徴点抽出の初期位置として利用する．

$$O(G) = \sum_{m \in G} (\|W_{m,j} - I_{m,j \times s_T}\|). \quad (3)$$

ここで，式 (3) の G はグラフ中のノードを表す． $W_{m,j}$ は参照グラフのノード m において，大きさ j のガボールカーネルで抽出されたガボール特徴量を要素とするベクトルである． $I_{m,j \times s_T}$ はシャドウグラフ G のそのベクトルであり， $j \times s_T$ は大きさ j のカーネルをグラフの伸縮率 s_T で拡大縮小させたものを表す．顔位置検出では，グラフの伸縮が行われないので $s_T = 1$ とする．

3.3 顔特徴点抽出

3.3.1 スケール変換

スケール変換は顔特徴点抽出に用いるグラフを，入力画像に対して効率的に適切な大きさに伸縮させる役割を果たす．これには，スケール辞書というデータベースを利用する．

まず，スケール辞書の生成について説明する．顔の大きさが参照グラフ（平均顔グラフ）とほぼ等しい N 枚の教師顔画像を用意する．システムはそれぞれの画像に対して，参照グラフを伸縮率 s_T で T 段階に伸縮させたグラフを顔位置検出と同様の方法で適切な位置へマッピングし，シャドウグラフを生成する．参照グラフに対するシャドウグラフのネットワーク応答は式 (3) によって求められ，これがスケール辞書に登録される．つまり，スケール辞書には，顔とグラフの大きさの合致具合を表す教師データが記録されている．教師データの総数は，教師顔画像の数 $N \times$ 参照グラフの伸縮率 s_T の段階数 T となる．

グラフの適切な大きさへの伸縮は，この辞書を用いて以下のように統計的に行なわれる．まず，入力画像に対して，顔位置検出により求められた顔位置で参照グラフからシャドウグラフが生成される．そしてネットワーク応答が求められ（式 (3)， $s_T = 1$ ），スケール辞書の教師データとの比較が行われる．比較は教師顔画像ごとに行われ，最も誤差の小さい教師データがそれぞれ検出され，その教師データに対応した参照グラフの伸縮率 r_i が導き出される（式 (4)）．ここで，式 (4) の $O(G_{ref} \times s_T)$ は教師データで， $G_{ref} \times s_T$ は参照グラフ G_{ref} を s_T で伸縮させたグラフである．次に，すべての教師顔画像から求めた r_i のヒストグラムが生成される．このヒストグラムから求まる r_i の確率分布に従い，入力画像に対するシャドウグラフの大きさの比率 t_{opt} が推定される．シャドウグラフの適切な大きさへの伸縮は，伸縮率 $s_{opt} (= 1/t_{opt})$ を求めて行われる．我々はこの処理を精度良く行うために，高い確率を持つ r_i を c 個選び， t_{opt} の候補 t_{cand} とする．そして，それらに対する伸縮率 $(1/t_{cand})$ で伸縮させたシャドウグラフと参照グラフのネットワーク

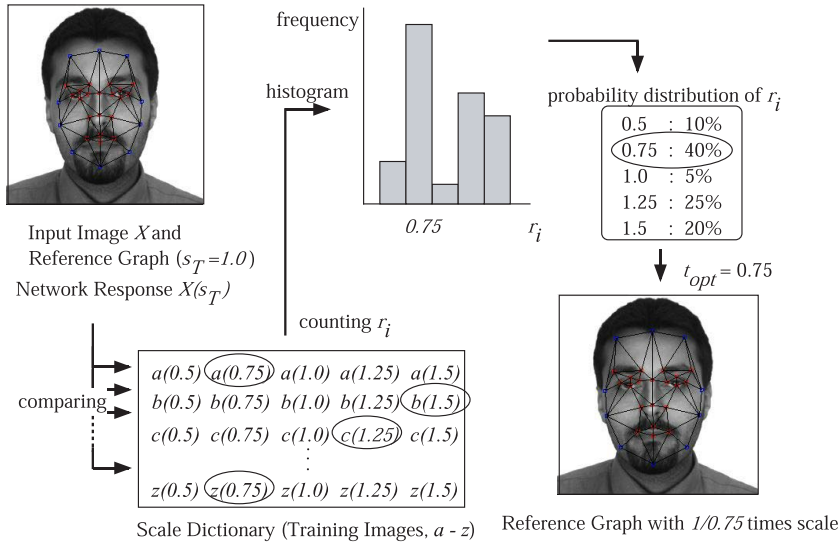


図 5 スケール変換

Fig. 5 Scale conversion process.

応答を求める (式 (3), $s_T = 1/t_{cand}$). 我々は, ネットワーク応答が最も小さくなった $1/t_{cand}$ を s_{opt} として採用する. c の設定については後述する (4.1 節). 図 5 にスケール変換の処理の流れを示す. この図は, 参照グラフの大きさを 1 として, 入力画像の大きさが $1/0.75$ である場合のスケール変換を表している.

$$r_i = \arg \min_{s_T} [O(G) - O(G_{ref} \times s_T)], \quad (4)$$

$$i = 1, \dots, N.$$

3.3.2 フレキシブル特徴照合法

フレキシブル特徴照合はまず, スケール変換後のシャドウグラフの各ノードをその近傍にランダムに移動させること (ランダム拡散過程) でシャドウグラフを変形させ, 新しいシャドウグラフの候補を生成する. シャドウグラフの 1 つのノードをその近傍にランダムに移動させている様子の略図を図 6 に示す. 新しいシャドウグラフにはこの候補のうち, 式 (5) のネットワーク応答が最も小さいものを採用する. この処理がシミュレーテッド・アニーリング法によって繰り返され, シャドウグラフと入力画像の全体的な調和が高められていく. なお, ガボール特徴量は, シャドウグラフの候補が生成されるごとにそのノード位置でのみガボールウェーブレット変換が施されて, 抽出される. また, ノードが顔の器官の空間位置関係を無視した位置に収束することを防ぐため, ノードの移動に簡単な制約を持たせる. この制約には一般的に知られている顔の構成知識^{20),21)} を利用している.

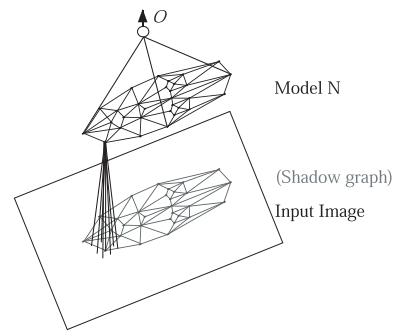


図 6 ランダム拡散によるシャドウグラフの生成

Fig. 6 The random diffusion mechanism for shadow graph generation.

$$O_{flex}(G) = \sum_{m \in G} \left[(\|W_{m,j} - I_{m,j \times s_{opt}}\|) + \lambda \sum_{n \in G-m} \left(\|v_{mn} - \frac{e_{mn}}{s_{opt}}\| \right) \right]. \quad (5)$$

ここで, 式 (5) の $j \times s_{opt}$ は大きさ j のガボールカーネルを s_{opt} で拡大縮小させたものを表す. これにより, 参照グラフとシャドウグラフが持つガボール特徴量が正規化される. v_{mn} は参照グラフの, e_{mn} はシャドウグラフ G のノード m と n の距離情報である. 距離情報 e_{mn} も s_{opt} によって正規化される. また, λ はガボール特徴量と距離情報のどちらを重視するかを決める係数である. λ は式 (6) に従って, 照合の繰返し処理が進むにつれて単調減少する. λ_{ini} は照合の初期段階のノード移動距離, λ_{fin} は最終段階のノード

ド移動距離, t_{max} は最大の繰返し数である. なお, t は現在の繰返し数である. また, 照合のノード移動距離も λ に従って単調減少する.

$$\lambda(t) = \lambda_{ini} \left(\frac{\lambda_{fin}}{\lambda_{ini}} \right)^{\frac{t}{t_{max}}} \quad (6)$$

フレキシブル特徴照合は, 参照グラフに平均顔グラフを利用することで, モデルデータベースに登録されていない人物の顔の特徴点もロバストに抽出することができる. また, 空間位置関係を考慮した特徴照合を行うので, ひげや眼鏡などを身に付けていても, 特徴点の抽出は可能である.

3.4 顔 同 定

顔特徴点抽出で得られた最終的なシャドウグラフは, 入力画像の顔の表現として利用できる. 顔の同定は, そのシャドウグラフと顔モデルデータベースとを比較することで行える.

同一人物の顔かどうかの判定は, 式 (5) で定義されるネットワーク応答から決定される. すなわち, シャドウグラフとモデルグラフのネットワーク応答が最小であるモデルが入力画像の顔の人物であると認識される. ただし, 応答があらかじめ設定されている閾値以下でなければ, システムはデータベースに登録されている人物とは判定しないとする. 閾値の設定については後述する (4.3 節). なお, 顔同定処理での式 (5) の $W_{m,j}$ はモデルグラフのノード m において, 大きさ j のガボールカーネルにより抽出されたガボール特徴量を要素とするベクトル, v_{mn} はモデルグラフのノード m と n の距離情報である.

4. 実 験

我々は, 提案手法とシステムの有効性を検証するために評価実験を行った. 評価実験システムの実装は SGI 製ワークステーション Onyx2 (CPU R12000 300 MHz, Memory 4 GB, OS IRIX6.5.10 m) 上で行われた. 画像は Purdue 大学の AR Face データベース²²⁾ のもの (256 × 256 画素, 8 bit 濃淡) を利用した. モデルデータベースには, ほぼ等しい大きさの 50 人の顔画像とそれぞれのモデルグラフが登録されている. 画像は 1 人につき 1 枚で, すべてが無表情の正面顔である. モデルグラフの特徴点位置は我々が目視によって顔画像中で定義した. 入力画像は, モデルデータベースに登録された人物のモデル画像, 登録されていない人物の画像がそれぞれ 50 枚と, それらを 80%, 90%, 110%, 120% に伸縮した画像の計 500 枚とした. 画像の伸縮は MacOS 9.2 上で Adobe Photoshop 5.0

を用いて bi-cubic 内挿法で行われた. スケール辞書は, モデルデータベース中の 20 枚 (= N) の画像と参照グラフを 11 段階 (= T) に伸縮させたもの (平均顔グラフを 1 として $s_T = 1/1.5, 1/1.4, 1/1.3, 1/1.2, 1/1.1, 1, 1/0.9, 1/0.8, 1/0.7, 1/0.6, 1/0.5$ で伸縮) からそれぞれ求めたネットワーク応答により作成された. また, ガボールウェーブレット変換 (式 (2)) は, σ を 1, a を 2.5, u を π , カーネルの大きさのファクタ数 N_j を 2, 方向数 N_θ を 8 とし, 総数 16 個のカーネルを用いて行われた. 顔位置検出は, x 方向, y 方向ともに 5 画素間隔でラスタスキャンした. 式 (6) については, λ_{ini} を 4, λ_{fin} を 1, t_{max} を 200 とした. なお, 精度良くスケール変換を行うために選ばれる比率 t_{opt} の候補 t_{cand} の数 c については後述する (4.1 節).

我々は, 本システムの比較対象として, 2 つのシステムを構築した. 1 つ目のシステムは, 11 段階の大きさの参照グラフ (平均顔グラフを 50% から 150% まで 10% 刻みで伸縮したもの) と入力画像を照合し, 最も適合する参照グラフを従来のフレキシブル特徴照合法に用いるもので, 従来システム 1 と呼ぶことにする. 2 つ目のシステムは, 利用する参照グラフは 1 つで, 入力画像をこの参照グラフの大きさに正規化するものである. 我々が構築したシステムは, あらかじめ入力画像を 11 段階 (1/50%, 1/60%, 1/70%, 1/80%, 1/90%, 1/100%, 1/110%, 1/120%, 1/130%, 1/140%, 1/150%) に伸縮したものを用意しておき, それらと参照グラフを照合して, 最も適合する伸縮画像を従来のフレキシブル特徴照合法に用いるというもので, 従来システム 2 と呼ぶことにする. このシステムは, 他の 2 つのシステムより顔位置検出の計算コストが約 11 倍必要であり, また, 画像の伸縮を行うための計算コストも必要になる.

4.1 スケール変換の有効性

ここでは, グラフと入力画像の顔の大きさを合わせるための一般的な方法 (従来システム 1) に対するスケール変換の有効性を, 比率 t_{opt} の候補 t_{cand} の数 c の設定と絡めて検証する. まず, スケール変換の開始から t_{opt} の候補となる c 個の t_{cand} を求めるまでの処理時間を t_c , 1 つのグラフと 1 枚の画像からネットワーク応答を求める時間を t_n , 従来システム 1 で用意する多段階の参照グラフの数を T (スケール変換ではグラフの伸縮率の段階数に相当) とすると, スケール変換と従来システム 1 にかかる処理時間はそれぞれ, $t_c + c \times t_n$ と $T \times t_n$ で表される. t_c はスケール辞書の教師データ数にほぼ比例して大きくなるが, 本実験

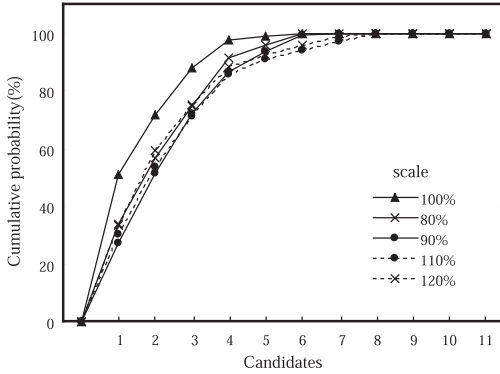


図 7 最適な比率 t_{opt} の候補が持つ確率の累積

Fig. 7 Cumulative probability of candidates for optimal scale ratio.

の設定 ($N = 20, T = 11$) で約 13 ミリ秒と微小である。これに対し、 t_n はガボールウェーブレット変換に要する時間を含むため、現状のシステムで約 470 ミリ秒と非常に大きい。すなわち、 t_c は t_n に対して無視でき、両手法の処理時間は c と T に依存することになる。そこで、 c についてであるが、本システムでは図 7 に示す統計を利用して設定する。図 7 は入力画像にスケール変換を適用したときに求まる r_i の確率を降順に累積したものである。この図から、第 5 候補までの累積確率が 90% を超えていることが分かる。また、5 つの候補の中に t_{opt} として最適なものが必ず含まれることが実験的に確認されている。これらのことから、本実験における s_T と $T (= 11)$ の設定でのスケール変換では、 c を 5 ($< T$) とすることで、従来システム 1 と同精度で t_{opt} を求めることができる。同精度といえる理由は、候補を選択した後の処理が従来システム 1 に基づくためである。ゆえに、スケール変換の方が計算時間が短いため有効であるといえる。

スケール変換の最も有効な点は、上述のように t_c が微小であるため、計算効率がスケール辞書の大きさには依存しないところにある。さらに、あらゆる顔の大きさを確率の形で瞬時に求められる点も有効である。ゆえに、顔の大きさの推定処理を行う大きさの範囲を限定でき、その処理を効率良く進めることができる。これはより細かな精度で顔の大きさを推定する必要がある場合に、特に有効であると考えられる。これに対し従来システム 1 では、大きさの推定精度は参照グラフの数 T に依存するので、計算効率とトレードオフの関係にある。

4.2 特徴点抽出の精度

図 8 に本システムを用いて顔特徴点を抽出した結果を示す。この結果から、特徴量を正規化しているにも

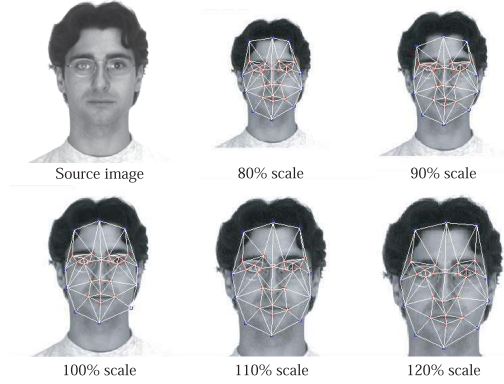


図 8 顔特徴点抽出結果

Fig. 8 Facial feature points extraction results.

表 1 モデルと入力画像のガボール特徴量の差

Table 1 Differences between the Gabor features of the model and those of the input image.

	80%	90%	110%	120%
equation (2)	0.53	0.55	0.50	0.48
equation (1)	0.74	0.57	0.58	0.66

かかわらず、抽出された特徴点の位置が入力画像間で一致しないことが分かる。この原因を表 1 から考察する。表 1 はモデルグラフが持つガボール特徴量とそのモデルに対応した入力画像 (80, 90, 110, 120% 伸縮) から式 (1) と式 (2) によって抽出される特徴量の差の平均を示す。なお、特徴点位置は我々が目視によって画像中で定義したものであり、式 (2) の大きさのファクタ j には画像伸縮率がかけられる。表 1 の上段を見ると値がほぼ同じとなっているため、式 (2) を用いることで顔画像の大きさの変化による空間周波数の変動に影響を受けないガボール特徴量が抽出されていることが確認できる。しかし、値が 0 になっていないことから、完全には正規化されていないことも分かる。これは画像の伸縮による情報落ちが影響していると考えられる。そして、この情報落ちがそれぞれの大きさの入力画像で抽出される特徴点の位置が一致しない原因となっていると思われる。また、この情報落ちの影響は、画像の伸縮によって顔の大きさを正規化する従来システム 2 の特徴点抽出の精度と認識率が低下する可能性を示唆する。

特徴点抽出の精度は、顔特徴点の抽出処理が終了した時点のシャドウグラフと、入力画像の人物に該当するモデルグラフの特徴点の位置誤差を指標とする。すなわち、入力画像はモデルデータベースに登録されている人物のものである。表 2 に 3 つのシステムの特徴点抽出の精度を示す。表中の 80%, 90%, 100%,

表 2 特徴点抽出の精度 (pixel)
Table 2 The accuracy of facial feature points extraction (pixel).

	80%	90%	100%	110%	120%
Proposed system	4.95	5.17	4.61	4.66	4.70
Previous system 1	5.94	5.43	4.77	5.20	5.13
Previous system 2	5.69	5.50	5.11	5.64	6.13

110%, 120%は入力画像の伸縮率を、数値は1特徴点あたりの平均誤差距離を表す。なお、精度の計算の際、本システムと従来システム1に関しては式(5)と同様の距離情報に対する正規化が行われる。この結果から、従来システムでは本システムより、入力画像の伸縮が大きくなるほど抽出誤差も大きくなる傾向があると分かる。これは、従来システム1ではガボール特徴量の正規化を行っていないこと、従来システム2では画像の伸縮による情報落ちが原因であると考えられる。一方、本システムでは顔の大きさの変化にあまり影響を受けずに特徴点を抽出できていることが分かる。なお、顔の大きさの変化に対する処理が適用されないはずの伸縮率100%の入力画像(つまり伸縮なしの入力画像)に対し、3つのシステムの精度に差があるが、これは伸縮率100%の入力画像の顔の大きさが完全に一樣ではないため、これらの画像にも顔の大きさの変化に対する処理が適用されるからである。

4.3 認識率

認識率は、モデルデータベースに登録されている人物の顔画像を入力した場合はその人物の同定成功率であり、未登録の人物の顔画像を入力した場合は、モデルデータベースに登録されていない人物であるとシステムが判定した割合である。図9に3つのシステムの認識率を示す。この図は、顔同定の処理で設定する必要がある閾値を変化させ、それに対する認識率をROC曲線の形で表したものである。この曲線はグラフの右上に位置するほど、その精度が良いといえる。したがって、元画像を伸縮させた入力画像(80%, 90%, 110%, 120%)に対して、本システムの精度向上が確認できる。従来システムにおける精度の低下の原因は、4.2節で述べたものと同様であると考えられる。

5. 結 論

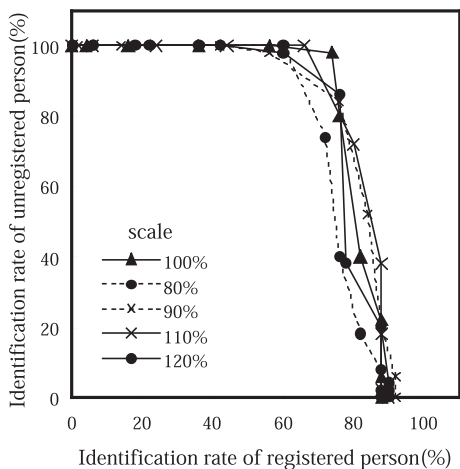
我々は、より良いヒューマンインタフェースの構築のために、任意の見え方の顔に対してロバストに認識できるシステムの実現をめざしている。本稿では、顔の大きさという観点での見え方に着目し、効率的な顔の大きさの推定と、顔の大きさの変化の影響を吸収した特徴量の抽出を行う新たな手法を提案した。そして、

この手法を顔位置検出、顔特徴点抽出、顔同定を一連の処理で行うシステムに導入した。我々は、このシステムの評価実験を行い、本手法によって入力画像の顔の大きさが変化する場合にも効率良く、かつ精度良く顔特徴点が抽出され、顔同定の精度の低下が抑えられることを確認した。また、実験結果から入力画像の顔の大きさの変化に対して、入力画像自体の大きさを正規化するのではなく、ガボールカーネルの大きさを正規化して特徴量を抽出することが有効であることが明らかになった。さらに、従来は5段階の大きさのガボールカーネルを利用して顔の大きさの変化に対応させるのが一般的であったが、本システムはそれを2段階に削減しても認識精度を維持することができた。ガボールウェーブレット変換の計算コストは非常に高く、カーネルの数に比例するため、この削減が計算効率の面に関して非常に有効と思われる。

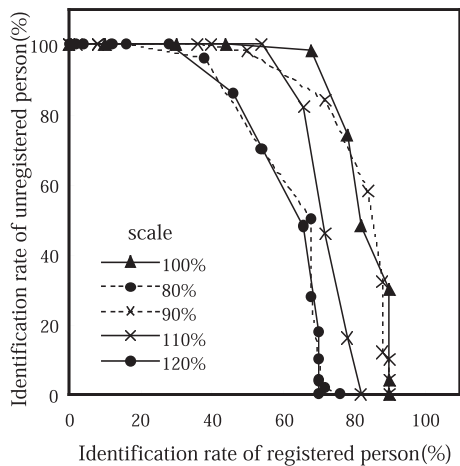
効率面で問題となるのは、顔位置検出における処理で、現在の手法で正確な顔位置を検出するためには画像全体にガボールウェーブレット変換を施さなければならないことである。これに対しては、ラスタスキャン以外の探索アルゴリズムを利用して、グラフが移動するごとに、そのノードまわりに対してのみガボールウェーブレット変換を施すような処理へと変更するか、もしくは画像をカラーに拡張してファジーパターン照合を用いた手法²³⁾を適用することなどが考えられる。

本システムは、モデルデータベースに登録されていない人物の顔画像を「未登録」と正確に判別できることから、未登録の人物の顔特徴点を正しく抽出できれば、モデルグラフをデータベースに自動登録することが可能である。今後は、本手法を応用し、顔向きの変化や顔部位のオクルージョンなどに対応した個人識別システムを構築する予定である。また、顔の表情動画像から連続的な表情変化を認識する手法^{11), 24)}に応用することも考えている。

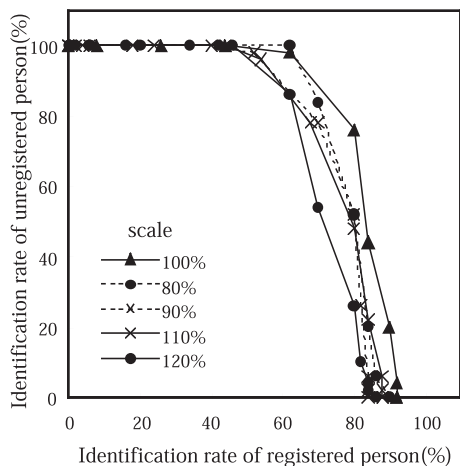
謝辞 本稿で述べた研究の一部は日本学術振興会未来開拓研究推進事業「感性的ヒューマンインタフェース」(プロジェクト番号 JSPS-RFTF99P01404)の援助を受けた。



(a) Proposed system



(b) Previous system 1



(c) Previous system 2

図 9 認識率

Fig. 9 Identification rate.

参考文献

- 1) 本名信之：ノンバーバルコミュニケーション，大修館書店 (1981).
- 2) Song, X., Lee, C., Xu, G. and Tsuji, S.: Extracting facial features with partial feature template, *Proc. Asian Conference on Computer Vision*, pp.751-754 (1994).
- 3) Turk, M. and Pentland, A.: Eigenface for recognition, *Journal of Cognitive Neuroscience*, Vol.3, No.1, pp.71-86 (1991).
- 4) Belhumeur, P., Hespanha, J. and Kriegman, D.: Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.19, No.7, pp.711-720 (1997).
- 5) Ayinde, O. and Yang, Y.-H.: Face Recognition Approach Based on Rank Correlation of Gabor-Filtered Images, *Pattern Recognition*, Vol.35, pp.1275-1289 (2002).
- 6) Daugman, J.G.: Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters, *Journal of the Optical Society of America A*, Vol.2, pp.1160-1169 (1985).
- 7) 中野宏毅，山本鎮男，吉田靖夫：ウェブレットによる信号処理と画像処理，共立出版株式会社 (1999).
- 8) Wiskott, L., Fellous, J.M., Krüger, N. and von der Malsburg, C.: Face recognition and gender determination, *Proc. International Workshop on Automatic Face- and Gesture-Recognition*, pp.92-97 (1995).
- 9) Wiskott, L., Fellous, J.M., Krüger, N. and von der Malsburg, C.: Face recognition by Elastic Bunch Graph Matching, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.19, No.7, pp.775-779 (1997).
- 10) Pramadihanto, D., Iwai, Y. and Yachida, M.: A flexible feature matching for automatic face and facial points detection, *Proc. 14th International Conference on Pattern Recognition*, pp.324-329 (1998).
- 11) Pramadihanto, D., Iwai, Y. and Yachida, M.: Integrated Person Identification and Expression Recognition from Facial Images, *IEICE Trans. Inf. Syst.*, Vol.E84-D, No.7, pp.856-866 (2001).
- 12) 土橋浩慶，岡崎彰夫，高木一圭：顔照合セキュリティシステム *FacePassTM*，東芝レビュー，Vol.57, No.8, pp.48-51 (2002).
- 13) Krüger, N., Pötzsch, M. and von der Malsburg, C.: Determination of face position and pose with a learned representation based on

labelled graphs, *Image and Vision Computing*, Vol.15, pp.665–673 (1997).

- 14) Kotropoulos, C., Tefas, A. and Pitas, I.: Morphological elastic graph matching applied to frontal face authentication under well-controlled and real conditions, *Pattern Recognition*, Vol.33, pp.1935–1947 (2000).
- 15) Lades, M., Vorbrüggen, J.C., Buhmann, J., Lange, J., von der Malsburg, C., Würtz, R.P. and Konen, W.: Distortion invariant object recognition in the Dynamic Link Architecture, *IEEE Trans. Comput.*, Vol.42, pp.300–311 (1993).
- 16) ダデットプラマディハント, 吳 海元, 谷内田正彦: 多様な姿勢を有する入力顔画像からの個人識別, 電子情報通信学会論文誌 D-II, Vol.80, No.8, pp.2232–2238 (1997).
- 17) Lyons, M.J., Budynek, J., Plante, A. and Akamatsu, S.: Classifying Facial Attributes using a 2-D Gabor Wavelet Representation and Discriminant Analysis, *Proc. International Conference on Automatic Face and Gesture Recognition*, pp.202–207 (2000).
- 18) Lades, M.: Invariant Object Recognition Based on Dynamical Links, Robust to Scaling, Rotation and Variation of Illumination, Ph.D. Thesis, Ruhr Universität, Bochum, Germany (1995).
- 19) Okada, K., Steffens, J., Maurer, T., Hong, H., Elagin, E., Neven, H. and von der Malsburg, C.: The Bochum/USC Face Recognition System: And How it Fared in the FERET PhaseIII Test, *Face Recognition: From Theory to Applications*, Springer-Verlag, Sterling, UK (1998).
- 20) 森 於菟, 小川 興三, 大内 弘, 森 富: 解剖学, 金原出版 (1965).
- 21) 藤田恒太郎: 人体解剖学, 南江堂 (1979).
- 22) Martinez, A.M. and Benavente, R.: The AR face database, CVC Technical Report24 (1998).
- 23) 吳 海元, 陳 謙, 谷内田正彦: ファジイパターン照合を用いた色彩画像からの顔検出システム, 電子情報通信学会論文誌 D-II, Vol.80, No.7, pp.1774–1785 (1997).
- 24) 王 梅, 岩井儀雄, 谷内田正彦: 動きモデルによる顔特徴点追跡と表情認識, 電気学会論文誌 D, Vol.119-D, No.5, pp.699–706 (1998).

(平成 14 年 5 月 31 日受付)

(平成 15 年 4 月 3 日採録)



平山 高嗣

平成 12 年金沢大学工学部電気・情報工学科卒業。平成 14 年大阪大学大学院基礎工学研究科博士前期課程修了。現在, 同大学院基礎工学研究科博士後期課程在学中。顔画像処理, パターン認識に関する研究に従事。



岩井 儀雄 (正会員)

平成 4 年大阪大学基礎工学部情報工学科卒業。平成 6 年同大学大学院基礎工学研究科修士課程修了。平成 9 年同大学院基礎工学研究科博士課程修了。同年同大学院基礎工学研究科助手。平成 15 年同大学院基礎工学研究科助教授。コンピュータビジョン, パターン認識に関する研究に従事。博士(工学)。IEEE, 電子情報通信学会, 日本ロボット学会各会員。



谷内田正彦 (正会員)

昭和 46 年大阪大学大学院工学研究科修士課程修了。同年同大学基礎工学部制御工学科助手。同助教授を経て同学部情報工学科教授, 平成 6 年同学部システム工学科教授。現在, 同大学大学院基礎工学研究科教授。昭和 42~43 年デンマーク原子力研究所留学。昭和 47~48 年米イリノイ大学にて Research Associate。昭和 55~56 年西独ハンブルグ大学 Research Fellow。昭和 57 年米ミネソタ大学 CDC Professor。ロボット学会, 人工知能学会等会員。著書「ロボットビジョン」(昭晃堂, 大川出版賞受賞)、「コンピュータビジョン」(丸善, 編著)等。コンピュータビジョン, 画像処理, 人工知能, 移動ロボット等の研究を行っている。工学博士。