

人間と協調して行動するロボットにとって、リアルタイム顔認識は不可欠な視覚機能である。一般に、顔の向きや表情変化はロボット側から制御できないために、ロボットへ搭載される顔認識にはこれらの変動に対するロバスト(頑健)性が要求される。ところが、1枚の静止画像を用いた従来法では、この要求を満足させることは難しかった。これに対して、当社では、異なる視点からの複数画像を用いて顔認識を行う。このような複数画像はロボットがみずから移動したり、逆に人間がロボットへ近づくなどの行動により獲得できる。この多視点画像を用いた認識は擬似的な三次元認識に相当し、安定な認識を可能とする。ロボットへの搭載を模擬した評価実験では、当社方式が従来法に比べて約3倍(当社比)の認識性能を示した。

Face recognition is an important function of robot vision. Face recognition implemented on a robot requires robustness to changes in facial expression and facial direction, because the robot cannot control these changes.

Toshiba has developed a face recognition technology that satisfies these requirements using sequential images. Our method can accurately recognize a face based on information on the three-dimensional shape from the sequential images. Experimental results have verified that our method is robust in comparison with the conventional methods using a single static image.

1 まえがき

社会生活を営む人間にとって、相手がだれであるということとは不可欠な情報である。このため、人間と協調的に働く機械にも顔認識⁽¹⁾を行わせたいという願望は古くから見られ、究極の機械であるロボットの描写には顔認識の機能が自然と組み込まれていた。

個人識別手段としては、顔、指紋、光彩、音声などの様々な方法が考えられる。これらのいくつかは既に実用化され、精度的にも顔認識より優れている。それにもかかわらず、なぜ顔認識なのだろうか。まず、顔画像認識はユーザーに意識させないで行えるという大きな利点を持っている。これは、顔が常に外部に露出しているために得られる特性で、他の手段にはない特長である。更に、接触動作を伴わないために、短い時間で距離に依存せずに認識が行える。他の利点としては、識別結果に加えて、性別、国籍、表情、視線などのノンバーバル(非言語)情報が同時に得られることが挙げられる。これほど豊富なノンバーバル情報を同時に獲得できる識別手段は考えられない。

このように顔画像認識は、識別精度という面では優位性に欠けるものの、総合的にはロボット向けの個人識別機能として極めて有効であることがわかる。

2 顔認識に要求される機能

顔認識のロボットへの搭載例の典型は図1に示すようになると考えられる。まず、ロボットの視覚部に埋め込まれたカメラから入力された画像中から、リアルタイムで顔領域を検出する。次に検出した顔パターンから、目の前の人物がだれであるかを認識する。ロボットと人間の位置関係により認識が難しい場合には、正しく認識できる位置にロボットみずから移動する。このような適用状況を考えた場合に、顔認識に要求される特性は、顔の向きや表情変化に対するロバスト性である。一般に、顔の向きや表情などはロボット側から制御できないために、これらの影響を受けにくい認識アルゴリズムは必須と言える。また、どのような環境下でも安定な



図1. 顔認識機能を搭載したロボット 目の前の人物がだれであることを認識する。

Face recognition for robot vision

認識を行うためには、照明条件の変動への対処も大きな課題である。更に、顔のリアルタイム学習機能も必要である。ここで、ロボットに搭載される顔認識に要求される条件を整理すると以下ようになる。

- (1) 顔の向きや表情変化に対するロバスト性
- (2) 照明変動に対するロバスト性
- (3) 顔の学習機能を持つこと

3 多視点画像を用いた顔認識

当社では、前章で挙げた条件を満足する顔認識アルゴリズムを開発した。この方法の大きな特長は、1枚の静止画像を用いる従来法に対して、異なる視点の複数画像を用いて認識する点にある。

3.1 基本アイデア

ロボットから見た顔パターンは、顔の向き、表情変化、あるいは人間とロボットとの相対位置関係により大きく変化する。このために1枚の静止画像だけを用いた従来法では、これらの変動を吸収できず安定な認識が難しい。これに対して当社方式では、複数の顔パターンを用いて認識を行う。このような複数パターンは、次の二つの方法で獲得することができる。まず、人間側のロボットへの接近、あるいはのぞき込みなどの行動に期待した方法である(図2)。ロボットが



図2．人間側が動く場合 人間が動くことで複数パターンを得る。
Scene I : In case of human movement

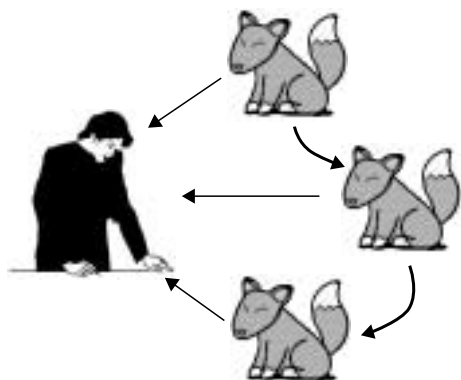


図3．ロボット側が移動する場合 ロボットが自律的に移動することで複数パターンを得る。
Scene II : In case of robot movement

音声を発して人間の注意を引く場合は、これに当てはまる。一方、自律移動可能なロボットでは、自身が移動することで同様に複数パターンを獲得することができる(図3)。このように能動的に望みの画像を獲得できる点は、移動機構を持たないパソコンなどの情報機器への搭載とは大きく異なる。

このような複数の顔パターンを用いた認識は、顔を多視点から見て三次元的に認識することに相当し、顔の向きや表情変化に対して安定な認識が期待できる。

当社方式と1枚の静止画像を用いた従来法との比較を図4に示す。顔パターンを一次元に展開したベクトルと考えると、ある視点から見た顔パターンは特徴空間の、ある1点(黒丸)に対応する。従来法では、入力パターンと各人物のパターン分布との統計距離を求め、もっとも近い分布に該当する人物を本人とする。ところが、顔パターンは視点及び顔の向き、表情により容易に変動する。これに応じて黒丸の位置も変動するため、入力パターンと各分布までの距離は不安定である。一方、当社方式では、複数視点1~Nの入力パターン分布と辞書パターン分布の類似度に基づいて識別を行うため、上記変動の影響を受けにくいことがわかる。

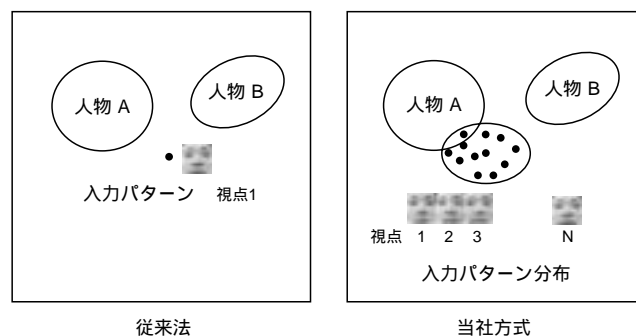


図4．従来法(左)と当社方式(右)の比較 従来法では1枚の静止パターンだけを用いて認識を行う。これに対して、当社方式では複数視点のパターンを用いる。

Comparison of conventional and newly developed methods

3.2 認識アルゴリズム

二つのパターン分布の類似度は相互部分空間法⁽²⁾を用いて計算する。このために入力パターン分布と辞書パターン分布を、図5に示すようにパターン空間中の線形部分空間PとQで表す⁽³⁾。これにより二つの分布の近さは、二つの部分空間の成す正準角で定義できる。この角度は二つのベクトルの成す角度の多次元への拡張であり、二つのN次元部分空間に対してN個の正準角が求まる。二つの部分空間が完全に一致しているときには、正準角度はすべて0度となり、両者が離れるにつれて正準角は大きくなっていく。正準角の累積値を二つの顔パターン分布の類似度とする。

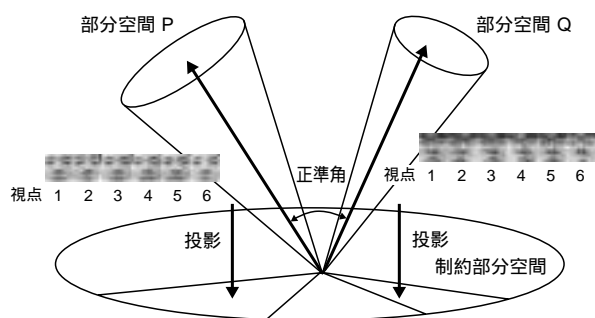


図5. 当社方式の概念 各人物のパターン分布を線形部分空間で表す。二つの部分空間の近さが本人らしさとなる。
Concept of newly developed method

3.3 顔の学習機能

当社方式では、何らかのトリガー信号に基づいて顔の学習を開始することができる。例えば、パソコンへの搭載時にはユーザーの登録ボタンを押す動作をトリガーとした。ロボットへの搭載時には、頭をなでるなどの人間側からのロボットへの働きかけをトリガーとして利用できると考えられる。いったん学習を開始すると、自動的に顔パターンを切り出し、所定枚数に達した時点で辞書部分空間を生成する。

更に、ロボットが認識結果を有効に活用するためには、記憶している顔パターンと名前などの他情報とをリンクさせる必要がある。このためには音声認識などが有効であると考えられるが、学習機能の実現は顔認識だけの問題ではなく、より高次の処理との連携が必要であると考えられる。

3.4 照明変化への対応

照明変化の影響に対しては、制約相互部分空間法⁽²⁾を適用することで対処できる。この方法では図5に示すように、あらかじめ生成した照明変動を含まない制約部分空間へ、入力部分空間と辞書部分空間を射影し、その後射影された二つの部分空間に対して相互部分空間法を適用して類似度を求める。

4 顔認識処理の流れ

当社の顔認識処理の流れを図6に示す。まず、CCD(電荷結合素子)カメラから入力された画像中から顔領域を抽出する。次に、分離度フィルタとパターン照合の組合せにより両瞳(ひとみ)、両鼻孔を検出する。更に、検出されたこれらの特徴点を基準に正規化された顔パターンを抽出する。この一連の処理を視点移動しながら繰り返し行い、正規化パターンが抽出されるたびに入力部分空間を更新する。入力部分空間と辞書部分空間を制約部分空間へ射影し、射影された二つの部分空間の成す正準角を類似度とする。全登録人物に対して上述の類似度算出を行い、得られた類似度の中でもっとも類似度が高い人物を本人と認識する。

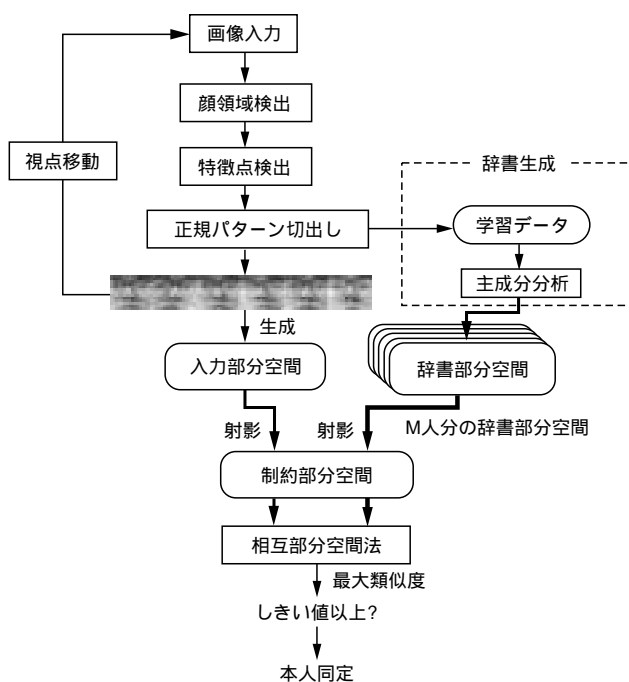


図6. 認識処理の流れ 画像入力から認識結果出力までの流れを示す。
Flow of recognition process

5 評価実験

まず、顔検出及び特徴点抽出の基本動作の確認を行った。このために、図7に示すような小型カメラを搭載したラジコン車を用意し、人間の操作により(I)人間が近づく状況(II)ロボットが動く状況を想定した画像列を収集した。図8は状況(I)に対する処理結果、図9は状況(II)に対する処理結果を示している。図中の矩形(くけい)は検出された顔領域、丸は瞳と鼻孔の位置を示している。ともに異なる視点の複数画像を獲得できているのがわかる。



図7. 無線カメラを搭載したラジコン車 この車を使って人間が近づく場合(状況I)と、ロボットが動く場合(状況II)の画像列を収集した。
Radio-controlled model car with camera



図8．状況Ⅰ 人がロボットに近づく際に獲得した画像列である。
Scene I : Sequential images of human movement



図9．状況Ⅱ ロボットが移動して獲得した画像列を示す。
Scene II : Sequential images of robot movement

従来の静止画像による認識法と動画像を用いた当社認識法の性能比較を図10に示す。発話しながら顔の向きを変化させた顔画像を対象にしているため、従来法では、顔の向きや発話の影響を受けて類似度が頻繁に低下している。これに対して当社方式は類似度が安定していることがわかる。

また、ロボットへの搭載を模擬して、カメラの前に着席した被験者に顔向きを上下左右均等に向けるように指示したうえで認識実験を行った。図11は50人に対する本人排除率と他人受率率の関係を示している。両者が一致するクロスポイントでは、当社方式は従来法に比べ約3倍(当社比)の性能向上が実現できた。この評価は人間側が動いた場合であるが、ロボット自身が動く場合でも同様の結果が期待できる。ただし、ロボットがどのような評価に基づいて移動すれば効率よく多視点画像を獲得できるかは、今後の研究課題と言える。

6 あとがき

この論文では、ロボットに搭載される顔認識に必要な機

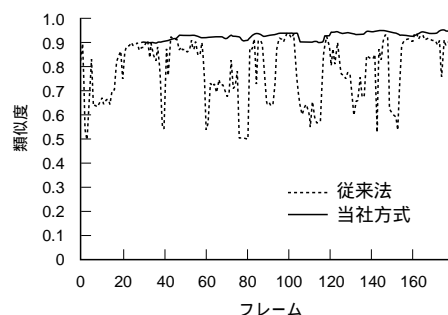


図10．類似度の時間変化 当社方式の類似度は安定している。
Changes in degree of similarity

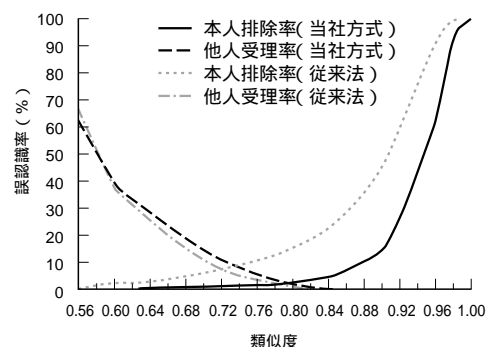


図11．認識性能の比較 クロスポイントにおいて、当社方式は従来法に比べて3倍の性能を示している。
Recognition performance of newly developed method

能を整理したうえで、動画像から得られる複数画像を用いた顔認識法を述べた。この方法の有効性は、ロボット搭載を模擬した評価実験において示した。今後は、多視点画像の効率の良い獲得アルゴリズムを検討する。

文 献

- (1) 赤松 茂 .コンピュータによる顔の認識の研究動向 .電子情報通信学会誌 . 80 ,3 ,1997 ,p.257 - 266 .
- (2) 福井和広 ,ほか . 制約相互部分空間法を用いた環境変化にロバストな顔画像認識 - 照明変動を抑える制約部分空間の学習 . 電子情報通信学会論文誌 . J82-D- II ,4 ,1999 ,p.613 - 620 .
- (3) エルッキ・オヤ(小川英光,佐藤 誠訳) . パターン認識と部分空間法 . 産業図書 ,1986 .



福井 和広 FUKUI Kazuhiro

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。
画像処理の研究・開発に従事。電子情報通信学会会員。
Multimedia Lab.



山口 修 YAMAGUCHI Osamu

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。
画像処理の研究・開発に従事。電子情報通信学会、情報処理学会会員。
Multimedia Lab.