

フレンドリーアーティファクトのための ロバストな顔認識システム

寺田 実†, 尾花 通雅†, 川田 明弘†, 白井 良明†

†東京大学 工学系研究科, ‡大阪大学 工学系研究科

人間と共存するロボットに必要となる, 背景, 照明条件, 向きに
ロバストな顔発見/識別システムについて, 処理の概要を性能評
価もふくめて報告する.

Face Recognition for the Friendly Artifact

Minoru TERADA†, Michimasa OBANA†, Akihiro KAWADA† and Yoshiaki SHIRAI‡

†University of Tokyo, ‡Osaka University

Friendly Artifact is an intelligent robot designed for the close interaction with human. It has various interaction channels such as vision, voice recognition and touch sensors. We implemented a face recognition system for this robot, which detects and identifies human faces within daily environment. In this report we describe the system including its performance.

1 フレンドリーアーティファクト

フレンドリーアーティファクトとは, 人間の通常の生活シーンにおいて, 人間と密に触れ合い共存することを目標とする自立知能ロボットである [1], [2]. 特徴としては, (1) 人間からの指示を待つのではなく, 自分から人間を観察して行動する, (2) 視覚, 聴覚, 触覚などの多様なインタラクションのためのチャネルを備える, などがあげられる.

したがって, 人間の認識が特に重要な役割を占めることになる. 画像認識システムとして考えた場合, 以下のような特徴がある:

- 運動能力を持つため, カメラ位置を自由に選ぶことができる
- 視覚以外のチャネルを通して, 認識対象とイン

タラクションが可能—たとえば顔認識において, 認識対象の人物に対して「もう少しこちらを向いてもらえませんか」などと発話することができる

- 実時間性が重要となる

フレンドリーアーティファクトのプロトタイプとして設計された二種類のロボットのうち, おもに視覚実験用としたのが「すえぞう」(図 1, 図 2)である. 想定した利用形態として, オフィスにおける留守番を考えた. このタスクでは, 視覚によって主人, 顔を知っている来客, 初対面の来客を認識し, それぞれに対して適切な対応をするものである. 通常は部屋の入口を見ており, 入ってきた人の顔を識別することでこの対応を行なうものとした. このタスクにおける顔認識は, 以下のように特徴付けられる:

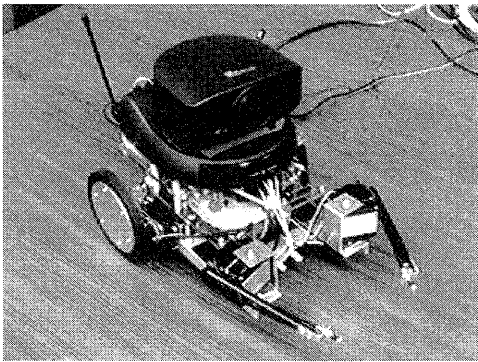


図 1: プロトタイプ「すえぞう」

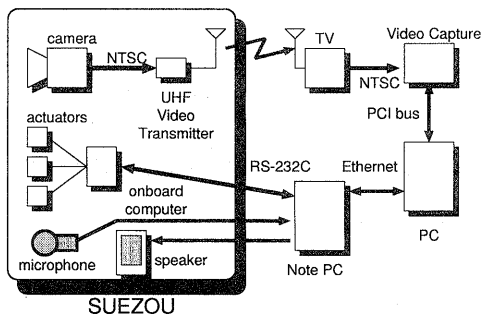


図 2: 「すえぞう」のハードウェア構成

- 背景が複雑
- 顔の位置、向きが変化
- 照明状況が人間の位置によって変化
- 連続画像が利用可能

本稿では、この顔認識システムを、顔発見と顔識別のそれぞれについてロバスト化の観点から説明する。またシステムの実装についても簡単に紹介し、認識システムの性能についても述べる。

2 顔発見

本節では、複雑な背景を持つ単眼カラー画像から、人間の顔を発見するモジュールについて述べる。入力画像は 320×240 ピクセルの大きさで、RGB 各 5 ビットをもつ。発見結果として、画像中の顔のそれぞれについて、両眼の位置と顔の上下端、左右端を出力する。処理結果の例を図 3 に示す。



図 3: 顔発見: 処理結果例

2.1 処理概要

顔発見の処理は、基本的には以下のように進行する:

1. 三値化
入力画像を RGB 基底から YIV 基底に変換し、Y 成分、I 成分に着目して、肌色、黒、その他の色に三値化する。
2. 顔候補領域の選定
黒領域の下に肌色領域がある部分を顔候補として選び出す
3. 候補領域内での目の発見
候補領域それぞれに対し、両眼の発見を試みる。
4. 顔の切り出し
顔の上下端、左右端を求める
以下では、画像条件の変化にロバストに対応するための手段について説明を行なう。

2.2 顔の位置、大きさ、顔の向き

我々が対象とする状況では、認識対象人物はカメラに対して一定の位置にいるわけではないから、位置、大きさ、顔の向きの違いを吸収する必要がある。さらに、標準的な顔のパターンを画面内で探すアプローチをとったとすると、こうした変化に対応するには多くのパターンを用意する必要がある。

我々は、最初は顔の形にとらわれず、「上に黒領域があり、接続して下に肌色領域がある」場所を候補として進めることにした。それ以降の段階で両眼の探索によって候補をしばっていくのである。

2.3 複雑な背景

想定している環境は一般のオフィスであり(実際の実験は大学の研究室で行なった), 背景には書架, 計算機など多様な物体が存在する。(図4(A)に対象画像例を示す。) それらの背景のうち, 三値化の結果として肌色や黒に分類される領域はかなり多い。(図4(B)に三値化の結果を示す。) これにより, 顔候補領域に多くの領域が含まれるようになり, 発見速度や性能に問題を生じる。また, 顔を無事発見できてもその切り出しの段階において, 背景と顔の領域が融合してしまい, 誤った切り出し結果をもたらすことも多かった。以上を解決するため, 我々が採用したのは以下の二つの方法である。

2.3.1 影響度伝搬

処理の最初に顔領域の候補としてとりあげるのは, 「上に黒領域があり, 接続して下に肌色領域がある」場所である。こうした基準を満たす場所だけを候補とすれば, 背景で単独に存在する黒領域や肌色領域を除外することができる。

この基準に合致する領域を捜し出すのに, 個々の黒領域や肌色領域に対して探索を行なう方法もあるが, 領域数によって所要時間が増大する問題点がある。そこで, 画像を2回走査することでこの条件に合致した領域だけを残すアルゴリズムを工夫した。

基本的な考え方は以下の通り: 各ピクセル (x, y) ごとに黒の影響度 $I_b(x, y)$ と肌色の影響度 $I_s(x, y)$ を定義する。 $I_b(x, y)$ の値は, そのピクセルの色(黒かどうか)と, 真上のピクセルにおける黒の影響度から計算する。そのピクセルが黒の場合には, (ある閾値を上限として) $I_b(x, y)$ に一定の値が加えられる。そのピクセルが肌色の場合には, 黒の影響度は真上のピクセルにおける値を保ち続ける。そのピクセルがその他の色の場合には, 一定の減衰率により減少させる。

このように画像を上から下に走査して計算される黒の影響度は, 黒領域とその下に隣接する肌色領域では高い値をとり, それ以外の場所では指数的に減少していく。肌色の影響度についても, 計算の方向が下から上にむかう以外は同様である。

こうして得られた二つの影響度の積を計算すると, 前述の基準を満たす顔領域候補だけを残すことができる。(図4(C)にその結果を示す。) 実際の処理においては, 横方向への影響度伝搬も考慮している。



図4: 顔発見:三値化

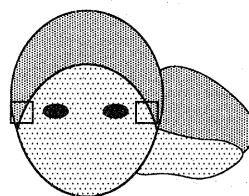


図5: 顔発見:顔の左右端の切り出し

2.3.2 顔切り出しにおける背景の除外

発見処理の最終段階において, 発見した顔に外接する長方形領域を求める必要がある。しかし, たとえば背景に黒い物体があり, それが認識対象の頭髮部分と重なっていた場合, 単純なラベリング結果に基づく切り出し処理ではそれを除外できない。

我々が採用した方法は, 頭髮領域に対して鉛直方向をパラメータとするヒストグラムをとり, もっとも急激に値が増加する場所(通常は頭髮領域の上端)をもって顔の上端とする。これにより, 細い縦線(柱, ドアの枠など)が重なった場合については対処が可能である。一般的な処理は難しい。

また, 顔の左右端の切り出しにおいても, 背景と融合した場合には大きな誤差を生じる。これについては, 眼の位置から左右に水平線をたどり, 頭髮領域との交点をもって顔の左右端とすることにした(図5)。正確な左右端とは差があるが, この方法では顔の内部にある情報のみから計算可能であり, 背景の影響を受けない点が長所である。

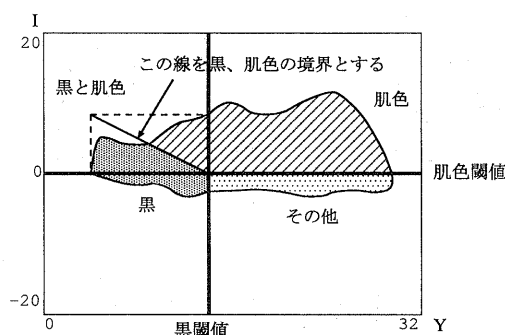


図 6: 顔発見: 「黒かつ肌色」領域の再分割

2.4 照明条件の変化

室内を対象にするとしても、天井の照明や窓からの光などにより室内での位置や顔の向きによって画像は変化する。直射日光と異なり、それほどはっきりした影は出ないが、認識に対する影響はそれでも大きい。そうした変化への対処について述べる。

2.4.1 三値化の閾値

YIQ 基底のもとで、Y のある値を黒かどうかの閾値とし、I のある値を肌色の閾値とした。これですべての色を四分割することができる。そのうち問題となるのが「黒かつ肌色」の領域である。この領域は、肌色の部分が陰になった場合におおく見られ、顔発見においては極めて重要であることが分かった。顔において、我々が目の周囲や鼻の周囲など通常「黒」と認識している部分はじつはこの領域である。(図 4 (D)) この領域を一括して肌色領域とみなすと、照明条件によっては顔の中が肌色一色になってしまう一方、一括して黒とみなすと目と眉が一体化したり、あるいは髪とくっついてしまう。

そこで、画像全体のピクセルの Y-I 平面上での分布図を考え、問題となる四半領域については分布から決まる直線を境界として黒領域と肌色領域に分割するようにした (図 6)。これによって、照明の変化に対してもっとも敏感な閾値をある程度自動的に調整することが可能になった。

2.4.2 目の周辺での再三値化

前述の閾値調整処理は画像全体に対して行なうため、顔の明暗とは必ずしも直結しない。本方式では、両眼の発見が非常に重要であるため、目が髪につな

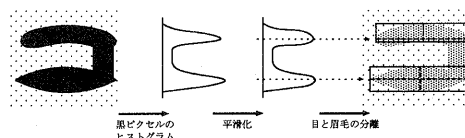


図 7: 顔発見: 目と眉の分離

がってしまう事態は避けなければならない。このために、目の探索に失敗した場合、探索範囲 (顔の領域とおおむね一致する) に対して、Y 成分のヒストグラムに基づいて黒の閾値を再設定する。

2.4.3 眉と目の関係

目の周囲の陰により、眉と目がつながってしまうことがある。前項の再三値化を行なっても解決されない場合、それらを別の方法で分離する必要がある。該当する領域に対して鉛直方向をパラメータとする黒のヒストグラムを作成し、中間に極小点が存在する場合には、そこで上下に領域を分割している (図 7)。

2.5 個人の髪型による問題点

頭髮の黒領域を手がかりとして発見を行なうため、髪型によってはいくつかの問題点があった。ここではその対処を述べる。

2.5.1 眉と前髪の関係

顔の中で目を探す場合、眉との区別が問題になる。単純なケースでは上から二番目のペアが両眼になるのだが、眉と前髪がつながってしまうのは非常に良く見られる現象であり、そうした場合には一番上のペアを採用しなければならない。

まず以下のルールを考慮して候補ペアを列挙する。

- ペアの大きさが近い
- 黒領域の横幅とペアの間隔の関係
- ペアを結ぶ線の傾き

こうして得た候補が複数ある場合、上に位置する二組に対して、位置関係、鼻や口との関係などにもとづくルールを適用して最善の候補を求めている。

2.5.2 分割された髪領域

髪を分けている場合や、上からの照明によって反射がおきている場合などには、髪領域が分かれてし



図 8: 顔識別: 処理結果例 (上段: 入力画像, 下段: 識別結果)

まうことがある。これに対しては、両眼からななめ上方向に黒領域を探索し、ことなる領域が発見された場合には統合する処理を行なっている。

3 顔識別

本節では、前節の顔発見モジュールによって得られた顔画像がだれの顔であるかを識別する処理について述べる。処理結果の例を図 8 に示す。

3.1 処理概要

顔識別の処理は、基本的には以下のように進行する:

1. テンプレートの採取
認識対象となる人物について、あらかじめテンプレートセットを用意する。テンプレートは方向や照明の変化を吸収するため一人につき 20 枚用意した (図 9)。
2. 入力画像の正規化
認識すべき顔画像の位置、大きさ、傾き、明るさを正規化する。
3. テンプレートとの照合
4. 結果の判定
未知の顔であるという結論も含めて結果を判定する。

各種の外乱に対してロバストに動作するよう、本研究においては (1) マッチングの対象領域の選択方式、(2) マッチングアルゴリズムに対する工夫、の二点を行なった。以下でこれらについて説明する。

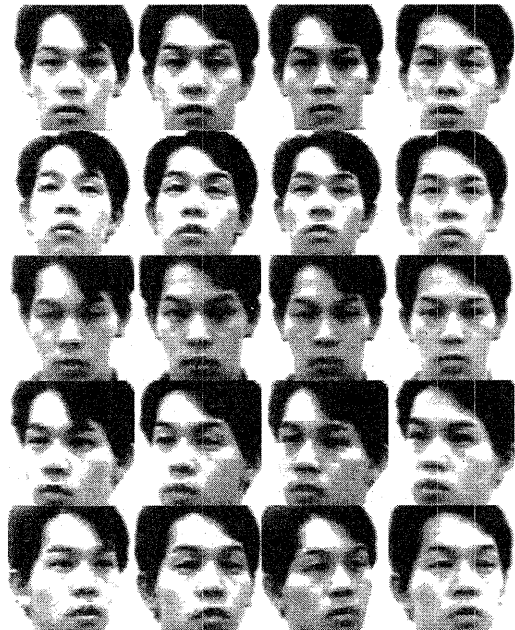


図 9: 顔識別: テンプレートセット



図 10: 顔識別: 正規化に用いる標準顔

3.2 標準顔とのマッチングによる正規化

テンプレート照合による識別について、結果を大きく左右するのは入力画像の諸条件 (明るさ、コントラスト、位置、大きさ、傾き) の正規化である。前節の顔発見により目の位置が得られるが、識別に際してあらためて正確な正規化を行なう。

まず、顔発見モジュールの提供する情報に従い、大まかな正規化を行なう (位置、大きさ、傾き)。次に、明るさを正規化し (肌色領域の平均、分散による)、そのあとあらかじめ用意してある標準顔とマッチングすることによって位置、大きさを正確に合わせる。このとき、顔の向き (顔発見モジュールによって提供される) にもとづいて三種類の標準顔 (図 10) を選ぶことによって、より正確な正規化が期待できる。

標準顔とのマッチングは二段階に分けて行なう。まず、両眼の中心付近 (図 11 (A)) だけを対象領域

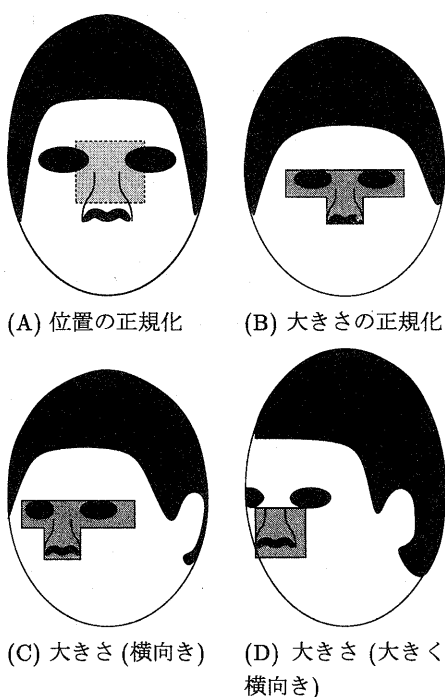


図 11: 顔識別: 正規化時のマッチング領域

としてマッチングを行ない、位置を正確に合わせる。対象領域をこのように決定した理由は、この領域は両眼の中心点を中心として放射状の形態になっているため、拡大率の差による誤差を小さくたもって位置合わせができるからである。

次に、両眼を含む T 字型領域 (図 11 (B)) を対象領域としてマッチングを行ない、大きさの正規化を行なう。こちらは大きさの差に敏感な領域を利用している。また、横向きの顔の場合には、用いる T 字型領域も非対称なものとしている (図 11 (C))。さらに大きく横向きになった場合、両目の間隔が小さくなり誤差が増大するため、鼻周辺だけを対象領域とする (図 11 (D))。

3.3 テンプレートとのマッチングによる識別

3.3.1 顔の変化への対応

顔は時間経過によって変化するものである。短期的には表情変化の影響が大きく、中期的には髪型の変化、そして長期的には加齢による変化がある。このうちの前二者に対応するため、以下の手法を行

なった。

まず髪型の変化に対しては、マッチング領域の適応的選択で対処した。具体的には、テンプレートとのマッチの際、髪領域とそれ以外とを分けて相関計算し (図 12 (A))、髪領域のスコアが良かった場合のみ全体のスコアに算入する方針をとった。

次に表情変化については、もっとも影響の大きい口の周辺について、通常の SSD ではなく、位置のずれを吸収するように修正した相関計算を用いた。具体的には、参照画像の単位 (タテヨコ L の領域を平均化して単位とする) と入力画像の単位の相関値を計算する際、入力画像を上下左右に $L/2$ 移動させた領域を (元画像から平均計算しなおして) 作成し、基本の位置のものと合わせて 5 つの単位との Y 成分の差の二乗のうちの最小のものを採用するのである。これは結果としてピクセルごとに独立にずらしを導入することとなり、変形による影響を回避できる。(この計算方法を SSD2 と呼んでおく。)

さらに、表情変化に対応するため、入力画像の一部を相関計算から除外する処理を行なう。これは、テンプレートセットとの照合を行なった結果において、一位、二位の候補テンプレートとの差が大きい入力画像の部分を除外するものである。

3.3.2 横向きの顔への対応

あらかじめ用意してあるテンプレートセットでは、顔の方向については正面と上下左右の合計五方向だけしか収録していない。そのため、想定した角度以外の入力画像についてはどうしてもずれが発生する。とくに大きく横を向いた顔に対しては、対策が必要になる。

まず、前節で述べた SSD2 による相関値を用いることにより、顔の三次元性に由来する構成要素のずれをある程度吸収できる。また、カメラから遠い方の半面はずれがさらに大きくなるため、大きく横を向いた顔に対しては近い側の半面だけをマッチングの対象領域とすることで精度を保っている (図 12 (B))。

3.4 未知顔への対応

既知の顔とのマッチングだけでは、未知であるという判断を行なうのは難しい。そこで、マッチングのスコアにおける上位二つに着目し、それらが十分離れている場合に限って首位の顔であると判定し、

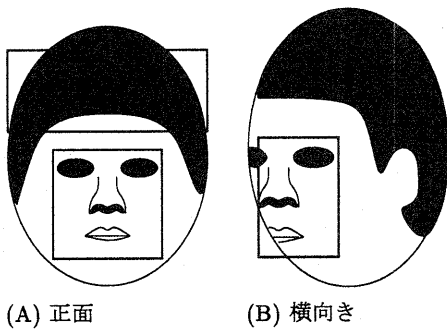


図 12: 顔識別: 識別時のマッチング領域

そうでない場合には未知と判定することにした。具体的には、一位のスコアを $minSSD$ 、二位のスコアを $2ndSSD$ として (二位の候補の選定には、一位と同一人物のテンプレートは除外する),

$$minSSD > T_1$$

の場合には認識不能と判定し,

$$minSSD/2ndSSD > T_2$$

の場合には未知顔と判定する。 T_2 については、実験によって 0.65 とした。

3.5 画像系列の利用

入力画像はビデオカメラからの連続画像であるので、それを利用して、連続して同一の結論が出た場合に限り最終判定とする方法をこころみた。ただし、この方法は現状の処理性能においては反応性をそこなう面もあり、なお検討を要するといえる。

4 実装

4.1 システム構成

本システムは PC 互換機 (Pentium PRO 200MHz) で Linux の上に開発した。画像入力は Argocraft 社製のボードを利用し、キャプチャのソフトウェアはカーネルモジュールとして動作する。そのうえに画像データの共有のためのプログラム *imgd* があり、認識モジュールからの画像データの要求に対応する。画像データは *shared memory* を利用して、コピーすることなくモジュール間で共有可能である。とくに、画像番号 (フレーム ID) を指定

して画像を要求できるため、複数のモジュールが同じ画像を処理することができる。

顔発見、顔識別はそれぞれ独立したモジュール (*unix* プロセス) として C 言語で実現され、*imgd* と通信することで画像の取得を行なう。それらモジュールを統合する形で、トップレベルにスクリプトと名付けたプログラムが存在し、各モジュールの動作開始、認識結果の取得、それに応じたアーティファクトの動作などを制御する。

独立したモジュールとして実装したことにより、発見モジュールが顔を追跡しつつ、識別モジュールが処理を行なうといったゆるやかに関連した動作が可能になった。

4.2 性能

4.2.1 顔発見

動作速度は、単純な背景の場合、3.3 frame/sec となった。さらに、前フレームでの発見結果を利用して探索範囲をしばる処理を行なうと、4.5 frame/sec に向上する。ただし、背景が複雑になる場合には必要時間はこれより増加する。

実験室において、11 人、合計 146 枚の画像に対する発見率を求めたところ、正確に目を検出できたものは 78% であった。

4.2.2 顔識別

テンプレートセットを 6 人ぶん (合計 120 枚) 作成し、それらの人物を対象として総数 284 枚の画像に対して識別実験を行なった。その結果、85.6% が識別に成功している。

また、未登録者 5 名の画像 56 枚について識別実験を行なったところ、正しく “unknown” と判定したのは 75% であった。

位置の正規化については、188 枚の入力画像に対して、97% が成功、大きさの正規化については 90% が成功している。

平均動作速度は、2.36 秒である。

4.3 課題

まず、評価のための実験データが少ないことが問題である。顔発見についてはテレビ放送に適用して実験を行なっている。今後は公開されている顔データベースなどへの適用を行なっていく必要があろう。

つぎに、処理のさらなるロバスト化を目指す必要がある。現状では、室内でのみ動作しているが、たとえば直射日光下などより厳しい条件を考えていく。また、眼鏡の着用や、白髪の人物など本アルゴリズムの前提をこえた条件にたいしても、新しい方法を考えていかなければならない。

5 関連研究

顔認識に関する研究は歴史が長く、非常に多岐に渡っている [3]。その中で本研究の位置付けについては以下のように考えている。

まず顔発見については、通常行なわれている方法のひとつは、輪郭線の抽出からはじめて顔の外形を構成する線分群を発見しそれらを統合するアプローチである。これは輪郭画像を利用する点から照明条件に対する依存性は少ないけれども、背景の複雑さには対処が難しい。

一方、平均的な顔のテンプレートを用意して画像内でのマッチングを行なうアプローチも一般的である。この場合、位置や大きさに関する依存性を解消するための操作が必要になること、背景の複雑さに対しても問題を生じることがあげられる。

色情報を利用した顔発見としては [4] がある。

本研究では、黒い頭髮という前提条件を利用して、色情報だけによる形に依存しないアプローチをとっている点が特色である。また、その後の認識処理で重要な役割を果たす目の検出について、ヒューリスティックなルールを含む多くの処理を行なっている。

[5] では、最初にまばたきの検出による目の発見からスタートするという興味深いアプローチをとっている。

つぎに顔識別については、顔の構成要素の位置を特徴量として比較するアプローチがあるが、これはその構成要素の位置認識をロバストに行なう困難がある。また、最近主流である K-L 変換を用いた固有空間法は、画像正規化の誤差に敏感であるという問題点がある。

本研究では、古典的なテンプレートマッチングをロバストに行なうために、マッチングアルゴリズムの一部修正 (前述の SSD2) と、照合領域のきめこまかい制御を行なったところに特色がある。

また、システムの利用目的として、ロボットのための実時間視覚をターゲットにしている点も特色で

ある。人間とのインタラクションを利用して認識の補助とするなどの新しい手段が可能になっている。

6 まとめ

人間と共存するロボットに必要な、背景、照明条件、向きにロバストな顔発見/識別システムについて、処理の概要を性能評価もふくめて報告した。

今後は問題点の解決とともに、PC クラスタを利用した並列処理による高速化などを行なっていくたい。また、携帯可能な計算機上で動作させることにより、人間の同伴者としてのシステムの構築も進めていきたい。

本研究は、新エネルギー・産業技術総合開発機構 (NEDO) の新規産業創造型提案公募事業によるものです。

参考文献

- [1] Yoshiaki Shirai, Hirochika Inoue, Masayuki Inaba, Minoru Terada, and Yoshisuke Tateyama. Robot as friendly artifact - human recognition and interaction -. In *Proc. of '99 ICAR*, 1999.
- [2] Yukiko Hoshino, Masaki Miyamoto, Yasutada Suzuki, Hideko Yamamoto, Masayuki Inaba, and Hirochika Inoue. Development of friendly artifact - prototype first: Penguin -. In *Proc. of '99 IROS*, 1999.
- [3] Rama Chellappa, Charles L. Wilson, and Saad Sirohey. Human and machine recognition of faces: A survey. *Proc. of the IEEE*, 83(5), 1995.
- [4] 徐 剛 and 杉本 岳夫. カラー画像を用いた実時間顔検出追跡システム. 信学技法, HIP97-11:9-16, 1997.
- [5] Lars-Peter Bala, Kay Talmi, and Jin Liu. Automatic detection and tracking of faces and facial features in video sequences. In *Picture Coding Symposium 1997*, 1997.