

実時間人物発見のための視覚と聴覚を統合した肌の動的学習

山本 俊一[†]戸田 充彦^{††}和田 俊和[‡]奥乃 博^{††}[†] 京都大学工学部情報学科^{††} 京都大学情報学研究科知能情報学専攻[‡] 和歌山大学システム工学部情報通信システム学科

1. はじめに

近年、人間とコミュニケーションを行うロボットや、人間の代わりに仕事を行う自律移動ロボットの研究が盛んに行われている。我々の研究室ではヒューノイドロボット Replie において実時間複数話者追跡³⁾の実装を進めている。Replie の稼働部は、左右の腕（肘の捻り、曲げ伸ばし）と肩（上下、開閉）、首（左右の回転、左右の傾け、傾き）の合計 11 自由度である。耳として SONY のマイク ECM-77S を左右の内耳モデルの内側に 1 台ずつ、目としてコニーのカメラ CN-68MC を目の部分に 1 台ずつ配置している。Replie の全体像を図 1、カメラからの画像を図 2 に示す。

実時間複数話者追跡により、声が聞こえればその方向を向いたり、顔だけで追跡、声だけで追跡、などどこちないものの非言語的なさまざまな挙動が実現できている。しかし、声のする方向を向くことにより照明条件が変わり、肌の発見ができず、音からの定位を基に顔を探すという動作が入ったりする。また、声する方向を向くためにモーターが作動し、その駆動音によって音源検出率が低下することも生ずる。このように、情報統合により、全体的には視覚情報、聴覚情報の信頼性向上があるものの、局所的には信頼性低下が少なからずある。それぞれのセンサーがロバストでなければ、視聴覚の統合が有効に働かないので、実時間話者追跡にはロバストな顔発見が不可欠となる。

単一画像での顔発見手法は、Knowledge-based, Feature-based, テンプレートマッチング, Appearance-based による主に 4 つのアプローチに分かれる¹⁾。我々は画像からの顔候補領域の抽出は、Feature-based の一つである肌色領域抽出を使用する。ただし、ロボット自身が動いたり、外部の要因などにより明るさが変動すると、肌の検出が困難となる。

本研究では、最近傍識別器を用いた色ターゲット検出を実時間複数話者追跡と組み合わせることにより、肌色識別器を動的に環境に適応させることで明るさの変動や個人差に対してロバストな顔発見のシステムを構築する。

本研究では、最近傍識別器を用いた色ターゲット検出²⁾を実時間複数話者追跡と組み合わせることにより、肌色識別器を動的に環境に適応させることで明るさの変動や個人差に対してロバストな顔発見のシステムを構築する。

2. 最近傍識別器を用いた色ターゲット検出

最近傍識別器を用いた色ターゲット検出は、学習と検出が



図 1 Replie 図 2 カメラからの画像 (左), 肌色抽出 (右上), フレーム間差分 (右下)

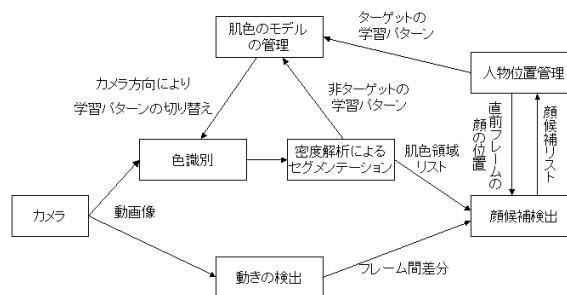


図 3 システム構成

高速に行える。すなわち、肌色をターゲットの色として、また、肌色に間違いやすい色を非ターゲットの色として人手で与えることにより、インタラクティブに肌色を学習する。

しかし、本システムを Replie や他のロボットに実装するには、人間の教示ではなく、ロボット自身が他のセンサー情報を基にターゲットと非ターゲットとを上記学習システムに教示することが必要である。特に、実時間複数話者追跡システムとの統合を考えると、次のような情報が教示のために使用できる：

- ステレオ視覚による顔定位
- 顔追跡による顔領域
- 2 本のマイクロフォンによる話者方向定位
- 混合音からの音源分離による音声検出
- 話者認識による個人同定

3. 顔発見のシステム構成

本研究で構築するシステムの構成を図 3 に示す。

3.1 色識別に用いる色空間

最近傍識別器を用いた色ターゲット検出においては、十分な数のターゲットと非ターゲットの学習パターンを教示すれば RGB 色空間を用いても識別を行っても問題ないと考えられる。よって、 $128 \times 128 \times 128$ のサイズに量子化した RGB 色空間を Look Up Table (LUT) として用いる。

3.2 顔候補検出

顔候補検出のモジュールでは肌色領域検出で得られた肌色

Dynamic Learning of Skin Color for Real-Time Human Detection based on Audio-Visual Integration
by Shunichi Yamamoto (Kyoto Univ.), Mitsuhiko Toda (Kyoto Univ.), Toshikazu Wada (Wakayama Univ.) and Hiroshi G. Okuno (Kyoto Univ.)

領域から顔候補を選択する．顔候補は次のような構成要素によって表現される．

- (1) スクリーン座標系における肌色領域の矩形．
- (2) 肌色領域の重心を顔の位置としカメラ方向を用いてカメラ座標系から絶対座標系に変換し極座標で表す．

顔候補検出の手順を以下に述べる．ただし，スクリーン座標系における肌色領域の矩形の左上と右下の座標をカメラ座標系において極座標で表したものを $(r, \theta_1, \phi_1), (r, \theta_2, \phi_2)$ とする．このとき r はまだ分かっておらず，顔候補確定後に顔の大きさに合わせて r を決定する．

- (1) 初期値として肌色領域検出で得られた肌色領域をすべて顔候補とする．
- (2) カメラから顔までの距離を R 以内とし，顔の肌色部分の横幅の長さを W とすると，

$$\sqrt{2}R\sqrt{1 - \cos(\theta_1 - \theta_2)} \geq W$$

となり，肌色領域の矩形の横幅がこの大きさに適さないならば候補からはずす．顔の縦幅 H についても同様にする．

- (3) これらの顔候補と直前のフレームの顔候補の顔の位置を比較し，距離が一定値以下のものは同一の顔とみなす．直前のフレームには存在したが現在のフレームでは存在しない顔候補があった場合は，直前の顔の位置から一定値以内で，肌色の画素密度が最大となるように局所探索を行う．ただし，顔の位置がスクリーンの端に到達した場合は顔が画面外に出たものとする．

3.3 肌色のモデルの管理

色識別器はターゲットと非ターゲットの学習パターンから構成され，学習パターンに基づいて学習を行う．よって，ターゲットと非ターゲットの学習パターンを切り替えることで複数の肌色のモデルを実現する．学習パターンには次の2つがある．

- (1) あらかじめ用意しておいたターゲットと非ターゲットの学習パターン．

- (2) 実行中の環境から抽出された学習パターン．
- 学習パターンを切り替える基準はさまざまなものが考えられる．カメラの向きによって照明の条件が変わるので，位置によって学習パターンを切り替えることが考えられる．教師なしで学習を行っているので誤った学習を行うことがあるが，時間がたつにつれて古い学習パターンを忘れていけば，誤ったままになるのを防ぐことができる．また，肌色は人によって異なるので，人によって学習パターンを切り替えることも考えられる．

本研究では，位置によって学習パターンを切り替える方法を提案する．(2) の学習パターンは抽出された位置も保存しておき，カメラの方向が変わった際には，カメラの視野の範囲内の位置で抽出された学習パターンとあらかじめ用意しておいた学習パターンをもとに色識別器の学習を行う．

3.4 肌色の動的学習

初期の学習パターンによってある程度肌色を検出できるようにしておくことで，肌色と背景色の識別境界付近の学習パ

ターンを教示して，現在の環境に適応させる．色識別において肌色を学習するにはターゲット (肌色) と非ターゲット (背景色) のデータを与える必要がある．ここでは顔候補の追跡中に学習パターンを収集する方法について述べる．学習パターンを収集するために色識別器，フレーム間差分による動きの検出，顔候補検出により得られた顔領域を用いる．

3.4.1 ターゲット (肌色) の学習パターンの抽出

肌色の学習パターンを収集する手順を以下に述べる．顔候補検出により得られた顔領域すべてに対して以下を実行する．

- (1) 動きの割合が一定値以下ならば肌色を抽出しない．
- (2) 顔領域内の画素すべてに対して色識別を行い，肌色と識別されたものを学習パターンとする．

追跡中の顔候補自体は肌色だと考えられるので，その顔領域の肌色を利用する．また，動いているものは顔候補である可能性が高いと考えられる．動きの割合が小さい領域はその人物が止まっているから，あるいは人物以外の動かない物体を顔候補として検出していると考えられる．誤った肌色を学習するのを未然に防ぐために動きの割合が小さい顔候補は肌色を抽出しないようにする．

3.4.2 非ターゲット (背景色) の学習パターンの抽出

背景色の学習パターンを収集する手順を以下に述べる．

- (1) 動きの割合が一定値以上ならば背景色を抽出しない．
- (2) 顔領域内の画素すべてに対して色識別を行い，背景色と識別されたものを学習パターンとする．

検出された肌色領域のうち顔候補の可能性が低い領域から背景色を抽出する．矩形領域のサイズが小さくかつ，フレーム間差分により得た動きの割合も小さい部分が顔候補の可能性が低いと考えられる．ターゲットの学習パターンの場合とは逆に，誤認識した肌色の領域では動かない物体の可能性が高く，動きが検出されないからである．

4. おわりに

本報告では，肌色領域に基づく実時間での顔発見を行い，色識別器の肌色のモデルを適応，および切り替えるシステムを提案した．今後は，このシステムによる顔発見の評価実験を行う予定である．また，視覚と聴覚のモダリティの肌色の動的学習への利用を検討する予定である．

謝辞 本研究は日本学術振興会科学研究費補助金特定研究 (C) 第14019051 号の助成を受けた．

参考文献

- 1) Ming-Hsuan Yang: "Detecting Faces in Images: A Survey", *IEEE Trans. on PAMI*, vol.24, no.1, pp.34-58, January 2002.
- 2) 和田 俊和: "最近傍識別器を用いた色ターゲット検出 - 「らしさ」に基づかない識別とコンピュータビジョンへの応用 -", 情報処理学会研究報告 CVIM No.134-003, 17-24, 2002.
- 3) Nakadai, K., *et al.*: Real-Time Auditory and Visual Multiple-Object Tracking for Robots, *Proc. of IJCAI-01*, 1425-1432, 2001.