

実時間性を考慮した人物検出処理の最適化についての検討

寺田 健吾[†] 内海 章[‡] 角 康之[‡] 間瀬 健二^{*‡}

[†]金沢工業大学 [‡](株)ATR 知能映像通信研究所

[†]tera@infor.kanazawa-it.ac.jp [‡]{utsumi,sumi,mase}@mic.atr.co.jp

1. はじめに

我々は、興味・状況を共有する人が集まる場所（博物館や学会イベントなど）での人の出会いや対話を促進することを目的としたエージェントサロンと呼ばれるシステムの開発を進めている[1]。実装中のシステムは複数のユーザが同時に利用できる大きなディスプレイを持ち、そこに表示される情報を制御することによって、ユーザ間の創造的議論を促進しようとしている。このようにユーザ間の対話促進を目的とするエージェントサロンでは、各ユーザの反応や操作者の移動・交代などユーザ行動を推定し、状況に応じた柔軟なサービスを提供することが重要となる。そこで、我々はユーザの反応をカメラで撮影し、ユーザの同定及び動き検出を非接触で行うシステムを検討している。今回は特に、画像からの人物検出処理に焦点をあて、計算量の制約に応じた最適化が可能な人物検出手法を提案する。我々は他方で、多数カメラを用いて、実時間で複数人物を追跡するシステムの開発を進めているが[2]、今回提案する人物検出手法は多数カメラを用いたシステムにも容易に応用できると考えられる。本稿では、ユーザ行動推定の概要について述べた後、確率画像を利用した人物検出の最適化について説明し、実画像を使った実験結果を示す。

2. ユーザ行動の実時間推定

実験システムは、大型ディスプレイの上部に CCD カメラを設置し、そのカメラから観測される動画像を用いて、画像からの人物検出処理や、その結果を用いた追跡処理及び人物同定処理などを行う。そしてそれらの出力結果である各ユーザの位置情報をエージェントサロンにフィードバックさせ、エージェントサロンはその後のシステムに活用する。概念図を図 1 に示す。

近年、画像情報から人物の動作を推定するシステムの研究が盛んに行われている[3,4]。これらのシステムの多くはマン・マシンインタラクション等、応用面からの要求により実時間処理を目指している。しかし、動作推定など多くの処理において得られる推定（検出）精度と必要な計算量の間にはトレードオフの関係があり、実時間

処理を実現するための安易な処理の簡略化は精度の極端な低下に繋がりがねない。高精度な動作推定を行うにはシステム全体での計算資源の最適な配分が必要になると考えられる。本稿では特に人物検出処理について、与えられた計算量の制約に応じて処理方法の決定を行う計算量削減方法を提案する。

以降の章では、人物検出処理に関して、処理の流れと計算量の削減方法について述べる。

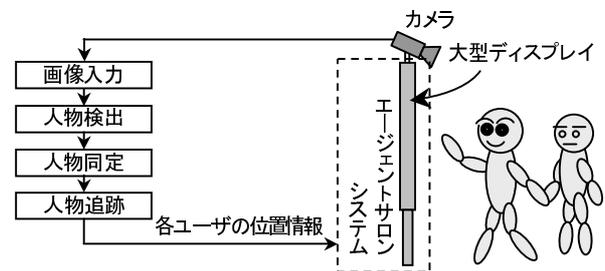


図 1：システム概念図

3. 人物検出

3.1 人物検出処理の流れ

ここでは以下のような人物検出処理を考える。

照明条件などの環境条件の変化に適応可能にするため、時系列画像を利用し動物体を抽出する[5]。具体的には、時系列画像において各画素に関する画素値分布を計算し、得られた各画素値分布を基に各画素について対象物体の存在する確率を示す確率画像を生成する。そして、その確率画像からテンプレートマッチングによって、人物位置を抽出する。テンプレートは人物とカメラの位置関係の変化に応じて数種類生成した。図 2 に処理の流れを示す。

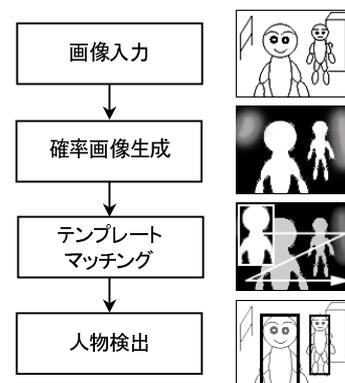


図 2：人物検出の流れ

3.2 計算量を制約条件とした処理の最適化方法

人物検出処理にはテンプレートマッチングを用いるが、テンプレートマッチングの処理手順を計算量に応じて最適化させることによって、最小限の精度低下で計算量を削減する。

本提案手法では、ベクトル量子化を利用し、コードブックの種類を変動させることによってテンプレートマッチングの計算量を調節する。具体的には、テンプレート画像を 2×2 画素や 4×4 画素などの小画像の集合として近似し、マッチングの前に入力画像の各部分に対してそれらの小画像との適合度を計算する。その結果をテンプレートマッチング時に用いることによって、計算量を削減することができる。

本手法のマッチングにおける計算量は小画像の種類数（以下、種類数）と、テンプレート画像を近似する際に用いた小画像の総数（以下、置換数）によって決まる。そのため、計算量に制約がある場合、その制約を満たす種類数・置換数を持つ処理手順のうち、最も検出精度が高くなる組を選ぶことによって、指定された計算量を満たし、また、検出精度の低下を最小限にできると考えた。本システムでは、計算量の制約を満たす処理手順を総当たりに調べ、元テンプレート画像と最も近似したものを作り出した処理手順を選び出す。そして、それらを使ってマッチングを行う。図 3 に本手法の流れを示す。

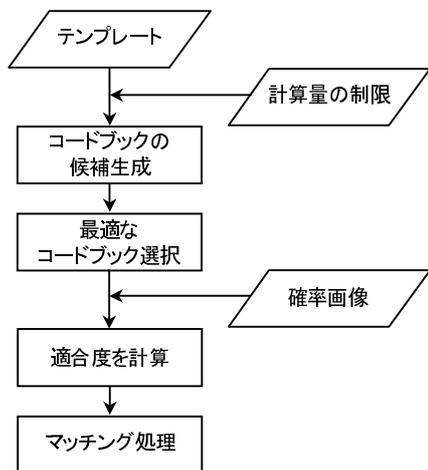


図 3：提案したテンプレートマッチングの流れ

3.3 実験結果

図 4 に計算量に応じて選択されたテンプレートを示す。ここで、計算量は PC (Pentium 933MHz, FreeBSD) 上での 1 フレームあたりの所要時間として与えた。入力画像は 160×120 画素、テンプレート画像は 32×80 画素である。尚、計算量の削減を行わない場合、1 フレームあたりの計算量は約 9 秒であった。また、同一の PC 上で、1 フレームの計算量を 0.3 秒として与えた場合の

人物抽出結果の例を図 5 に示す。

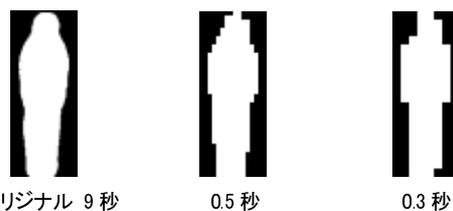


図 4：計算量に応じて選択されたテンプレート

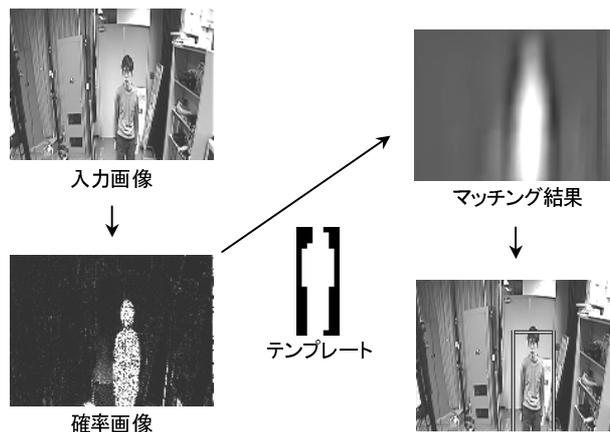


図 5：人物検出例

4. むすび

テンプレートマッチングの計算量を調節することによって、計算量の制約に応じた最適化が可能な人物検出手法を提案した。本手法により与えられた計算時間の中で最適な検出処理を選択することができるようになると思われる。今回は確率画像を対象としたが、同様の最適化は確率画像に限らず広く適用できる。今後は本手法の人物検出処理を人物追跡システムなどと統合し、システム全体としての実験を行うとともに、各処理への計算時間の最適な配分方法等について検討を行う予定である。

参考文献

- [1] 角 康之, 間瀬 健二. エージェントサロン: パーソナルエージェント同士のおしゃべりを利用した出会いと対話の促進. ソフトウェアエージェントとその応用 特集ワークショップ (SAA2000), 電子情報通信学会, 2000.
- [2] 内海章, 大谷淳. 非同期多視点画像による人物追跡手法の検討. 電子情報通信学会ソサイエティ大会予稿集, pp.D-12-52, 1998.
- [3] Q. Cai and J. K. Aggarwal. Tracking Human Motion Using Multiple Cameras. Proc. of 13th International Conference on Pattern Recognition, 68-72, 1996.
- [4] Ismail Haritaoglu and David Harwood and Larry S. Davis. W4: Who? When? Where? What? a real-time system for detecting and Tracking People. Proc. of the 3rd IEEE International Conference on Automatic Face and Gesture Recognition, pp.222-227, 1998.
- [5] 内海章, 大谷淳, 中津良平. 画素値分布の推定による時系列画像からの動物体領域抽出. 信学論, Vol.J81-D-II, No.8, 1768-1775, 1998.