

環境の変化に対し頑健性を有する
 5 J-2 実時間動画像並列認識システム

長谷川 修, 横澤 一彦, 石塚 満
 東京大学生産技術研究所

1. はじめに

本稿では、環境の変化に対して頑健性を有する実時間動画像並列認識アルゴリズムと、その試作システムについて述べる。ここで「環境」とは、照明環境及び認識対象となる周囲の環境の経時的変化とする。認識対象は、我々が構築を進めているヒューマン・インタフェースへの応用を目指し、通常室内環境下での複数の人物像とする。本アルゴリズムによれば、画像入力用CCDカメラ(1台)の方を向いている人物(必ずしもカメラの近くにいる人物ではない)を画像中より抽出可能であり、また室内照明の半分を消す等の環境の変化に対し、実時間に反応して処理を続けることができる。また本システムでは、ボトムアップ処理により対象人物の画像上の位置を確認すると、以後はトップダウン処理により、より効率的・高速に対象人物のトラッキング処理を行うことができる。本処理アルゴリズムは、生体の視覚系の処理方式より知見を得て、複数の画像特徴抽出モジュールからの実時間出力を統合する手法を基本とし、石塚研究室独自の並列画像認識・合成システム(TN-VIT)上で実現された。

ールには、それぞれ、①画像全体の明るさ、②色、③初期設定画像との差分、④動き、⑤エッジを用いる。また、各モジュールからの出力は、履歴(ヒストリ)として保存される(図1)。

2. 2 各モジュールからの出力 画像全体の明るさの処理には1台のVIT(後述)を用い、エッジの抽出に3台、その他の情報の抽出に4台を用いている各処理は、入力画像からx軸、y軸方向にそれぞれ10ピクセルおきにサンプリングした、51×51ピクセルサイズの画像に対して行う。処理時間は、エッジの抽出処理には約0.15秒を要し、その他のモジュールでは平均約0.1秒を要する。明るさの抽出モジュールからの出力は、画面全体の平均である。色情報の抽出モジュールでは、肌及び髪の色を持つデータを抽出している。初期設定画像との差分の抽出モジュールでは、処理開始時に入力し設定した画像との差分画像が出力される。また色情報及び初期設定画像との差分の抽出モジュールでは、抽出されたデータの密度を算出し、あるしきい値以上の密度を持つデータを、密度に応じた重みを付加して出力している。このように、膨大な画像データの中から意味のある情報のみを選択的に統合過程に伝達する。

動き情報の抽出モジュールでは、フレーム間差分による画面上各部の動きを8方向に抽出・出力している

2. 3 特徴統合過程 各特徴抽出モジュールからの出力は、それぞれの履歴部において重みの調整が行われる。この履歴情報を用いると、定常的に抽出され続ける認識以外の情報の重み(画像中の風にそよぐ窓辺の人物のポスター等)を小さくすることができる。

特徴情報の統合過程を図2に示す。特徴情報の統合処理は、まず履歴部を経た色情報と初期設定画像との差分情報の統合が行われる。統合は、各出力情報の重みのコンポリューションにより行う。この結果、処理の開始時に画面上に存在せず、かつ人間の頭の色情報を持つ部分のみが抽出され、その大まかな位置のデー

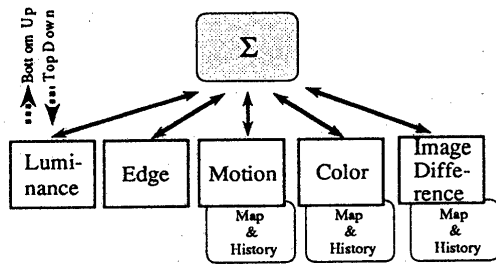


図1. 基本的アプローチ

2. 処理アルゴリズム

2. 1 基本的アプローチ 本システムにおける画像処理は、複数の画像特徴抽出モジュールからの実時間出力を統合する手法をとるが、画像特徴抽出モジュ

A Parallel Moving Image Recognition System Which is Robust to Environmental Conditions
 Osamu Hasegawa, Kazuhiko Yokosawa, Mitsuru Ishizuka
 Institute of the Industrial Science, University of Tokyo

タがマップに記入される。次に、このマップ上に動き情報が書き込まれる。動き情報は抽出速度が速く、頭が動いた先の位置の検出に用いている。

明るさの情報は、これらとは別に初期の段階で常に各モジュールにモニタされており、明るさに大きな変化が現れた際の各モジュールにおけるしきい値の再設定を行うのに用いる。

このように本システムでは、並列処理の利点を活かし、逐次的に処理を進める処理体系とは異なり、各特徴抽出モジュールに変化が検出されると実時間に処理に反映される特徴を有する。

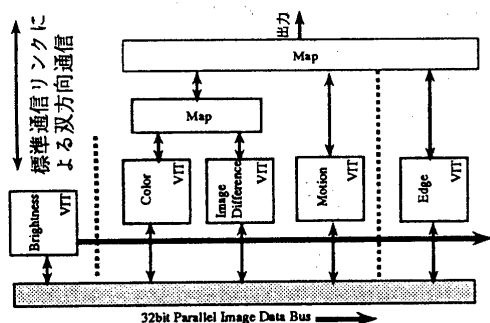


図2. 特徴情報の統合過程

2.4 トップダウン処理 トップダウン処理は、頭が存在すると推定された部分に対してウィンドウを設定し、その内部のエッジの抽出に基づいた、顔の向きの判定を行う処理として実行される。本トップダウン処理と特徴情報の抽出処理（ボトムアップ処理）は、並列に行われる。

3. VITとTN-VIT

3.1 VIT VIT (Visual Interface for Transputers)は、トランスポータのローカルメモリに直接データ入出力でき、ビデオレートで画像を転送できる、本研究室独自の32bit並列高速画像データバス付きトランスポータボードである。各VITは、T800トランスポータ (1.5MFLOPS, 10MIPS) を1台、2MByteのプログラムメモリ、1MByteのローカル画像メモリを有する。通信系では、標準の4本のシリアルリンクの他に、画像データ転送用の32-bitパラレルバスを有し、画像データの転送速度は、最大100Mbyte/sec (40nsec/pixel)である。また各VITは個々に独立して、カメラから画像データバスに入力された画像データの処理、さらにローカルメモリへの描画も可能である。

3.2 TN-VIT TN-VITとは、VITを中心とした並列トランスポータネットワークと画像の入出力のための周辺装置の総称である。さらに、通常の複数台のトランスポータと結合することにより、知識ベースに基づく推論等による知的機能を付与することが可能である。現在、VITが32台、標準のトランスポータ (T800) 16台が利用可能となっている。このうち本稿で述べる動画像の認識には、16台のVIT、及び16台の標準のトランスポータを用いた。

4. 実験及び結果

実験の結果、3人程度の人物が存在する画像中よりカメラの方向を向いている人物を抽出する処理が、約0.2秒で実行可能であった。ここでカメラの方向を向いている人物が複数存在する場合は、その中でカメラに最も近い人物とした。画像中に人物のポスターを混入した実験では、実際の人物像のみを容易に識別可能であり、処理速度等に影響を及ぼすことは無かった。照明条件を変化させた実験では、約0.2秒で各モジュールのしきい値の修正、トラッキングの続行が可能であった。また、ある環境で本システムを稼働させ続けた場合、履歴機能により処理速度が改善されることも確認できた。

5. まとめ

本稿において、照明環境の変化及び画像各部の経時的変化に対し頑健性を有する、新たな動画像認識手法を提案し、その試作システムについて述べた。本研究は、並列処理を意識した画像認識アルゴリズム研究の第1歩であり、今後さらに改良を進める予定である。

参考文献

- 1) O. Hasegawa, W. Wongwarawipat, C.W. Lee, M. Ishizuka: "Visual Software Agent built on Transputer Network with Visual Interface", Transputing '91, pp. 813-827, IOS Press, (1991)
- 2) W. Wongwarawipat, M. Ishizuka: A Visual Interface for Transputer Network (VIT) and its Application to Moving Image Analysis, Transputer/Occam Japan 3 (Proc. Third Transputer/Occam International Conference), IOS Press (1990)
- 3) O. Hasegawa, C.W. Lee, W. Wongwarawipat, M. Ishizuka: "A Real-time Visual Interactive System Between Finger Signs and Synthesized Human Facial Images Employing a Transputer-based Parallel Computer", Proc. CG Int'l '92, (1992)