8H−9

視覚的感性情報を考慮した人物動画像による ヒューマンインタフェースの試作

長谷川　修，　藤木　真和，　陳　履恒，　石塚　満
東京大学

1．はじめに

我々は次世代のヒューマン・インタフェースとして、自然な人間の姿を有して実時間で動作し、限定はされているが人間とのコミュニケーションが可能な「ビジュアル・ソフトウェアエージェント(Visual Software Agent : VSA)」を提案し、その構築を進めている。この人物像（エージェント）は、実際の人物のテクスチャを3次元ワイヤフレームモデル上にマッピングしたものであり、自然感が高い。またこのエージェントは、画像認識技術の利用により実時間でユーザを検出して反応し、視線を一致させることが可能である。本稿ではこれに続き、ユーザの感性を喚起するためのエージェントの「人間らしい反応・挙動」を構築と、そのインタフェースにおける利用形態について検討した。

2．並列コンピューティングシステム：ＶＩＴとＴＮ−ＶＩＴ

2．1　ＶＩＴ

ＶＩＴ(Visual Interface for Transputers)は、トランスピュ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ピュータボードである。各ＶＩＴは、T800トランスピュータ（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。

2．2　ＴＮ−ＶＩＴ

ＴＮ−ＶＩＴとは、ＶＩＴを中心とした並列トランスピュータネットワークと画像の入出力及び周辺装置の総称であり、現行のシステムでは画像認識と画像合成を同一のハードウエア上で並列に行うことができる。またネットワークの環境としては、ＶＩＴが32台、標準のトランスピュータ16台が利用可能となっている。これらの構成は、トランスピュータの標準のリンクを活用することによって任意に設定・拡張が可能であり、目的に応じた編成で利用することができる。その他の主要な構成装置は、パーソナルコンピュータ：2台，ＣＣＤカメラ：1台，合成動画像表示用モニタ：1台，テキスト・静止画像表示用モニタ：1台，マウス：1台，である。

3．エージェントの合成

本研究で用いたワイヤーフレームモデルは、1147個のバーテックス（頂点）よりなる516個のポリゴン（微小3角平面）から構成されている。人物モデルでは、人物の顔の各部分（左右の瞼・左右の瞳・上下唇及び顎）を1つまたは複数選択して反応させることができる。これにより、人物モデルに数種の表情をさせることができる。動画の合成には時間分割型パイプライン方式を用いた。

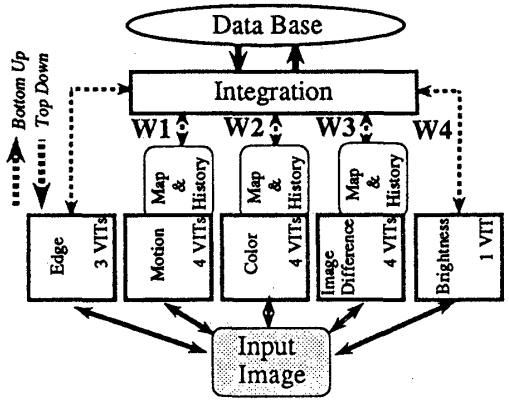

図1．特徴情報統合手法の基本的アプローチ

4．実時間並列画像認識手法

本動画像認識処理法は、基本的にボトムアップとトップダウンの2つの処理過程よりなり、これらの処理が並列的・協調的に動作する特徴を有する（図1）。ボトムアップ処理では、複数のモジュールにより画像特徴を抽出し、統合する。具体的には、①．画像全体の明るさ，②．動き情報，③．色情報，④．設定画像との差分，⑤．

A human interface system with moving human-like agent which considers emotional visual information
Osamu Hasegawa, Masakazu Fujiki, Chen Lieu-Hen and Mitsuru Ishizuka
University of Tokyo

エッジ情報，の各５つの情報抽出モジュールを設定している。また②③④のモジュールはマップとヒストリの機能を有している。トップダウン処理では、ボトムアップ処理で推定された認識対象の画像上の位置にウィンドウを設定し、データベースとの照合を行っている。ここで認識対象は、通常の室内環境下における複数の人物動画像とした。各モジュールにおける処理内容と統合過程の詳細については既に報告済みであるので、ここでは概説するにとどめる。

4．2　特徴情報の統合

①．色情報と設定画像との差分情報の統合　　ここでは設定画面上に存在せず、かつ人間の髪と肌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タが算出される。この際、頭全体に対する肌の色の割合を算出し、トップダウン処理をかける際の優先順位として利用している。

②．動き情報の統合　　次に動き情報が統合される。動き情報は抽出速度が速く、主として画像上の各人物の動きの計測（頭の位置のトラッキング）と、新たに画面中に現れた人物像の検出のために用いる。

③．明るさ情報の統合　　明るさの情報は、画面全体の明るさに大きな変化が現れた際にのみ、変化量を他のモジュールに通信するという形式で統合される。通信されたデータは、各モジュールにおけるしきい値の再設定に利用される。

④．エッジ情報の統合　　エッジの情報は設定されたウィンドウ内について抽出され、トップダウン処理過程においてデータベースとの照合時に利用される。

5．感性的なエージェントの挙動の検討：視線の利用

5．1　入力に対する反応　　一般に人間は「視線」の挙動に敏感であり、ここではこれを利用する。視線の合成は、主として各モジュール出力に対する重みWi（図１参照）を統合時に調節することにより行う。すなわち各重みを大きくすると、ユーザに対する反応を以下のように変化させることが可能となる。

W1 → 画像上の動きに対する反応が敏感になる。

W2 → 顔の色情報に強く反応するため、ユーザの顔を見つめ続ける。

W3 → 新たに画面に入ってきた対象に対し、敏感に反応する。

W4 → 大きな明るさの変化があった場合、瞬きする。

5．2　ユーザの注意の誘導　　これとは別に、ユーザの注意を他へ誘導することを目的とした視線の合成を行った。具体的には、ユーザに見せたいデータなどを表示している部分にエージェントの視線を向け、そちらにユーザの注意を向けさせることを試みた。

6．実験と結果

ＶＩＴを含むトランスピュータのプログラミングは、簡易ＯＳであるTDS2上でのOccam言語を用いた。

図２．出力画像例

実験の結果、画像入力に実時間で反応（画像出力まで約240msec，表示は毎秒約13ｺﾏ）しての多彩な視線の合成が可能となった。また定量的評価は困難ではあるが、エージェントの視線を用いたユーザの注意の誘導は効果があることが確認できた。

7．まとめ

今後は、画像認識と合成に関する研究を進める他、ＶＳＡへの音声機能や知能機能の付加、すなわち音声言語による対話機能や、知識ベースの結合によるエージェントのインテリジェント化などが課題である。

謝辞：本研究で用いたワイヤーフレームモデルは、東京大学工学部・原島（博）研究室から御提供頂いたものをベースに作成したことを記し、感謝致します。

参考文献

[1] W.Wongwarawipat and M.Ishizuka : "A Visual Interface for Transputer Network (VIT) and its Application to Moving Image Analysis", 3rd International OCCAM conference, pp.65-76,(1990)

[2] 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", Visual Computing, Springer-Verlag, pp.77-94, 1992

[3] 長谷川・横澤・藤木・石塚：「実時間動画像並列認識システムによる画像上人物モデルの視線移動の実現」，第８回ヒューマン・インタフェースシンポジウム講演論文集，pp.49-54, 1992