特 集 3
研究解説
UDC 681.3：621.397.3：165

新しいヒューマンインタフェースへ向けての並列コンピュータ（TN-VIT）上のビジュアル・ソフトウェアエージェント（VSA）

A Visual Software Agent (VSA) built on a Parallel Computer (TN-VIT) for Advanced Human Interfaces

石塚　満*・土肥　浩*・長谷川　修*・藤木真和*
Mitsuru ISHIZUKA, Hiroshi DOHI, Osamu HASEGAWA and Masakazu FUJIKI

並列コンピューティング，実時間画像認識，実時間画像生成等の情報技術を融合し，新世代のヒューマンインタフェースの形態を具体化するものとして研究を進めているビジュアル・ソフトウェアエージェント（VSA）について紹介する．現在までに，トランスピュータを要素プロセッサとする並列ビジュアル・コンピューティングシステムTN-VITと，外界の状況を認識して実時間で反応して動作する女性像のVSAと金魚版のVSAを作成している．

1. まえがき

本稿では，筆者等がこの4，5年に渡って進めている新しいヒューマンインタフェースの形態としてのビジュアル・ソフトウェアエージェント（VSA：Visual Software Agent）と，これを実現するハードウェアとして製作した並列ビジュアル・コンピューティングシステム（TN-VIT：Transputer Network with Visual Interface to Transputers）について紹介する．

近年，情報システムにおけるヒューマンインタフェースの重要性は増大しつつある．しかし，よいヒューマンインタフェースを実現するための基本的考え方，基本的指針は必ずしも明らかにされている訳ではない．

ヒューマンインタフェースを正しくとらえる困難さは，その総合的，複合的性格に起因する．第1に，十分に解明されていない存在である人間が一方の重要な対象になっていることである．第2に，関係する技術が多くのメディア技術，操作性の感覚（体感）に関する技術，知能機能の技術と非常に多岐にわたり，しかもこれらを総合的観点から融合する必要がある点である．

因難ではあるが，先端技術を真にわれわれのためのものにするためには現時点でヒューマンインタフェースの基本的考え方，目標，指針を明らかにしておくことは是非必要である．そのような目的で，筆者の1人（石塚）が主査を務めて郵政省の研究会で活動し，いくつかの新しい考え方の提示などの有意な成果を挙げている[1]．

要約してその成果の一端を記すと次のようになる．

1) ヒューマンインタフェースを構成するアーキテクチャの上位概念として「機能サイフル」を提示した．
2) 下位概念としてヒューマンインタフェースで考慮すべきことを「メディア」，「体感（英語では perception）」「知能」の3要素にまとめ，それらを統合する構成イメージを示した．
3) ヒューマンインタフェースの良さを表す総合評価尺度として「IA度（Degree of Intension Achievement ——意志達成容易度）」を提示し，それを算定するチェックリストを示した．
4) 今後目指すべき人間中心主義のヒューマンインタフェースを有する「メディアステーション」のモデルシステムを提示した．

いくぶん堅いまえがきが長くなったが，上記の研究会で提示した未来型ヒューマンインタフェースの具体例として，図1に示すようなTele-Museumがある．ここではメディアステーションの3次元仮想空間に美術館が実現されている．図1のシャガールの絵の横に立つのは知識を有し，言語を理解して話し，かつ感性的な顔，姿を有する合成されたガイド嬢である．「シャガールはどのような画家でしたか？」と問うと「ロシア生まれの画家で，絵の中にはロシアの農村の生活のイメージが現われています」などと答えてくれる．本稿で紹介するVSA（Visual Software Agent）は，このようなガイド嬢を具体的に実現することを目的とした研究であると理解してもらってよい．

さて，良いインタフェースはという問いに対し，きわめて簡単に重要な真理を述べると，

"The simple is the best"

となる．しかし情報システムのヒューマンインタフェースを考える場合，simpleだとsimpleなことしかできず，これだけでは不十分である．ボタン1つのインタフェースはsimpleでわかりやすいが，それだけではできることは3，4のタスクの範囲内となる．simpleでかつ多様なタスクを行える形態が必要なのである．

*東京大学生産技術研究所　第3部

図1　Tele-Museum とそこに現われるソフトウェアロボットとしてのガイド嬢のイメージ画

そこで simple というのを物理的 simplicity から拡張し，“人間にとって simple とは”と考える．そうすると，日常生活の動作は人間にとって特別に覚えたり意識したりしないという意味で simple である．この性質を利用したアイコン（Icon）インタフェース，より広義にはメタファ（Metaphor; 隠喩）によるインタフェースが現代のコンピュータのインタフェースの主流になっている．たとえば，デスクトップ・メタファといって机上の作業環境を図的に提示し，ゴミ箱図形へファイルを移すとファイルが消去されるといった具合である．

VSA は並列コンピュータ等のコンピューティング・パワーを活用し，このアイコンインタフェースを超え，顔，姿を有したアシスタント，秘書……をインタフェースの接面に登場させ，より人間の日常的コミュニケーションに近い形という意味で simple で円滑な次世代ヒューマンインタフェースを実現しようとする研究である．図2は Bright Star Technology という会社が描いたヒューマンインタフェースの進化の図だが，ここに述べたのと同様な思想を表わしている．

2. 研究の背景

繰り返しになる面もあるが，本研究の目的は次の二つの側面をもっている．

第1は，動画を含むビジュアルデータの実時間処理など，高次ヒューマンインタフェースの実現に要求される並列コンピュータの開発である．今後の情報処理システムの高速化に向けて並列コンピュータはきわめて重要な役割を果す．現在主流の逐次型コンピュータはノイマン型という基本的に単一のアーキテクチャにまとまっているが，並列コンピュータに関しては今後7，8年は単一の形態を決めることはできず，応用分野毎に複数の形態を追求することが重要であると考えられる．

並列処理の応用として画像処理・合成は処理量が多いことから重要な位置を占める．低レベル画像処理用の高速専用プロセッサはすでに存在するが，今後を見通して，中位，高位レベル処理まで含めて画像処理・生成に適用可能な高速並列（ビジュアル）コンピュータの形態を具体化して示したい．処理の高度化の背景には必ず強力なコンピューティングパワーを必要とする．画像のデータ構造に適合する2次元メッシュ状結合が容易で，他の汎用 CPU チップと異なり並列処理を前提として設計されたトランスピュータ（英国 Inmos 社製）を，われわれは，われわれの並列ビジュアルコンピュータの要素プロセッサに選択した．並列コンピュータの接続法としては，一時多くの並列処理形状に効率的にマッピング可能なことで，ハイパーキューブ結合が有望と考えられていたが，最近では並列処理の要求が高い画像データ処理や物理シミュレーションなどに適合するメッシュ状結合も盛り返してきている．

トランスピュータは4本の標準通信リンクを備えていて，並列接続が容易であるが，この通信リンクは速度が20Mb/s のシリアルリンクで，画像データの転送用にはきわめて不十分な速度である．そこで，トランスピュータのローカルメモリに直接入出力するビデオレートで動作する32ビット高速ビジュアル・データバスを開発し，この問題に対処することにした．このような機能実現に際し，並列処理言語 Occam やオペレーティングシステムなどトランスピュータのソフトウェア環境をそのまま利用できる設計とした．これによって，われわれの並列コンピュータは実用的にも十分使用でき，以下に記すような応用も進められる形態になっている．（大学で製作された並列コンピュータはハードウェアとしては動作したとしても，ソフトウェア環境を整備できないことから大部分利用されないで終っている．）

研究の第2の目的は，複数の情報技術を融合し新しい時代のヒューマンインタフェースを具体的に示すことである．筆者らは以前から人工知能の研究と共に画像処理・認識の研究を行ってきたが，7，8年前から画像処理は近い関係があるが別の発展の歴史をもつ画像生成（コンピュータグラフィックス）技術と合わせて，新しいビジュアルテクノロジーを形成していくのが時代的要請であると感じてきた．一方でまえがきで述べたように，人間と情報システムとの接面を形成するヒューマンインタフェースはますます重要になってきており，ここで画像の果す役割はきわめて大である．そして，動画像は静止画にない感性情報を含むが，これまでのコンピュータ

図2　コンピュータのヒューマンインターフェースの進化（Bright Star Technology 社による）

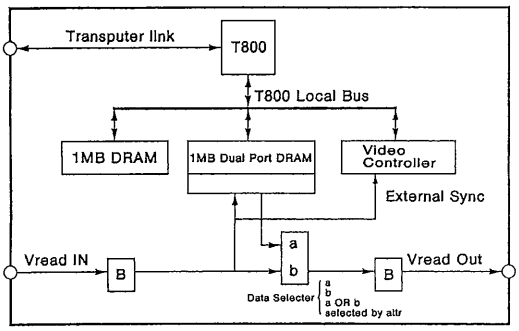


図3　VITのブロック図

では実時間処理はなかなか難しかった．

そこで，本研究で並列コンピュータから得られるコンピューティングパワーを活用し，動画像処理・合成の実時間処理技術を基盤とした新しいヒューマンインタフェースの形態としてVSA（Visual Software Agent）の実現という目標を設定し，その具体化を図る．VSAは音声機能や知識ベース機能を次第に付加していき，感性的でしかも知性あるエージェントに育てていこうとするものである．

3. 並列ビジュアル・コンピューティングシステム：TN-VIT

3.1 VIT（Visual Interface to Transputers）

並列コンピュータの開発に当り要素プロセッサとしてトランスピュータを選択した理由は，4本の標準通信リンクを具備していて並列接続が容易なこと，並列プログラミング言語Occamを基本言語としてサポートしていることであった．OccamはC.A.R. Hoareによる並列プログラミングの一つの重要な基礎的考え方であるCSP（Communicating Sequential Processes）に基づく並列処理言語である．トランスピュータはこの言語を実行するCPUチップとして英国Inmos社で開発された．われわれが使用し始めた1986年にはまだT-400であったが，1987年以降は32ビット（RISC形式）のT-800になっている．

トランスピュータの標準通信リンク（20Mb/s）では画像データ転送には不十分な速度であるため，われわれが開発したのがVITである．VITは図3のような構成であり，32ビット幅の並列データバス（VITバス）からVITのデュアルポートのローカルメモリへ直接アクセス（入出力）できる機能をもつ．並列データバスは片方向で，ビデオ画像の実時間処理・生成に適するようにビデオレート（40nsec/pixel）のクロックで動作している．したがって，約100MByte/secのデータ転送速度を実現していることになる．トランスピュータから見ると，この並列データバスへのアクセスを有するローカルメモリも，他のメモリと変りなく見え，データの書込み，読出しは自由である．図4は製作したVITボードであり，ダブルハイトのユーロカード規格となっている．

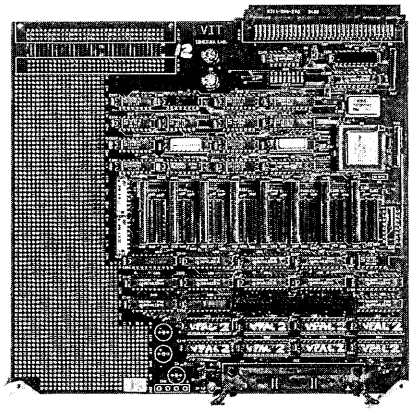


図4　製作したVITボード

要素プロセッサのT-800は32ビットCPUだし，VITバスも32ビット幅なので，メモリも1語32ビットで，処理は1語32ビット長を単位としているので画像の1画素をこれに対応させる．カラー画像データも扱うのでR，G，Bに各8ビットを割り当てても計24ビットで8ビット余る．そこでこの8ビットをアトリビュート（attribute）と称し，付加的データの記述に当てる．多くの場合，アトリビュートは3次元像の奥行き（depth）データ表現に当てている．

各VITは32ビットのビットバスに流れる画像データを独立してそのローカルメモリに画素位置を指定した画素単位で入力することができ，また逆にローカルメモリの画像データを単素単位でVITバスに出力することができる．この出力に際しては，ソフトウェアでスイッチ可能な次のような出力モードを備えている．

1) VITバスからのデータをそのまま出力
2) ローカルメモリのデータをVITバスに出力（VITバスのデータを書き換える）
3) VITバスのデータとローカルメモリのデータをビット単位でOR演算して出力
4) VITバスのデータとローカルメモリのデータのアトリビュートを比較し，小さい方のデータを出力

4）の出力モードによりVITバスはZバッファの役割を果すことができ，図5に例示するような画像の合成が実現できる．3）や4）の出力モードを利用すると，任意の領域の画像をVITバスに出力することができ，多彩な画像合成が可能となる．VITバスは片方向バスであるが，共有メモリ型通信の観点から見ると，メモリアクセスの競合がない共有メモリとも見なせる．すなわち，

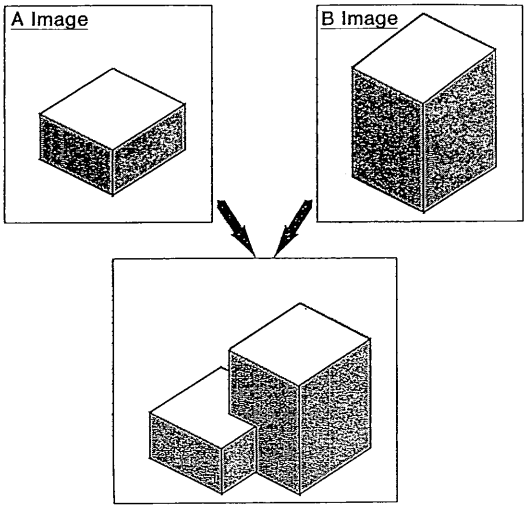


図5　奥行きを表すアトリビュートの比較によるVITバス上での画像の合成

各VITは通常の共有メモリで生ずるアクセス競合によるボトルネックを伴わずに，同一の画像データに高速でアクセスできる．以上のように，VITバスは単なる画像データの転送バスでなく，実時間の画像合成機能もサポートする機能バスとなっている．VITバスとローカルメモリとのインタフェース部は，図4の写真の下部に見られるPLA（Programmable Logic Array）で実現されている．

VITバスは以上のように各VITでバッファリングを行いデータ演算を行うバスであるが，各VITでの遅延はビデオレートで1画素の時間，すなわちインタレース・モード時で33nsec，ノンインタレース・モード時で66nsecと非常に小さい．これによってVITをVITバスで接続したシステムは，任意個のVITを接続可能となり，scalable（拡張性を有する）システムとなる．たとえ512のVITを接続しても，合計の遅延は約60μsecである．

3.2　TN-VIT（Transputer Network with VIT）

TN-VITはVITボードをVITバスで接続し，他の必要な機能を付加した実時間画像処理・画像合成に適する並列ビジュアル・コンピューティングシステムである．図6はその標準的な構成を示している．

TVカメラからの信号をA/D変換した画像データがVITバスへの標準的入力となる．画像のフィルタリングやエッジ抽出といった低レベル画像処理は，色々と高速専用プロセッサが開発，市販されているので，これらはVITで処理するよりも効率が良い．そこで図6に示すように，VITバスの先頭に専用画像プロセッサであるTAICHI-CORE（イーゼル社製）を挿入している．VITバスの最終的出力はA/D変換を通してカラーTVモニタに表示される．

画像データ以外のデータ交換はトランスピュータの標準通信リンクを介して行うようになっており，VITバスへのアクセスを有しない通常のトランスピュータ・モジュールも接続できる．図6のホストコンピュータはプログラム開発やプログラムの各プロセッサへのロード，ファイル管理等を行うPC-9801であり，OSにはトンスピュータ用の簡易OSであるTDSを使用している．

現在のTN-VITの規模はVIT 32台，それ以外の通常トランスピュータ16台で，合計48台からなる並列コンピュータとなっている．図7にTN-VITの外観を示す．

並列処理としては，以下のような方法をプログラミン

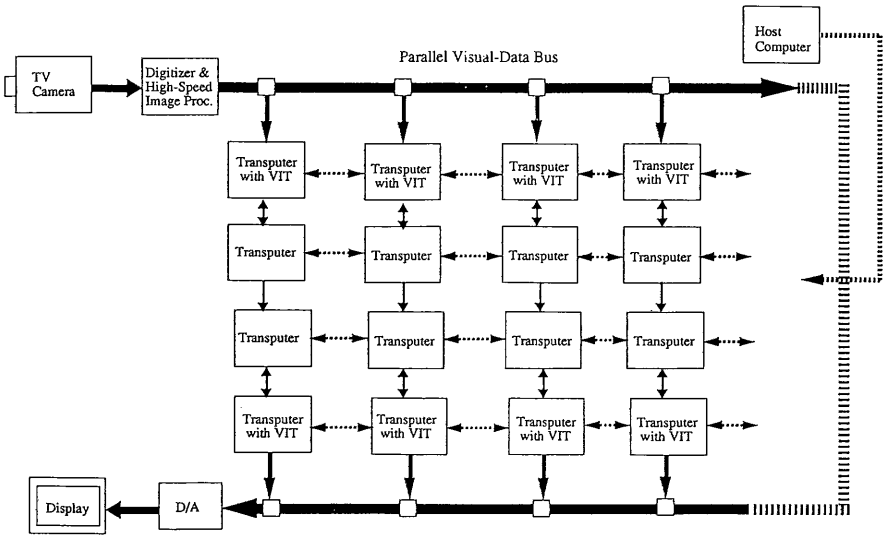


図6　並列ビジュアル・コンピューティングシステムTN-VITの標準構成

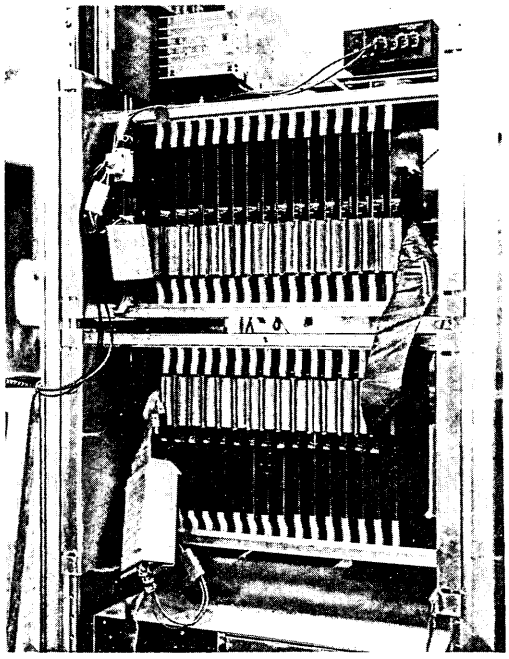


図7　並列ビジュアル・コンピューティングシステム TN-VIT の外観

グにより選択して使用できる．

a）平面分割処理（各プロセッサが画像の異なる領域を処理・生成する．）

b）対象分割処理（各プロセッサが異なる対象物を処理・生成する．）

c）時分割処理（各プロセッサが異なる時刻の画像を処理・生成する．切り換えて連続して出力すると動画となる．）

高速画像処理を実現するために高速並列データバスを備えた並列トランスピュータ・システムとしては，他に新日鉄製の製品などがある．新日鉄の製品は DSP（Digital Signal Processor）による高速画像処理を可能とするために，3本の32ビット並列データバスを備えた構成となっており，われわれの TN-VIT の方がシンプルで，時間画像処理と実時間画像合成の両者に適合する構成であるといえる．

4. 自然感の高い人物モデルとその動画像の実時間合成[3)～6)]

4.1　3次元表面モデルとテクスチャマッピング

VSA（Visual Software Agent）はディスプレイ上で動作する擬人化エージェントであり，これを情報システムとユーザの接面とするヒューマンインタフェースを実現する．通常のグラフィックスによる人工的な合成像と異なり，VSA では高次のビジュアル技術を導入し，3次元モデル上に実際の人物像から得たテクチャ（表面パターン）を張り付けるテクチャマッピングにより，高い自然感を実現する．これは最近の知的画像符号化で導入された技術と同じである．

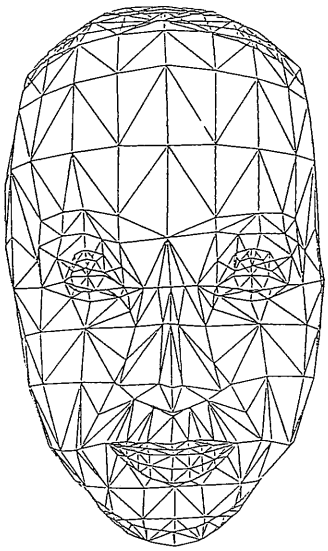


図8　顔の3次元表面モデル

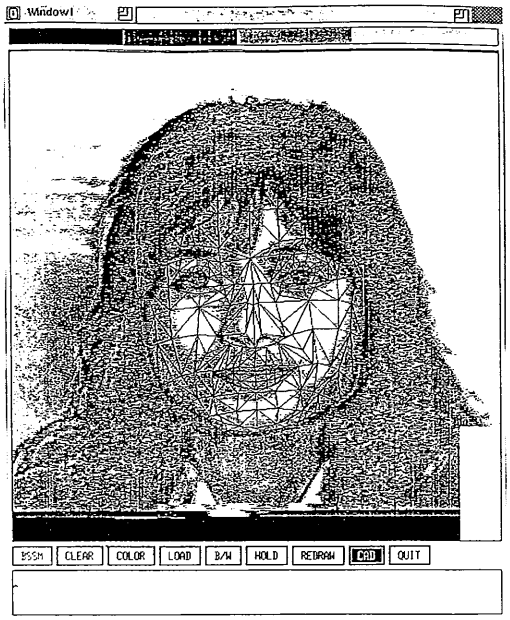


図9　3次元表面モデルの実画像へのフィッティングによるテクチャ・データの取り込み

われわれの人物顔形状の基礎3次元表面モデルは図8のものであり，約600の頂点，約500の3角形パッチから構成されている．これは東大工学部電子工学科の原島博教授からいただいたデータを基にして作成した．髪の長い部分は別途のモデルを用いている．

3次元表面モデルを図9に示すように，TVカメラから得た実際の正面顔画像にフィティングし（高さ，幅，傾き，さらには目と口の位置を合わせる），各3角形パッチ毎に対応するテクスチャ・データを取り込んで管理する．合成像の生成時には，3次元モデルを移動，回転させ，各3角形パッチ毎に対応するテクスチャをマッピングする．このテクスチャマッピングによる生成像は非常に自然感は高いが，通常のグラフィックスによるレンダリングと比べると計算コストが高く，時間がかかることになる．

4.2　TN-VITによる実時間動画像合成

応答速度は良いヒューマンインタフェースの重要な要件の一つである．VSAは計算コストの高いテクチャマッピングによる人物像の動画を，TN-VITの並列コンピューティングパワーを利用して，ほぼ実時間で生成する．

画像合成の並列処理モードとしては，いくつかの方法を試みたが，現在使用しているのは，時分割型の処理である．すなわち，異なるVITに異なる時刻の画像を描画させ，出力を順次切り換えて動画を生成する方法である．n台の＃1～＃nのVITを1つのグループとすると，これを仮想的にリング状の構成として，表示時には＃iのVITが所定のインターバルの間表示を行った後，次の＃(i＋1)のVIT(＃nの次は＃1)に表示を行わせるようにする．表示を行っていない間のVITは次の時刻の描画を行う．したがって，1枚の描画に要する時間をT_dとすると，1つのVITの表示時間は$T_d/(1+n)$以上に設定する必要があり，動画としての単位時間（毎秒）当りの表示枚数は$(1-n)/T_d$以下となる．顔画像の場合は頂点座標の計算，テクスチャマッピングを含めて29台のVITに描画させると，毎秒約25枚の動画生成速度を達成できる．

ホストコンピュータあるいは他の動き制御プロセッサは，各描画VITに動き制御情報を伝達するのは煩雑であるので，＃1～＃nのVITの代表となる＃1のVITだけと交信し，動き制御情報を送る．＃1のVITは顔の大局的な位置，回転と，左目，右目，口の後述する代表点の動きを，残りの＃2～＃nのVITに伝える．各VITはこの情報と該当の表示時刻に基づき，顔の各頂点座標を計算し，テクスチャマッピングして画像を生成する．

目口の動きは表情を形成する上で非常に重要な要素であるが，それらの個々の頂点の動きを伝えるのは複雑である．そこで図10で黒丸で示しているような代表頂点のみを動かすと，他の関連する頂点も移動するモデラとしている．図10の例では，このようにして上まぶたを閉じる動きと，眼球の移動が実現されている．

図11(a)(b)(c)は，それぞれ正面像，まぶたを閉じ口を若干開いた像，手前にうなずいた合成像を示している．

5. 外界の認識による反応[7),8)]

親しみのあるヒューマンインタフェースとしてのVSAは，自律的な外界とのインタラクションが必要であり，そのためには外界の実時間認識能力が必要となる．

最初に簡単な機能として，外界の動き領域を検出し，そちらに視線を向ける機能を作成した．動き領域は，

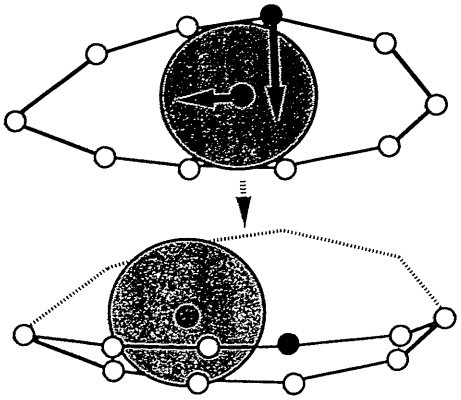


図10　代表頂点(黒丸)の移動による他の関連する頂点の移動

(a)　(b)　(c)

図11　女性像VSAの動作例(実際はカラー)

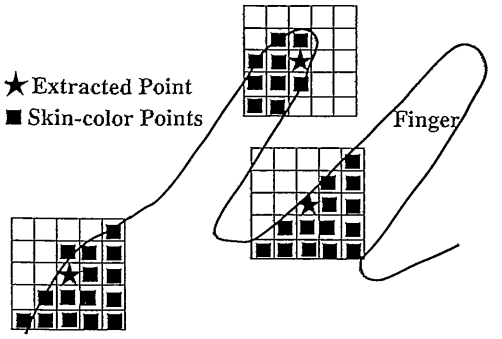
VIT バスの先頭に設置された専用画像プロセッサがフレーム間差分信号を出すようにし，これを処理することで実時間で検出できる．表示モニタの前に位置するユーザは動きを伴うので，この機能により VSA とユーザとの近似的な視線一致を実現できる．視線一致によりインタラクション時の親しみ易さは増大する．

実現されている幾分複雑な他の外界認識機能は，指サイン認識とこれによるコマンド伝達である．手や指によるコマンド伝達の手段としては，最近ではデータグローブが人工現実感（Virtual Reality または Artificial Reality）システム等の入力系によく使われるようになった．データグローブは違和感のある器具を手に装着しなければならない点が，最も大きな問題点である．ここでの指サイン認識は非装着で簡単なコマンド伝達を実現しようとするものである．これまでにあった指先の認識によるポインティングよりは，いくぶん進んだ機能を有しており，TN-VIT の並列処理を利用して高速認識を達成している．

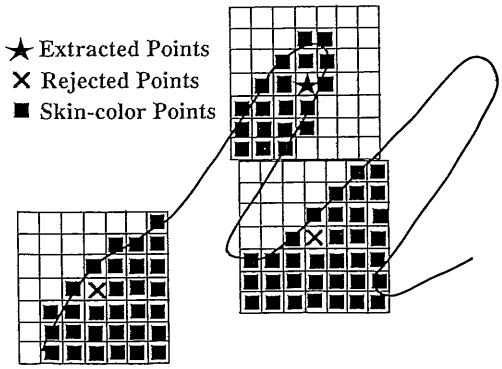
われわれの指サイン認識は，立てられた指の本数とその位置を認識する．右手，左手，人指し指，中指等の区別はない．TV カラーカメラからの入力を行うので，指サインによるコマンド入力はカメラに対して直角に近い角度に掌を位置させる必要がある．掌の表裏位置，回転は認識に関係しない．

指サインの認識法は以下の通りである．

1) カメラから手を含むカラー画像を入力
2) 細部の変動の影響と雑音を抑圧するために，平滑化する．
3) データ量を減らし処理速度を向上させるため，10×10画素領域を1画素にした圧縮画像を作る．
4) 圧縮画像から手領域の候補として肌色領域を抽出する．
5) 5×5画素のウィンドウを走査して，中心が肌色でかつウィンドウ中に10～16画素の肌色データが存在する部位を見つける．これは図12(a)に示すような肌色領域の境界部を抽出することに相当する．（10～16画素といった値は実験に基づいて設定したものであり，以下も同様である．）
6) 上記で抽出された部位に関し，ウィンドウ・サイズを7×7画素に拡大したとき，肌色データが21～35画素となる部位を図12(b)のように抽出する．これは指先き部分ではウィンドウ・サイズを25から49に拡大しても肌色データの増加率は少ないことを利用し，指先き部のみの抽出を図っている．
7) 上記6）で抽出された点を，8近傍隣接している場合には1点に併合する．これにより1点が1つの指先きに対応することになる．



(a) 5×5サイズのフィルタリング



(b) 7×7サイズのフィルタリング

図12　指先きの抽出法の説明図

肌色といったように色データによる識別，5×5，7×7ウィンドウによるフィルタリングといった簡単な処理の組み合せにより，高速に指先きの位置と数を検出することができる特徴をもつ．複数の指先きがある場合には，それらの平均位置を算出してコマンドのポインティング位置とする．

この指サイン認識の認識率を実験により確認した．掌の平をカメラにほぼ直交して提示し，半袖シャツ着用で100回，長袖シャツ着用で100回の認識実験を行った．その結果，半袖シャツ着用時には指先き数の正しい認識率は1～5の本数に関し平均すると95％以上であるが，長袖シャツ着用時には約75％に低下した．さらに長袖シャツが肌色に近い色を含む場合には，約45％程に低下した．認識時間は1台の VIT で約0.5秒である．認識率は必ずしも十分に高い訳ではないが，認識結果が反応となり相手に即座にフィードバックされ，誤認識の場合には容易に再入力することが可能な応用に対しては，簡単なコマンド入力法として有効である．

VSA の応用では2台の VIT に0.25秒間隔でこの指サイン認識を行わせることにより，毎秒4回の指サイン・コマンドの認識を行わせることにした．認識結果はトラ

図13　女性像VSAとのインタラクションの様子

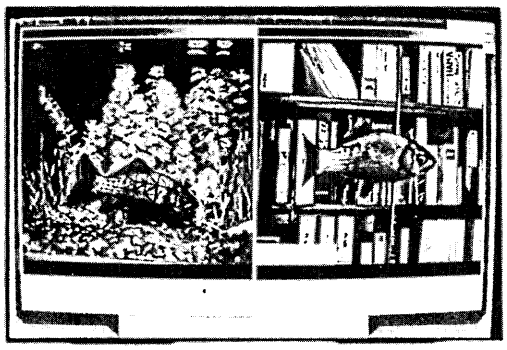


図14　金魚版VSA

ンスピュータの標準通信リンクを介して描画用VITに伝達するようにする．

デモンストレーションでは，女性像VSAの動きと以下のような対応付けを行い，指サインとのインタラクションを実現している．

- にぎりこぶし ―― 正面位置に静止
- 指1本 ―― 指の方向を向く
- 指2本 ―― 指の方向を向き右眼を閉じる
- 指3本 ―― 指の方向を向き両眼を閉じる
- 指4本 ―― 指の方向を向き口を開く
- 指5本 ―― 指の方向を向き両眼を閉じ，口を開く

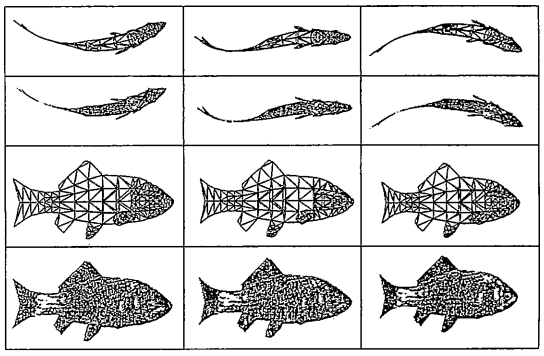


図15　BSSMを用いた金魚モデルの柔軟な変形例

指サインの認識時間が約0.5秒で，1台のVITの描出時間が約1.1秒なので，現状では指サインから動きの反応までに若干の遅れがあり，今後改善を考えている．

図13は女性像のVSAとモニタの前の人間とのインタラクションのようすを示す写真である．TVカメラはモニタ上に置かれており，VSAの眼の役割を果たしている．

外界の認識機能としては，指サイン認識の他に，複数の人の中でVSAに顔を向けている人を認識する方法等の開発も行っている．

6. 金魚版のVSA[9)~11)]

ヒューマンインタフェースの接面としてのVSAとしては，これまでに紹介してきたような擬人化エージェントとしてのVSAが最も重要である．しかし，人間像VSAは，能力が不足している点があると，そこが非常に目立つことになり，技術の完成度が高くないと実用には不十分で，実用までにはいくぶんかの年月を要する．

そこで，われわれは女性像VSAの開発と並行して，それ程高い知的能力を要求されず，より早期に実用化が可能と考えられる金魚版のVSAも開発している．熱帯魚等の存在は自然の美しさ，心の安らぎを与えるものであるが，餌を与えたり，水を換えたりしなければならず，楽ではない．われわれの金魚版VSAは，水槽，あるいはそれ以外の任意の空間に生息し，外界の変化に反応して感性的な動きをする仮想金魚像である．高い自然感を出すために，人間像VSAと同様に，3次元表面モデラ上に自然の金魚から得たテクスチャをマッピングして生成する．

図14の左画面は3次元モデルを実際の金魚にフィティングさせ，テクスチャを取り入れているところであり，右画面は本棚という現実にはあり得ない仮想空間に再生させたところを示している．

金魚の場合，自然に近くするためには体の曲げやひれの動きを出す必要があるが，個々の頂点を制御してこれらの動きを生成するのは大変である．われわれはこのような問題に対処するために，BSSM（Bone-structured Solid Modeler）[9)]という変形可能（ディフォーマブル）な物体表現のためのモデラを開発した．金魚の場合にはまず通常の3次元表面モデルを作り，次に動き制御のために体の中心付近に節点（ジョイント）を有する骨（ボーン）を挿入する．3次元モデルの各頂点を近くの節点に関連付け，骨を動かすとこれにつれて各頂点が動くようにしたモデラである．基本は頂点，パッチから成る3次元表面モデルなので，この上に元のモデルと変り

図16　金魚版VSAとのインタラクションの様子

なくテクスチャマッピングすることができる．図15はこのBSSMによる金魚モデルの柔軟な変形の例を示している．

外界の認識機能としては，動領域の実時間検出や，5節に述べたのと同じ指サインの認識機能を結合している．そして，動領域や指サインの位置に向かって，指サインの場合は指定の金魚が指定された動作で動くデモンストレーション・システムを実現している．図16はこのような金魚版VSAによるインタラクティブ・システムを示す写真である．

7. む　す　び

並列コンピュータ，実時間動画像認識，実時間動画像生成等の情報技術を融合し，新世代のヒューマンインタフェースとしてのVSA（Visual Software Agent）の構築を図っている研究について紹介した．

VSAについては今後，音声・言語コミュニケーション能力の付与，会社の受付等の対応に必要な知識ベースの結合，感性的な動作，表情の実現を図り，感性的で知的なVSAに育てていきたいと考えている．

謝　　辞

本研究の初期において元大学院博士課程学生のW. Wongwarawipat氏（現：㈱イーゼル），李七雨氏（現：イメージ情報科学研究所）の貢献が大であったことを記し，感謝いたします．また本研究はカシオ科学振興財団，村田学術振興財団，㈱イーゼル，郵政省電気通信フロンティア研究，文部省科研費重点領域研究（感性情報処理）から助成を受けたことを記し，感謝いたします．

（1992年9月1日受理）

参 考 文 献

1) ヒューマンインタフェース――人間中心のメディアステーションに向けて――（郵政省電気通信局電気通信システム課監修）：㈶日本データ通信協会，（1990.6）

2) W. Wongwarawipat, M. Ishizuka: A Visual Interface for Transputer Network (VIT) and its Application to Moving Image Analysis, Proc. 3rd Int'l Transputer/Occam Canf., IOS Press, pp. 65–76 (1990)

3) W. Wongwarawipat, C-W. Lee, O. Hasegawa, H. Dohi, M. Ishizuka: Visual Software Agent built on Transputer Network with Visual Interface, Trasputing '91, Sunnyvale, Ca. (P. Welch et al. eds.), IOS Press (1991)

4) M. Ishizuka, O. Hasegawa, W. Wongwarawipat, C-W. Lee, H. Dohi: Visual Software Agent (VSA) built on Transputer Network with Visual Interface (TN-VIT), Proc. Computer World '91, Osaka (1991)

5) O. Hasegawa, W. Wongwarawipat, C-W. Lee, M. Ishizuka: Real-time Moving Human Face Synthesis using a Parallel Computer Network, Proc. IECON '91, Kobe (1991)

6) 長谷川，李，W. ウォンワラウィパット，石塚：VSA（Visual Software Agent）のための入力画像に反応する実時間人物表情の合成，計測自動制御学会第7回ヒューマンインタフェース・シンポジウム論文集，pp. 279–286 (1991. 10)

7) 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 (T. L. Kunii (ed.)), pp. 77–94, Springer-Verlag (1992)

8) O. Hasegawa, C-W. Lee, W. Wongwarawipat, M. Ishizuka: Realtime Synthesis of Human-like Agent in response to User's Moving Image, Proc. 11th Int'l Conf. on Pattern Recognition, Vol. IV, Hague (1992)

9) C-W. Lee, O. Hasegawa, W. Wongwarawipat, H. Dohi, M. Ishizuka: Realistic Image Synthesis of Deformable Living Thing based on Motion Understanding, Jour. of Visual Communication and Image Representation (Academic Press), Vol. 2, No. 4, pp. 345–354 (1991)

10) 李，長谷川，土肥，石塚：動き認識に基づくディフォーマブルオブジェクトの画像合成，計測自動制御学会第7回ヒューマンインターフェース・シンポジウム論文集，pp. 279–286 (1991. 10)

11) M. Fujiki, O. Hasegawa, C-W. Lee, W. Wongwarawipat, M. Ishizuka: A Prototype of Goldfish Software Robot with Real-time Response Function by a Parallel Computer, Proc. IEEE Int'l Workshop on Robot and Human Communication (RO-MAN '92), Tokyo (1992)