博士学位申請論文

実時間並列動画像認識・合成による
ヒューマンインタフェースの研究

指導教官

石塚 満 教授

東京大学大学院

工学系研究科電子工学専攻

07095

長谷川　修

目　次

第１章　序　論

１．１　概要

人間が相対してコミュニケーションする場合、互いに目を見、言葉に抑揚をつけ、時には相づちや身振り手振りを加える。このような視線の一致(Eye Contact)や挙動(Gesture)は、言外の情報を大量に含み、円滑な意志疎通のために極めて重要な役割を果たしていると考えられる。すなわち、人同士のコミュニケーションにおいて互いに伝達しているのは、単なるテキストデータとしての言語ではなく、感性を含む「情意」そのものであると考えられる。

これに対し、抑揚がなく無機質で非人間的なコミュニケーションを「機械的な」会話と表現する場合がある。これは従来のコンピュータを含む機械が繰り返し作業の遂行を中心とするために用いられる比喩であるが、現在のマン－マシン・インタフェースの未熟さを端的に表しているとも考えられる。つまり、近年の工学全般に渡る研究・開発の進展は目ざましいが、

それらは主として機械中心の発展の歴史であり、それらを使うべき人間（ユーザ）を中心とした情報科学の立場からの研究が立ち遅れていることを示している。言い替えれば、人間が機械に合わせており、機械とのコミュニケーションにおける無機質性にいらだちを感じつつも、それが機械の持つ特性の一部であるかの如くにあきらめと同時に受け入れているのが現状である。ボタンが多数付いた高機能の家電製品を持っているが一度も使ったことのないボタンがある、という経験は誰にでもあろうが、これはこのことを如実に物語っている。すなわち高度に技術が発展した今日、理想的なインタフェースのあるべき姿を模索し、人と機械との円滑なコミュニケーションを具体的に実現することが強く求められている。

現在このような現状認識に基づき、各機関で個別的な研究・技術開発が進められている。例えば文部省科学研究費総合研究（Ａ）として、「感性情報の抽出・検索・表現に関する総合的研究」が大阪大学・辻三郎教授を中心に平成元年度より実施されている[1,2,3]。ここでは人間の感性的側面の情報科学の立場からの探求を研究目標とし、従来の知識科学に対して「感性科学」の基礎を確立することを目指している。具体的には、画像・図形・音響などのパターン情報を研究対象領域とし、研究分担者には大学・企業の日本における第一線の研究者２１名が名を連ねている。

またこれに先立ち、東京大学・石塚満教授が主査を務め1988年から1990年3月まで活動した郵政省のヒューマン・インタフェース研究会では、情報通信機器の今後の良いヒューマン・インタフェースへ向けての理念、基本的な考え方がまとめられた[4,5]。ここでは上位概念としての「機能サイクル」、このサイクルを円滑に回すために重要な３つの技術要素としての

「メディア」、「体感」、「知能」機能、ヒューマンインタフェースの良さの評価尺度としての「ＩＡ度（degree of intension achievement）」など、いくつかの有用な概念が提示された。そして今後数年はメディア技術が表立った役割を果たすが、次第にその上に知能、人工現実感的な体感を有する技術が重要になると論じられている。

これらの他にも、ＡＴＲ，ＮＴＴ，ＮＥＣ，松下，富士通，ＳＯＮＹなどの大手企業は独自の研究体制を擁して活発に研究を進めており、学会などを通じて成果が報告されつつある[6,7]。

一方で、飛躍的な進歩を続けているコンピュータの能力を、このような人間と機械との良い接面の実現のために活用することは、重要な観点である。特に、人間のコミュニケーションにおいて重要な役割を果たす画像情報は、データ量が膨大であり、従来その取扱いが困難であったが、近年実時間に近い高速での取扱いが可能となりつつある。中でも並列処理は各種画像処理アルゴリズムの効率的実装に適し、並列コンピュータの利用と並列アルゴリズムを含む技術の確立が求められている[29]。

本研究ではこのような背景に基づき、人と機械との円滑なコミュニケーションのためのヒューマン・インタフェースの具体的事例を、並列コンピュータ上に構築することを目標とした。

１．２　本研究の特徴

本研究ではヒューマン・インタフェースの具体的事例の構築を目標としたと述べた。そこでまず構築のための実際的なモデルが必要となるが、このモデルとして本研究では自然な人間の姿を有して実時間で動作し、限定

はされているが知能機能を有して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)」を提案する。

具体的には、コンピュータの画面上に人物動画像を合成し、人間と実時間でインタラクトさせ、人間と機械との間に人間同士のコミュニケーションをシミュレートする。この人物像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、レンダリングにより合成された画像に対し、より自然感・現実感の高い画像となっている。またこの仮想的「人物」は、ユーザを検出して視線を一致させ、眼球や頭部は人間に近い動きで動作する。このような人間的な動作は人間（ユーザ）の感性を刺激し、従来のインタフェースと比較してより親しみやすい印象を与え、コンピュータの存在を感じさせず、あたかも人と対話するかの如くに機械とコミュニケートすることが可能となる。

ＶＳＡの重要な要素であるユーザからの情報の入出力には、主として画像の認識及び合成の各技術を用いた。実時間での画像認識及び合成には極めて大きな負荷がかかるが、この問題の解決のために並列コンピュータシステムを利用した。並列コンピュータシステムの要素プロセッサとしては、4本の通信リンクを具備したトランスピュータ（英国Inmos社製）を採用した[31]。しかし、トランスピュータの通信リンク速度は動画像データの転送には不十分であることから、トランスピュータに32bit高速画像データバスを付けた石塚研究室独自のトランスピュータボード：ＶＩＴ(Visual Interface for Transputers, 1990年度博士修了のW.Wongwarawipat氏設計)[32]を32台製作し、これに標準のトランスピュータを16台加えて独自の並列コンピューティングシステム：ＴＮ－ＶＩＴを構築した。このような画

像の認識と合成の実時間処理を中心とするビジュアル・ヒューマン・インタフェースは、必要とされるハードウェア環境の整備が困難なことから従来例が少なく、本研究の大きな特徴の一つとなっている。

1．3　本論文の構成

1章以後の本論文の内容を以下に示す。

2章において、従来のインタフェース研究の概要について簡単に触れ、本研究の背景を述べる。

3章では、本研究において提案するビジュアル・ソフトウェアエージェント(Visual Software Agent : VSA)の概念と構想について述べる。

4章では、本研究で用いた並列コンピュータシステム、ＶＩＴ及びＴＮ－ＶＩＴのハードウェア及びソフトウェアの概要について述べる。

5章では、システムへのデータ・コマンド入力部の処理について述べる。特に、並列動画像認識処理アルゴリズムを中心に述べ、キーボード、マウスなどの利用形態についても触れる。

6章では、システムからの出力部の処理として、動画像合成過程やその他の周辺機器からの出力形態について述べる。

7章では、ＶＳＡにおける感性情報処理の検討について述べる。

8章では、ＶＳＡのプロトタイプシステムと、その性能評価のための実験結果について述べる。

9章では、本論文の結論を述べる。

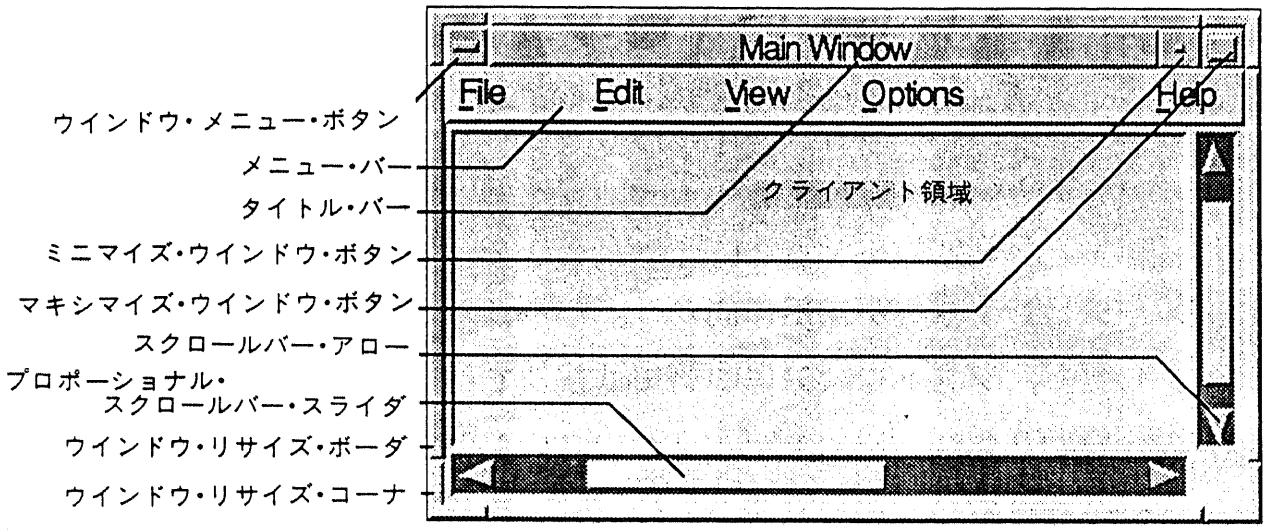
第２章　研究の背景

２．１　従来のインタフェースの研究

２．１．１　ＧＵＩからＭＭＩへ

本節では、コンピュータと人間をつなぐインタフェースの歴史を簡単に振り返る。

1960年代の初期のインタフェースは、プラグボード、パンチ・カードとライン・プリンタであった。この頃のコンピュータは専門家のものであり、多くの一般人には遠い存在であった。1970年代に入ると、パンチ・カードとライン・プリンタなどがテキスト情報（英数字）主体のインタフェースへと変わり、1980年代にはXerox PARC 社 Alto においてはじめてＧＵＩ(graphical user interface)が採用された。ＧＵＩはその後の十数年間において、インタフェースの主流となった（図１）。ＧＵＩにおいては、主としてトラックボールやマウスを使用し、静止画中心の出力が行われ、デスクトップ・メタフォアが利用される。このＧＵＩの登場により、インタ



グラフィカル・ユーザ・インタフェースを構成する要素　OSF（オープン・ソフトウエア・ファウンデーション）の Motif は，図のような構成のウインドウを利用する。図は Addison-Wesley Publishers Inc. の提供。

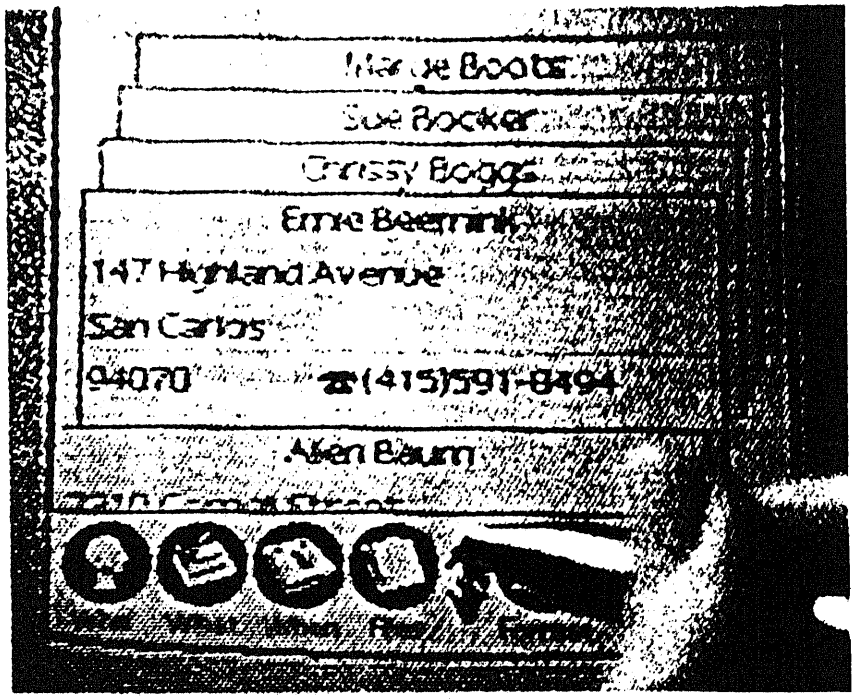
図１．ＧＵＩの例[8]

フェース環境はそれまでのコンピュータに比べ格段に進歩したが、それらにしてもコンピュータ中心のインタフェースであり、マウスやウインドウもコンピュータの都合から提案された概念であった。すなわち、ユーザはマウスの操作やウィンドウ、アイコンの概念などの「特別な技能」を学習しなければ利用できなかった。

1990年代に入るとＭＭＩ(multimodal interface)が提案され（図２）、コンピュータ・ハードウエアの飛躍的な進歩にも助けられ、主流となった。

マルチモーダルとは、コンピュータの入出力方式が多様になることを指す。ＭＭＩにおいては、手書き（ペン）入力や，ユーザ視線の検出，音声認識・合成，動画像の出力などを行うとしている。すなわち、文字や図形を書く、ものを見る、音を聞く、会話をする、といった人間の通常の生活でのコミュニケーションの形態をそのままコンピュータと人間との間に実現することにより、人間にとってより自然なインタフェース環境を提供することを目的としている。ＭＭＩは、「人間中心の」ユーザ・インタフェースの実現に向けての具体的な取り組みの第一歩とも言える。

例えば、マッキントッシュ社・W.W.Gaver氏の作成した"Sonic Finder"では、「音」を利用することにより従来のマックのアイコンに工夫を加え、新たなデスクトップ環境を提供している。この音が加えられたアイコンは、オーディトリ・アイコンと呼ばれている。ユーザがアイコンをマウスでクリックすると、ファイル容量に合わせて高低が設定された音がする。またクリックして引きずると、ものを引きずる音がする。ファイルのコピーでは、コップに水を注ぐような音がする。コピーが進むにつれて音のピッチが高くなり、ピッチが上がりきったところでコピーが終了する。この例で



音とアニメーションでメタフォアを強化する　米アップル社の Newton は，アニメーションと音を連動させて，メタフォアの現実感を高めている。

図２．ＭＭＩの例[9]

は音を使い、従来のＧＵＩをより親しみやすいものへと強化していると言える。また1991年から1992年にかけて米アップル社のQuick Time，米マイクロソフト社のマルチメディアWindpwsという２つのマルチメディアＯＳが相次いで発表されたが、これらはＧＵＩからＭＭＩへの移行に拍車をかけた。

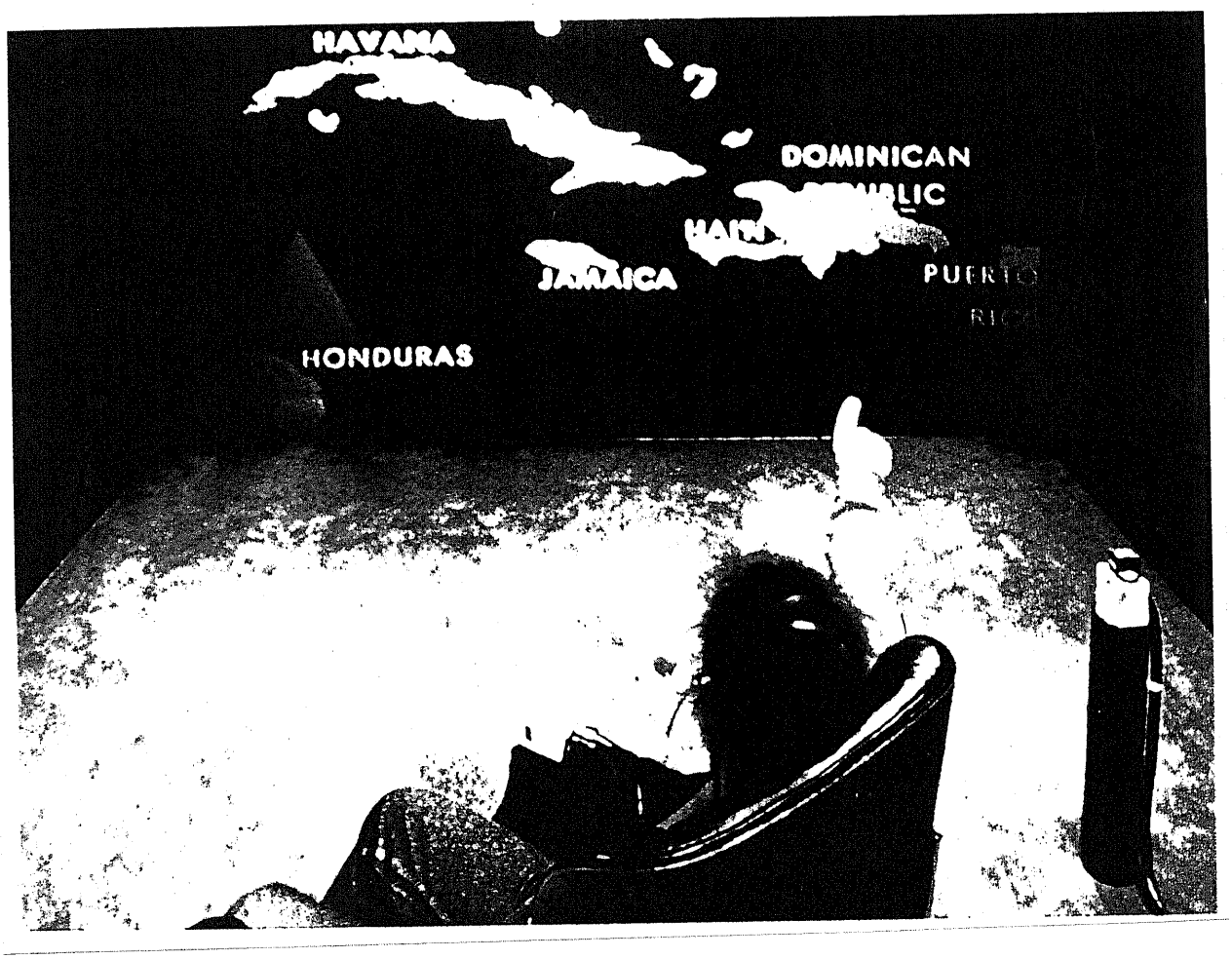
現在は、アニメーションや動画、音、ペン入力などの最も効率的な組み合わせは何か、また組み合わせるタイミングはどうするか、といったことが主要な研究課題となっている[8,9,10]。やみくもな組み合わせは、かえってユーザに対する負担となる危険があるからである。

２．１．２　ＭＩＴメディア・ラボ

これらの大きな流れとは別に、ＭＩＴのメディア・ラボでは、早くから多くの意欲的・挑戦的なインタフェースの研究が行われた[11]。

中でも手に磁気センサを付け、また音声入力を加えて、画面を手で直接指示することによりモニタ上の画像のエディットを行う"Put-That-There"システム（図３）は広く知られているが、これはデータグローブ（図４）が発表される前の1980年に発表されている。

また遠距離にいる人物の顔をモニタ上に立体的にローカルに合成し、テレビ電話のごとく利用する"Talking Heads"は、1981年に発表されている（図５）。"Talking Heads"では、立体的で顔の形をした半透明のスクリーン上にフル・カラー画像を背面投影して人物像を合成する。画像はローカルなビデオ・ディスクから取り出し、音声は電話回線経由で届く。また電話回線からの信号には、遠くの話し手の頭部に取り付けた方位・位置セン



それをそこに置け。音声と指示によって船をカリブ海のあたりに配備する　ここで操作されるアイテムは，艦船や小艦隊を表わす円や菱形で，カリブ海図を背景に移動される。利用者は軽量のヘッド・マイクロフォンを身につけている。2つの空間感知立方体の小さい方(利用者の袖に隠されている）が利用者の腕——二重露光で撮影されている——にくくりつけられている。利用者の右側の台座には，空間感知立方体の大きい方を載せたルーサイトのブロックがある（写真著作権：1980, Association for Computing Machinery, Inc. 許可を得て使用)。

図３．Put-That-Thereシステム[39]

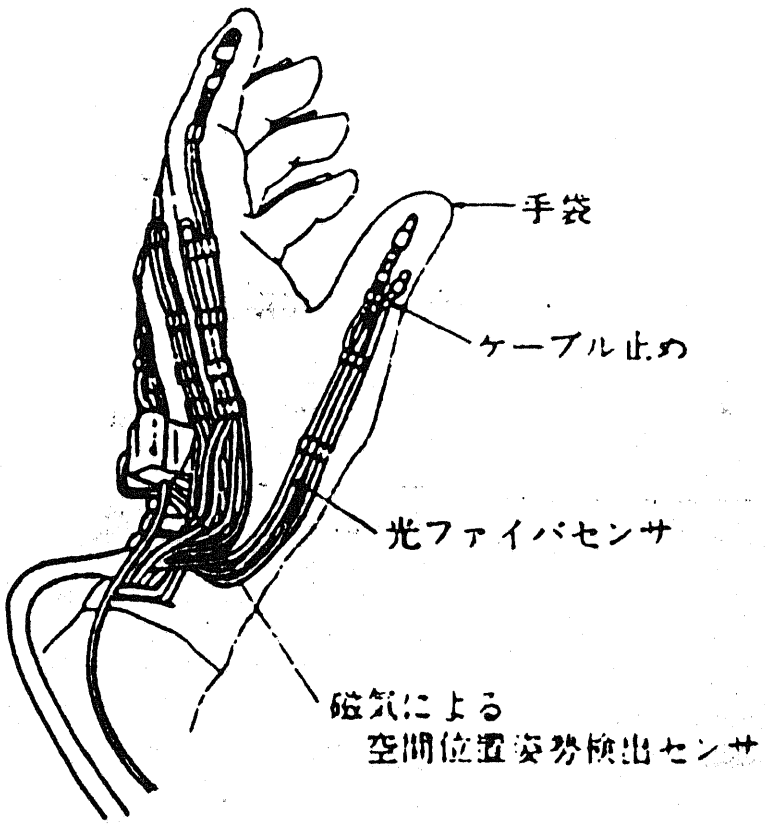


図４．データグローブ

顔の形をした半透明のスクリーンに，顔のフル・カラー画像を背面投影して，いきいきとした「存在」感をつくり出す。テレビ会議では，テレビ画像はローカルなビデオ・ディスクから取り出し，音声は電話回線経由で届く。遠方の話し手の音声信号は，頭に取り付けた空間感知立方体からの頭部位置情報と混合される。顔のスクリーンは遠方の話し手の頭部の動きと完全に連動して，振ったり，うなずいたりする。唇の動きは，説得力のある印象をつくり出すために，送られてくる音声に合わせてローカルに生成される。目はローカルに自然で任意な動きをつけてもよいし，原則的には遠隔眼球追跡によって動かすことも可能である。

図５．Talking Heads [12]

サの情報も含まれており、話し手の頭部の動きと連動して頭を振ったり、頷いたりする。口は音声信号と同期させて開閉させ、眼は適当に動かす。

これらの研究は発表された当時、あまりに独創的・先駆的な研究であり、また当時のコンピュータのパワー不足もあって周囲の理解が必ずしも十分に得られなかったが、今日のデータグローブや顔を用いた各種システムの発展を見るにつけ、当時の研究・開発者達の先見性の高さに驚かざるを得ない。

他にもメディア・ラボでは、早くから、眼球追跡，注視点を中心とした画像の漸進的送信システム，部屋そのものが情報端末となるメディア・ルーム，データランドなど、数々の興味深い研究が行われている。

2．1．3　合成顔画像を用いたインタフェース

合成顔画像を利用したインタフェースの研究は、前節で触れた"Talking Heads"などが初期の研究例と考えられるが、他の主なものに1980年代後半から東京大学・原島教授らのグループによって始められた研究がある[13, 14]。

同グループでは画像の知的符号化を主目的とし、遠距離の人物の挙動や表情をモニタ上にローカルに合成している。この合成画像は、ワイヤフレームモデル（図６）上に実際の人物の画像をテクスチャマッピングしたものである。豊かな表情の合成には、"Facial Action Coding System:FACS"[15]を用いて行っており、自然感が高い。

原島教授らの研究に続き、成蹊大学の森島助教授らのグループは、豊かな表情を有する顔の動画像をモニタ上に合成し、音声入力に反応して動作させたり、メイルの読み上げなどをさせている。また同グループでは、表情エディタを作成しており、表情の変化量や変化速度等の設定を容易に変

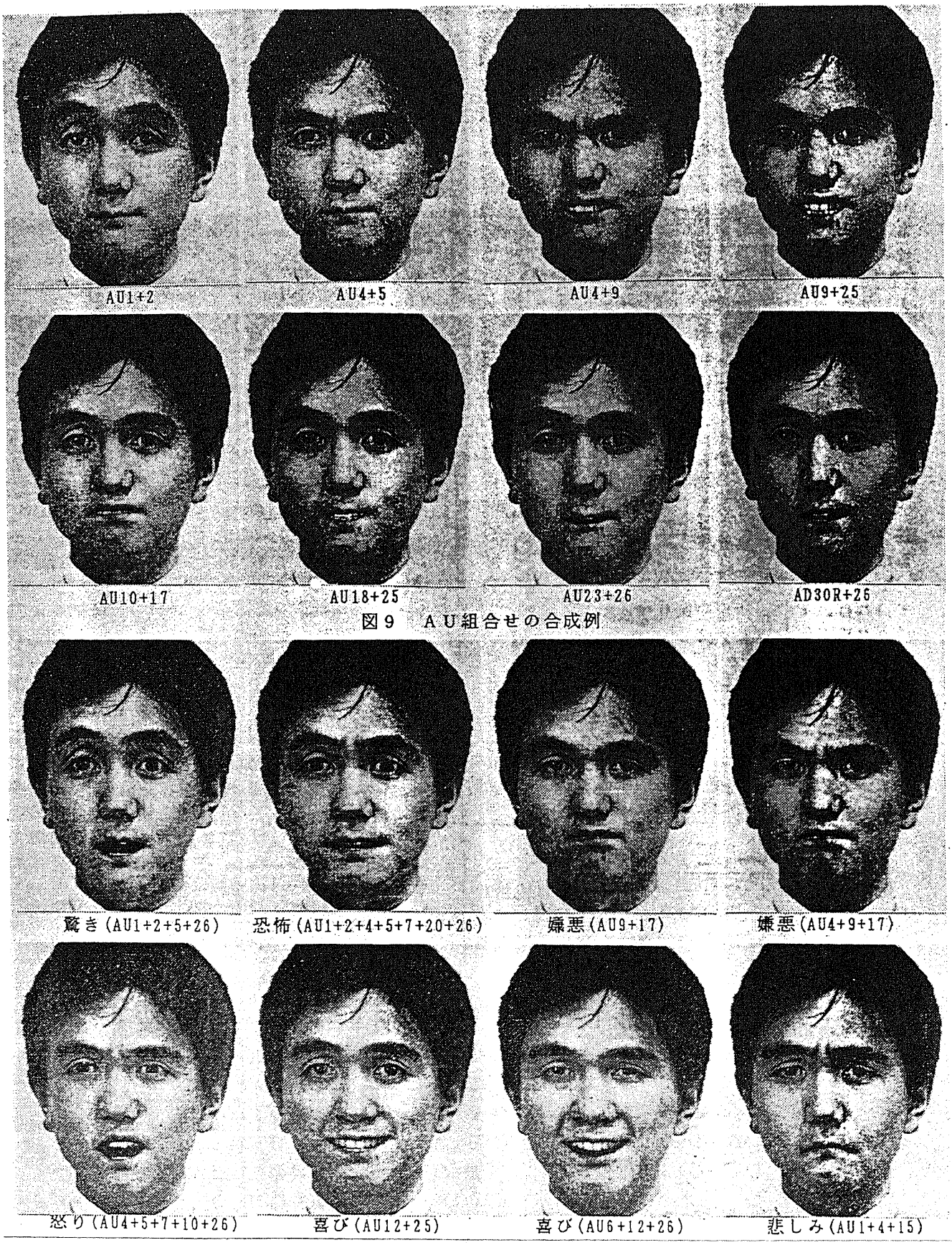


図9　AU組合せの合成例

図6．原島教授らのによる表情合成例

更可能としている。この他、最近では合成顔画像の品質の向上を目指し、テクスチャマップを施すモデルの精度を向上させたり、テクスチャマップの顔画像とＣＧの毛髪との合成の試みを報告するなど、活発に研究を進めている。

SONYコンピュータサイエンス研究所では、男性の顔画像が画面上に登場し、ユーザと対話するコンピュータを開発・発表した。この顔画像は１６本の”筋肉”を持ち、２６通りの表情を作ることが可能である。対話は音声入・出力技術を用いて行っており、現在３４の名詞と８つの動詞、４つの形容詞、その他２２を認識でき、SONY製品の紹介を行うデモが可能である。

これらの他にも、合成顔画像を利用したインタフェースの研究は早稲田大学やKDDなど、多くの研究機関において進められている。「顔」は情報の伝達などにおいて、単にデータやメッセージを伝えるという以上の活用形態があると考えられるが、現状では顔画像の利用による心理的効果などにおいて未知の部分が多い。多くの「顔」の研究者達は顔の未知の領域に、新たな可能性や神秘的とも思える魅力を感じているのかも知れない。

直接インタフェースにかかわる研究ではないが、近年「顔」や「人間」像の合成に関する研究事例は多く、枚挙に暇がない。一例のみ紹介すると、N.Thalman教授（スイス）[21]の人物像の合成の研究は有名である。教授らは早くからワイヤフレームモデルをレンダリングした精密な人物像を合成し、人物の歩行動作や衣類が風になびく様子の動力学的にも厳密なシミュレーションを行っている。

２．２　２章のまとめ

２章においては、従来のパンチ・カード，ライン・プリンタにはじまり、ＧＵＩ，ＭＭＩへと続くインタフェースの研究動向について概観した。またインタフェースの一つの究極の姿として期待される、画像上に合成顔画像（エージェントと呼ばれる）が登場するインタフェースに研究に関する報告例についても紹介した。

インタフェースの研究の進展の歴史は、常に進歩するハードウェアと共にあったと考えられる。すなわち機械（特にコンピュータ）が高機能・高性能化するにつれ、それらを平易で使いやすくするための高度なインタフェースへのニーズは高まるが、同時にその構築のためには、高機能で高性能なハードウェアが不可欠であることが理解できる。

ハードウェアの劇的な進歩を身近なパソコンの世界で見てみると、1982年当時のNECのパソコンに登載されていたＣＰＵと現在のそれとを比較すると、演算処理能力は実に180倍になっている。また最近富士通が試作に成功した２５０メガビットＤＲＡＭは、百円玉ほどのサイズのこのチップ１つにカラー静止画像が80-100枚記憶可能である。これらに見られるようなハードウェアの進歩は、現在Work Stationレベルの計算機を用い進められている多様で高度なインタフェースの研究の成果を、近い将来必ず身近に登場させるであろう。

近年米アップルコンピュータ社から発表され、ＭＭＩ的要素を取り入れた携帯型情報機器 Newton はその先駆的存在と言える。Newtonは手帳サイズでありながら、ペン入力が可能であるばかりでなく、画面に表示したカードを繰るときには音を出し、文書を捨てるときには紙をクシャクシャに

してごみ箱に入れるアニメーションが出るという。

合成顔画像を利用したインタフェースは、動画の合成機能のみをとっても大きな処理能力を必要とし、従来は十分な処理スピードが得られなかったが、ハードウェアの高速化・大容量化により実時間処理が可能となりつつある。ジャケットの内ポケットに入るようなポケコン（ポケット　コンピュータ）のディスプレイ上でも、合成顔動画像が登場するインタフェースが利用できる日は、そう遠くないと考えられる。

ところでＭＭＩの構築においてマルチメディアは必要不可欠であるが、逆にマルチメディアを効率的に使うためにも、よいインタフェースの確立が求められているという点は興味深い。

ＭＭＩでは多彩なデータの入・出力形態[17-19]を利用するが、逆にＭＭＩは多彩なデータをユーザに効率よく利用させるための手法と見ることも可能である。現在、音とアニメーションについては、使いやすいインタフェースの構築のために有効であることが確認されているが、動画については未確認である。いち早くＤＳＰを登載したアメリカ・ＮｅＸＴコンピュータ社においても、1992年 5月の段階で「マルチメディアの効果的な利用法とユーザ・インタフェースの在り方については検討中」とのことである[9]。いずれにせよインタフェースとマルチメディアは、相互依存、相互補完関係にあるといえよう。

第３章　新しいビジュアルヒューマン・インタフェースとしての Visual Software Agent (VSA)

３．１　ＶＳＡの概要

本節では、画像認識・画像の視覚化・知識ベース・並列処理技術を統合したヒューマンインタフェースとしての、ビジュアル・ソフトウエアエージェント（Visual Software Agent, 以下ＶＳＡ）の基本的概念について述べる。

人間と機械とのコミュニケーションを考える際の、情報処理のレベルと情報通信量の関係を図７に示す。図中左側の３角形は人間を示し、右側は機械である。ここでは、人間にとり快適なコミュニケーション（インタフェース）環境は左側の三角形最上部の情意レベルのコミュニケーションであるが、その実現のために必要とされる機械側の情報処理量は膨大であり、また極めて高度な情報処理技術が要求されることが示されている。このような膨大かつ高度な情報処理は現状では困難であるが、ＶＳＡでは段階を

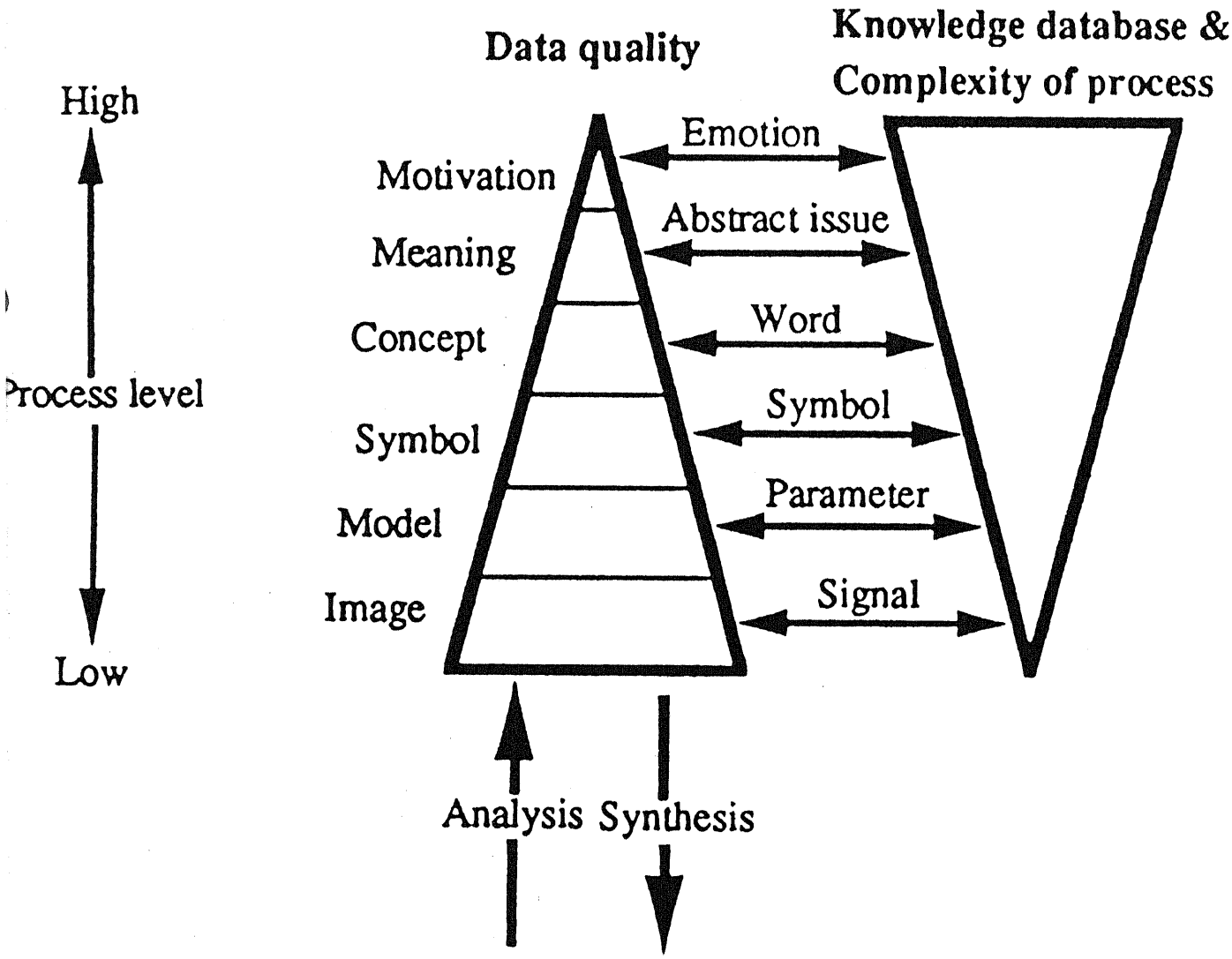


図７．マン・マシンコンミュニケーション

踏み、画像情報処理技術を中心として徐々にこれを実現することを目指している。

情報の視覚化は、人間に対する情報の伝達には極めて有効な手段である。ここで言う情報の視覚化には、単に静止画像や文書だけではなく、３次元動画像も含んでいる。ＶＳＡでは、一種のソフトウエアロボット（３．２節参照）とも言える、リアリティを持った実時間３次元動画像（人物像）をコンピュータとユーザとの接面とする。このようなアプローチは、必要とされるコンピュータ能力等の点から、これまでのヒューマンインタフェースには見られず、ＶＳＡの大きな特色である。この他にもＶＳＡには、以下に示すような特色を持たせる。

（１）高速な応答

ディジタル時代における情報システムの品質を考える際に、応答速度は非常に重要なファクタである。人間とコンピュータの間の良いインタフェースには、ビジュアル情報も含めた十分な応答速度、遅滞感のない迅速なコミュニケーションが不可欠である。ＶＳＡでは、並列トランスピュータ及びＶＩＴ（４章参照）を用いて高速な処理と応答を実現する。

（２）自然に近いコミュニケーション，操作感

ＶＳＡでは実際の人物画像を基に、自然に近い動画像人物像を合成してユーザとコンピュータの接面とし、自然感のあるコミュニケーションの実現を目指す。ここで人物動画像としては、人物の上半身像を用いる。また知識ベースとの結合やＡＩ技術の利用により、円滑なコミュニケーションに必要な「知能」を計算機内の「人物」に与える。この「人物」との「対話」には、ユーザの挙動等をＣＣＤカメラから画像入力する他に、音声入

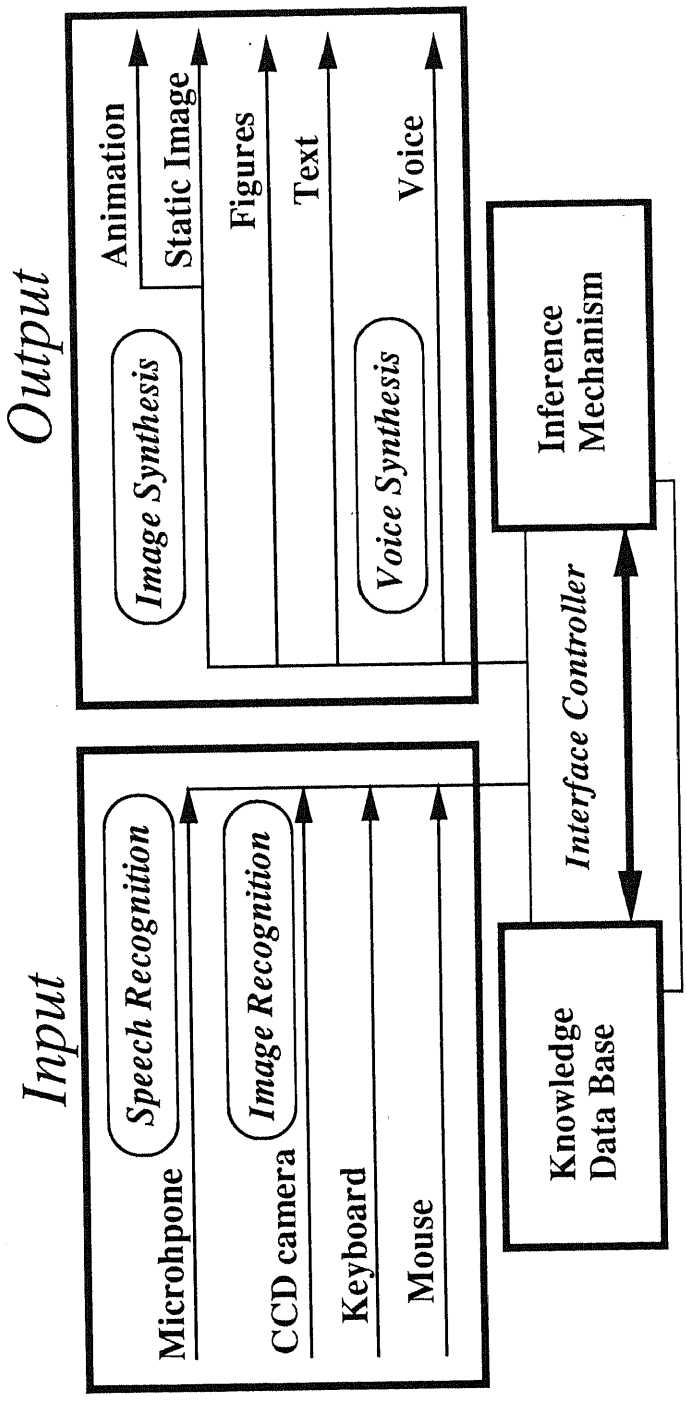


図８．ＶＳＡのシステム構想

出力も行う。ただし音声認識・合成には既成の技術を利用し、市販の装置を組み込む。図８にこのように多くの技術を統合するＶＳＡシステムの将来的なシステム構想を示す。（図８中で、本論文で言及するのは主として画像入出力関連であり、音声は対象としていない。）

（３）豊富な情報の提供方法・手段

通常複雑な処理を行えば様々なデータが生成されるが、それらをユーザに効率的・効果的に提供するために、システムには豊富な情報の提供方法・手段を持つことが求められる。また同一のデータに対し、異なったアプローチをとることも考えられ、システムには状況に応じた柔軟な対応が求められる。ＶＳＡではユーザへの情報の提供において、将来的には３次元人物動画像の他に、文書・静止画像・動画像等を目的や状況に応じて選択可能とする。

３．２　ソフトウェア・ロボット（Software Robot : SR）

本節では本研究におけるソフトウェアロボットの定義について述べる。ソフトウェアロボットは、英語ではSoft Roboticsと呼ばれ、１つ研究分野として徐々に認知されつつある[20,21]。

従来コンピュータグラフィックスによりロボットの動作のシミュレーション等が行われてきたが、近年のコンピュータハードウェアの進歩は、リアルな画像の実時間合成を可能としつつある。このような状況を背景として、合成画像を単なるシミュレーションの道具としてではなく、画像上の仮想のロボットとして積極的に利用しようとする動きが出てきた。これが

ソフトウェアロボット研究の発端であると考えられる。

現在ソフトウェアロボット（あるいはその概念）は、ヒューマンインタフェースをはじめ、人工現実感、アミューズメントなどの分野で幅広く用いられるようになってきている[22]。

しかし現状ではソフトウェアロボットに対する定義は研究者により異なるため、本研究におけるソフトウェアロボットを以下のように定義する。

－ ソフトウェアロボットの定義 －

［Ⅰ］ソフトウェアロボットは3次元データを有する画像上の仮想のロボットとする。

［Ⅱ］ソフトウェアロボットは実時間で動画像として合成される。

［Ⅲ］ソフトウェアロボット外部からの入力に実時間で反応する。

ＶＳＡでは、人間型ソフトウェアロボットをコンピュータと人間の間をつなぐエージェントとして用いた。この他にも本研究では金魚型ソフトウェアロボットも試作したが、これについては6章の6．2節において触れる。

富士通研究所におけるソフトウェアロボットの試作例を図9に示す[23, 24]。この例では、仮想生物と名付けられたソフトウェアロボット群が人間（ユーザ）とインタラクトする。具体的には、8ｍ四方、高さ4ｍと想定されたドーム内に3匹のクラゲ（仮想生物）と1匹の魚が泳ぎ回っており、ユーザはＶＰＬ社のアイフォンを呼ばれる眼鏡をかけ、この世界に入り込む。ユーザの挙動はデータグローブを使ってシステムに伝えられ、コマンドに変換されて仮想生物達とのインタラクションに利用される。クラゲ達

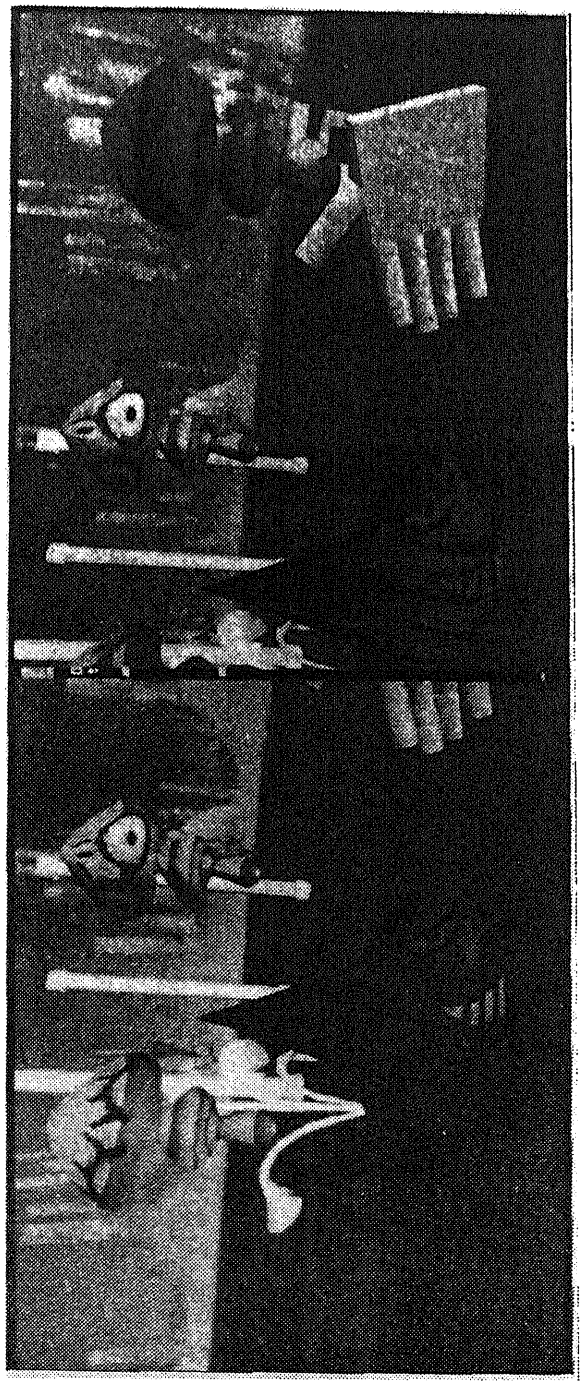


写真1　体験者の見る映像（左：左目用、右：右目用）

図9．富士通が開発した人工生物とのインタラクション・システム

の動きは行動シミュレーションによって自律的に制御されている。

本システムはアミューズメント的色彩が強いが、ソフトウェアロボットの具体的構築例として興味深い。

3．3　3章のまとめ

本章では、本研究の根幹をなすＶＳＡの概念とシステム構想について述べ、ソフトウェアロボットの概念と定義についても述べた。また、他の研究機関によるソフトウェアロボット構築の例として、富士通研究所の仮想生物を紹介した。

本研究で提案するＶＳＡは、本研究のみで完結するものではない。ＶＳＡは画像の認識と合成以外にも多方面から研究される必要があり、音声認識・合成技術や、知識情報処理技術などをはじめとする各種関連技術との統合が進められる必要がある。

ソフトウェアロボットは、近年のハードウェアの進展に伴って現れてきた概念であるが、アミューズメントの分野を始めとして既に身近になりつつある。ＳＲは機械的実体を伴わないために、ものを運ぶなどの物理的な作業は実行できないが、人の感覚や感性に訴えかけることが可能であり、今後感性工学などの進展と共に活用の場が広がるものと考えられる。

第４章　並列ビジュアル・コンピューティングシステム：ＴＮ－ＶＩＴ

４．１　ＶＩＴ（Visual Interface to Transputers）

ＶＩＴに関する詳細については、W.Wongwarawipat氏の1990年度，東京大学博士論文[32]を参照のこと。ここでは概略のみを述べる。

４．１．１　並列画像システム概観

従来並列画像処理システムは多く開発されてきたが、それらは主として処理の高速化を目的として設計されている[27,28]。中でも、画像の空間分割による並列処理を用いた高速化手法は効率が良く、一般に広く用いられている。

一例を挙げれば、新日本製鉄の「Fire Pip」システム[26]は、鋳片プリントにおける製品検査のための画像処理を行うが、画像を空間分割し、さらに特徴抽出・認識処理をＭＩＭＤ型の並列処理により高速化している。

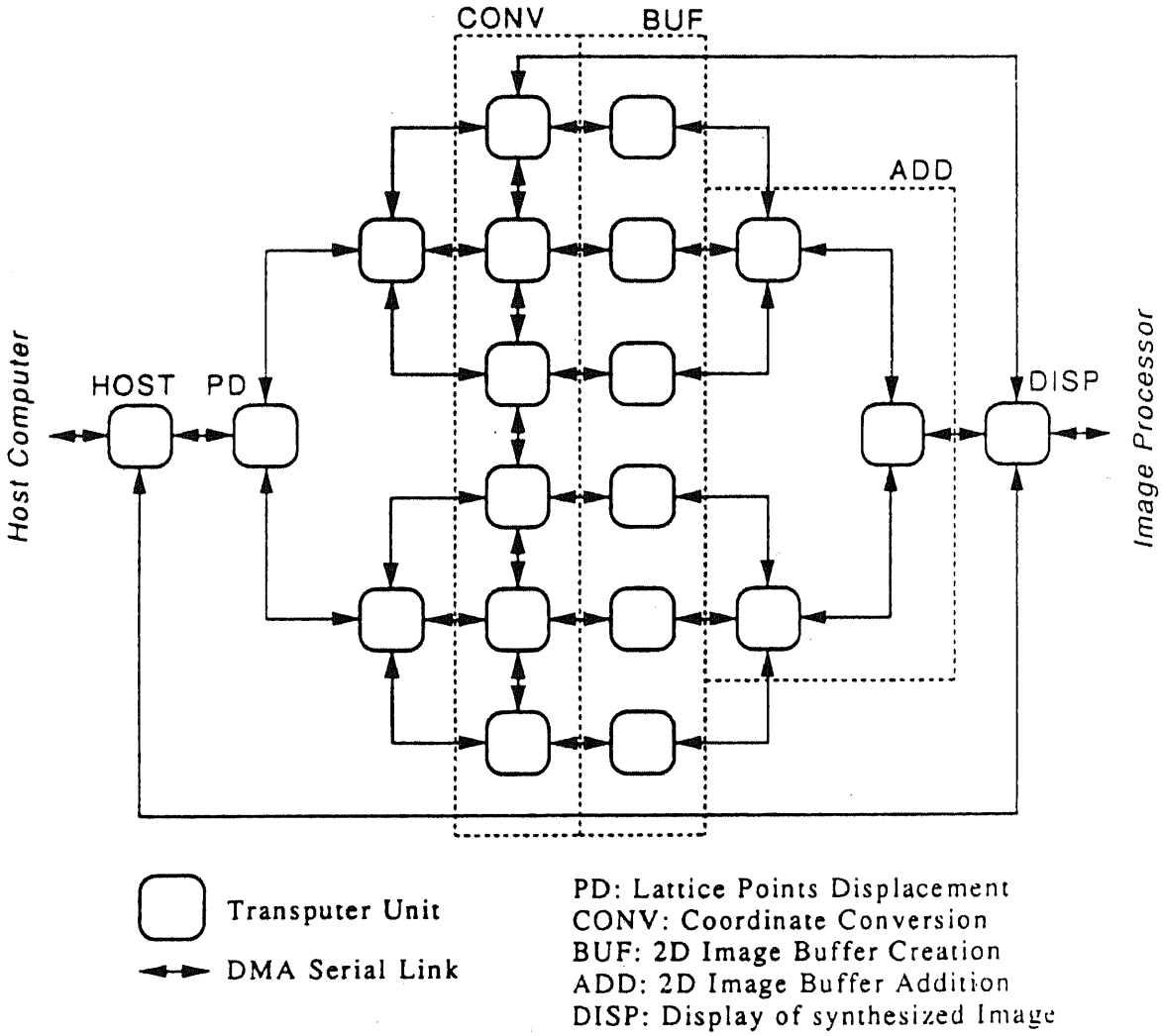


図１０．森島らの試作したトランスピュータによる

並列顔画像合成システム[16]

これにより、従来の逐次型処理システムでは画像の解析処理のみで１０分以上要していたが、スキャナによる画像入力（17秒を要する）を加え、３分程度で処理を完了することが可能となったと報告している。

従来の並列画像合成システムにおいても、ほぼ設計の主目的は画像合成の高速化にあると考えられる。画像の合成においては、パイプライン処理による並列化が一般的に用いられている。例えば成蹊大学の森島助教授らのグループは、トランスピュータ20台を用いてリアルタイムの人間の顔（表情）の合成を試みている（図１０）[16]。

４．１．２　ＶＩＴのハードウエア構成

画像認識と画像合成の高速化のために、ＶＳＡでは並列トランスピュータ・ネットワークを採用したが、現状ではトランスピュータのリンクの通信速度は画像データの転送には不十分である。そこで石塚研究室では、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32bit並列高速ビジュアルデータバス付きトランスピュータボードを開発し、ＶＩＴと名付けた[33]。

各ＶＩＴは、T805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、通常の４本のシリアルリンクの他に、32-bit画像データバスを有し、画像データの転送速度は、100Mbyte/sec(40nsec/pixel)である。

ＶＩＴの画像データバスは、マルチプロセッサにおける読み出しだけの共通メモリと同様に働き、どのＶＩＴでも高速にバスにアクセスできるが、アクセスの競合問題は起こらない（図１１参照）。

また各ＶＩＴは、カメラから画像データバスに入力された画像データの

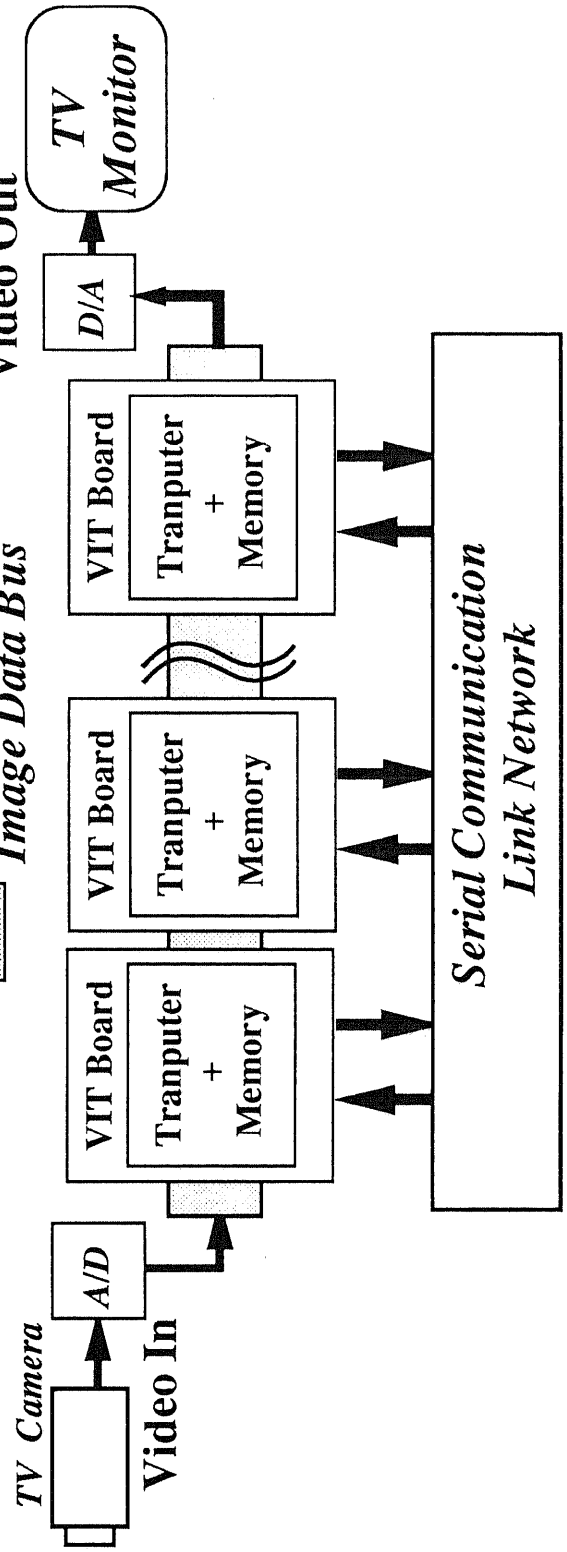


図１１．ＶＩＴにおける画像データバスの働き

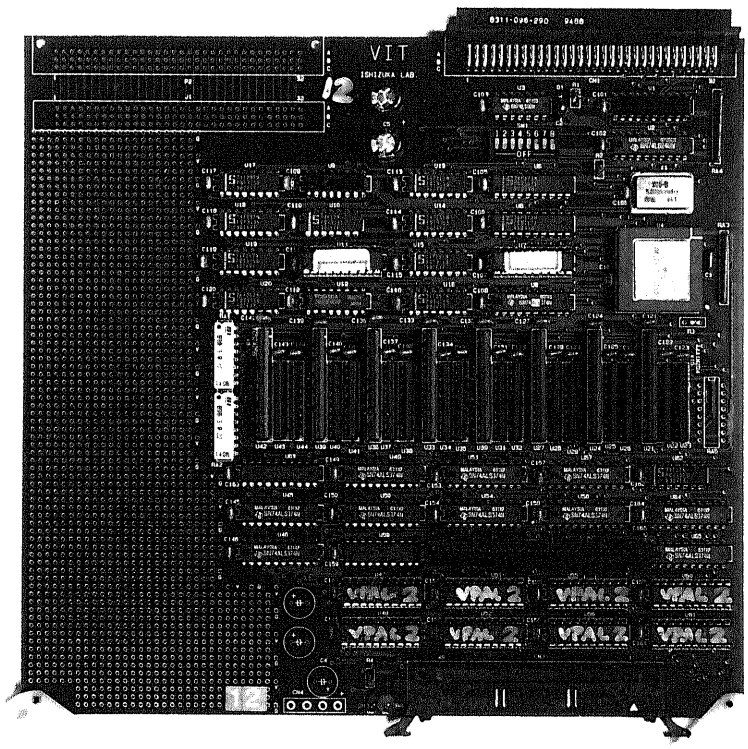


図１２．製作されたＶＩＴボード

処理、及びローカル画像メモリへの描画が可能である。画像の表示時には、ＶＩＴは隣のＶＩＴからの画像データと自らのローカル画像メモリ中のデータを、６つの画像合成モードから選択してピクセル単位で合成し、出力することができる。各ＶＩＴにおいて画像データの入出力時に要する時間は、１ピクセルクロック（ノン・インターレスモードで 33ns, インターレスモードで 66ns）であり、ほぼ無視できる。実際に作成されたＶＩＴは、ダブルハイト・ユーロカードの大きさであり、23.25×22.00cm である（図１２）。

４．１．３　ＶＩＴのソフトウェア構成

トランスピュータの利点は、整った環境とソフトウエアのサポートを有することのみならず、並列性の高い複雑なプログラムを比較的容易に書くことができるという点にある。またハイレベルの処理とローレベルの処理を同一のハードウエア上で実現でき、システムの汎用性が高い点も挙げられる。

ＶＩＴでは、このようなトランスピュータの開発環境をそのまま利用することができる。即ち、トランスピュータの標準言語 OCCAMとその環境を、ＶＩＴにおけるプログラミングにおいても何らの変更を加えることなく利用することができる。これは高速画像データ転送の性能と並びＶＩＴの大きな特長となっている。

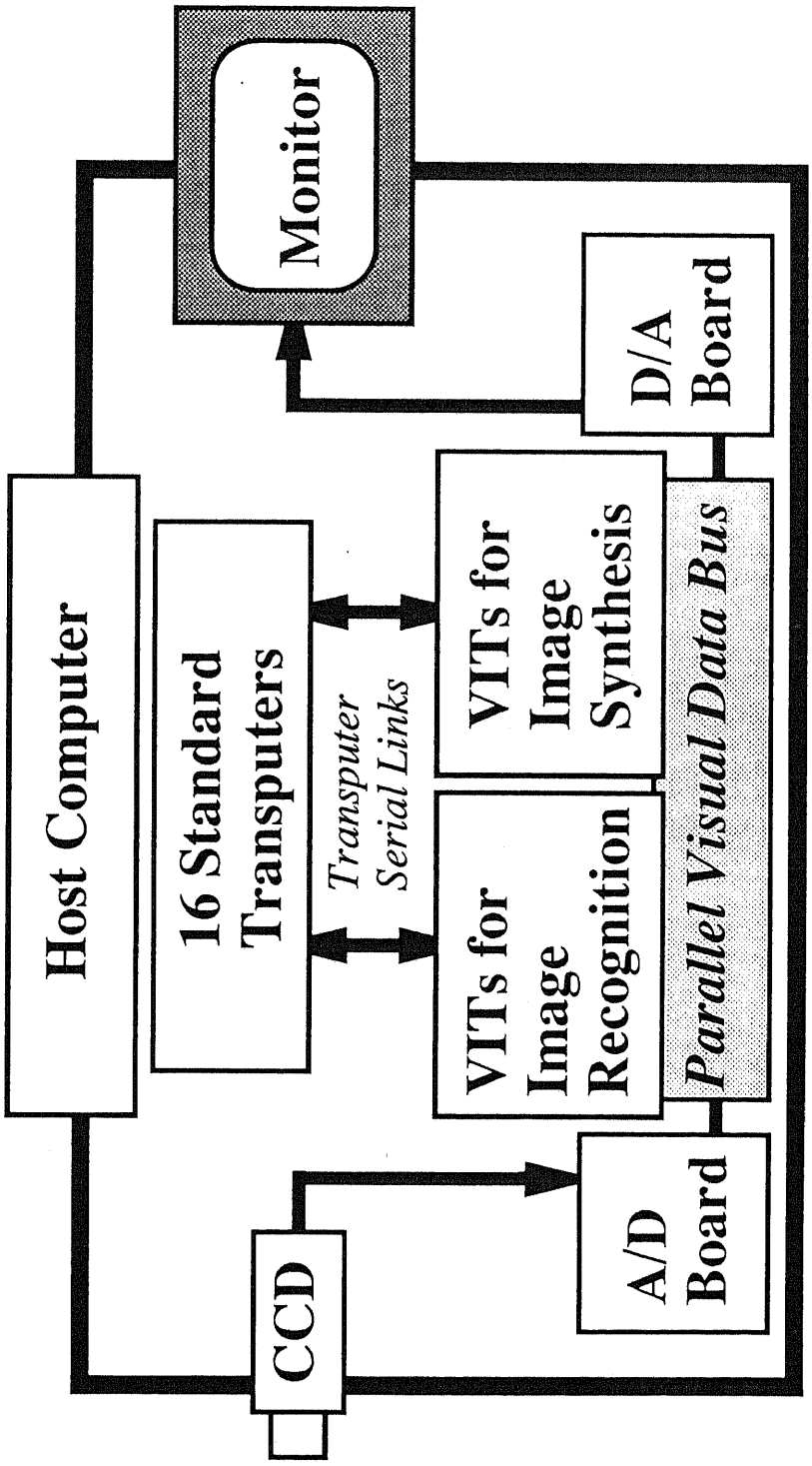


図１３．ＴＮ－ＶＩＴシステム構成

４．２　ＴＮ－ＶＩＴ　(Transputer Network with VIT)

本研究で構成した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、「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」点に、最大の特徴がある。

４．２．１　ＴＮ－ＶＩＴの構成

現在のネットワークの環境としては、ＶＩＴが３２台、通常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。また必要に応じ、ネットワークを拡張することも可能である。

その他の本システムの構成装置の主なものは、

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。図１３に、これらの装置の現在の接続状況を示す。この構成は、必要に応じ任意に拡張が可能である。

パーソナルコンピュータのうちの１台（NEC PC-9801RA）は、トランスピュータネットワークのホストコンピュータとして用いる。ホストコンピュータは、プログラムのエディット及びプログラムのトランスピュータネッ

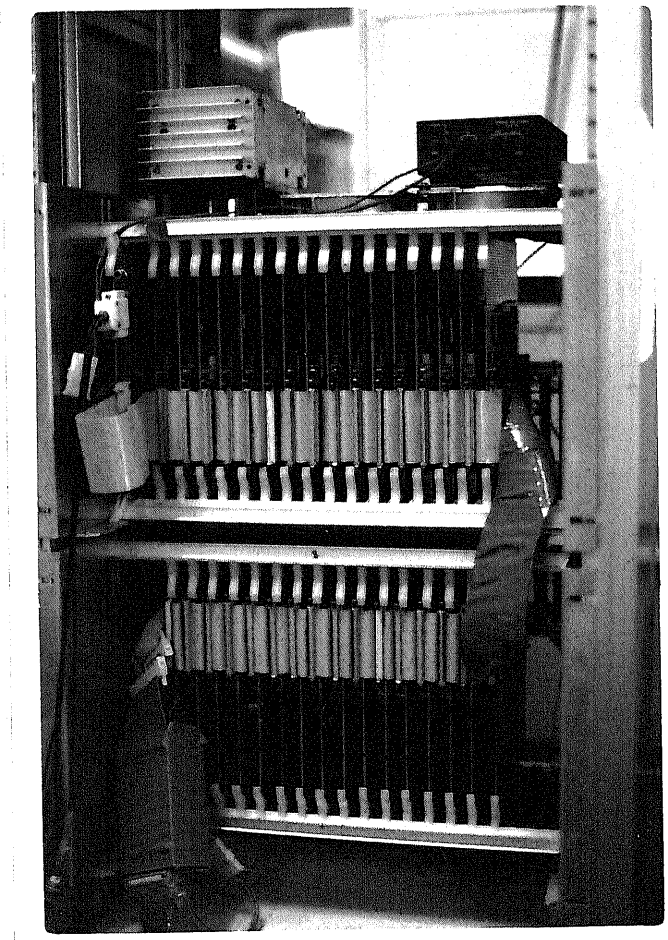


図１４．３２台のＶＩＴの外観

トへのダウンロード、またファイルの管理等を行う。

ホスト以外のパーソナルコンピュータは、テキスト画面の表示やマウスの制御用として用いる。このパーソナルコンピュータにはトランスピュータボードが装着されており、トランスピュータ標準リンクを介して最大20 Mbit/secで双方向通信が可能である。図１４にＴＮ－ＶＩＴシステムのうちの３２台のＶＩＴの外観を示す。

４．２．２　ＴＮ－ＶＩＴの特徴

（１）全体的特徴

ＴＮ－ＶＩＴでは、パソコンレベルのコンピュータからワークステーションに至るまで、TRAM (Tansputer Module)が接続可能ならば、標準リンクを介して双方向通信が可能である。すなわち、必要な機能を必要に応じ、容易にシステムに付加することが可能である。これはトランスピュータの機能とその環境が、そのまま利用可能であることによる。

このように、ＴＮ－ＶＩＴにおいては、システム構成のフレキシビリティの高さが第２の特徴として挙げられる。

（２）画像処理部の特徴

ＴＮ－ＶＩＴのうち、ＶＩＴを中心とする画像処理部では、ＶＩＴの各種機能を活用し、画像認識と画像合成を同一のハードウエア上で行うことができる。また通常のトランスピュータに知識ベース等を容易に付加し、高速でかつ高度な画像の認識・推論処理が可能である。この点において、ＴＮ－ＶＩＴはシリコン・グラフィックス社のＩＲＩＳなど、市販の高速画像合成コンピュータとは機能が異なっている。

さらに、画像処理部と画像合成部の構成は、ソフトウェアによりプロセ

スの実行中においても、任意に変更が可能である。すなわち、処理対象に応じ、画像認識部と画像合成部の構成をダイナミックに変更することが可能である。これらの点がＴＮ－ＶＩＴの第３の特徴である。

４．３　４章のまとめ

本章では、まず入力画像の認識・及び動画像の合成のために用いた、32bit高速画像バスつきトランスピュータボード：ＶＩＴ(Visual Interface to Transputers)について概説した。次に本研究において構築し、ＶＳＡにおいて重要な役割を果たしている、ＴＮ－ＶＩＴの構成と特徴について述べた。ＴＮ－ＶＩＴは32台ＶＩＴを中心とし、計48台のプロセッサよりなる並列コンピュータシステムであり、ＶＩＴの特徴を活用して画像の認識と合成が同一のハードウェア上で並列に実行可能である点が最大の特徴となっている。

画像認識と画像合成が同一のハードウェア上で実行可能であることの最大の利点は、ＴＮ－ＶＩＴ上の画像認識に用いるＶＩＴと合成に用いるＶＩＴの構成比が任意に設定可能であることにある。これは認識対象や合成対象に応じた負荷分散の設定が容易であることを意味している。さらに、ＴＮ－ＶＩＴではロードバランスの設定はプロセスの実行中においても、ソフトウェアによりダイナミックに変更可能である。画像認識用と画像合成用のコンピュータをつなげただけでは、このようなフレキシビリティの高さを実現することは極めて困難である。

このＴＮ－ＶＩＴの特長は、画像入力に反応する動画像の合成、すなわちソフトウェアロボットの実現には最適である。実際本研究においても、

認識対象や合成対象に応じ、認識部と合成部の構成比を変えている。

画像の認識という観点から見た場合でも、ＴＮ－ＶＩＴは他にないユニークな構造を有している。ＶＩＴバスを活用すれば、入力画像の処理に関してＳＩＭＤ型の動画像並列処理が可能な構造となっているからである。言い替えれば、カメラから入力されたカラーのビデオレート動画像に対し、複数のプロセッサが同時にアクセスすることが可能である。

本研究では、この利点を活かした並列動画像認識手法を提案している。これについては第５章において詳細に述べる。

他に画像の合成の観点からは、ＴＮ－ＶＩＴはＶＩＴの有する５つの画像合成モードを活用し、

１）空間分割による並列処理，

２）オブジェクト毎の並列処理，

３）時間分割型並列処理

などが簡単に実行可能な構造となっている。

また発展させて考えれば、ＴＮ－ＶＩＴは以下のような研究などにおいても、その特長を十分に発揮する事が可能である。

１）現在ＣＶとＣＧにおける共通の課題である、「モデルの共有問題」の研究に活用する。

２）カラー動画像のＳＩＭＤ処理が可能であることから、「生体の視覚機能のシミュレーション」のためのハードウェアとして用いる。

第５章　実時間画像認識による コマンドの入力法

５．１　ハンドサインの利用と対象物の実時間認識・抽出

５．１．１　本研究における画像認識の目的

冒頭にも述べたように、本研究では画像認識及び合成の技術を中心としたＶＳＡのプロトタイプを構築した。

本研究における画像認識処理は、ヒューマンインタフェースとしてのＶＳＡにおける利用を主目的とし、以下のような認識処理の実現を目標仕様とした。

①．複数の人物が随時出入りする動画像中から、システムを利用しようとする人物の顔のみを抽出する。

②．①で抽出された各人物の顔の向き（システムに注意を向けている，顔を向けている）が検出可能とする。

③．ＶＳＡの利用は、会社などの受付や、通常の室内環境であることを

想定しているため、照明条件や背景の設定などの制約は可能な限り設けない。例えば、昼間（窓から入る太陽光と室内照明）と夜（通常の室内照明のみ）という程度の照明条件の変化には耐え得るものとし、背景を統一するなどの操作は一切加えないものとする。

④. 画像処理のみに基づいた、ＶＳＡへの簡易なコマンドの入力が可能とする。

⑤. 利用者毎の初期設定などの煩雑な処理は省略可能とし、さらにハードウェア上の制約から、画像入力は全て単眼にて行うものとする。

⑥. ユーザに遅滞感を与えないために、認識処理の全てを実時間で実行する。ここで実時間とは、理想的には0.2秒未満での処理をいう。

以上の仕様を満たすために、従来提案されてきた手法[34-38]に加え、新たに幾つかの手法を提案してシステムに実装した。

まず５．１．２節では、上記の④・⑤・⑥に対する処理アルゴリズムを示す。次に５．１．３節では、①・③・⑤・⑥に対する処理を、また５．２節では、①・③・⑤・⑥に②を加えた処理アルゴリズムをそれぞれ示す。

５．１．２　ハンドサインによるコマンドの入力

（１）ハンドサインの利用

本研究では、画像処理を用いたシステムへの簡易なコマンドの入力手法として、ハンドサインを利用した。

ハンドサインを利用したシステムは、古くはMITメディアラボの "Put That There"システム[39]（磁気センサ利用，音声入力併用）があり、近年

も研究が盛んであるが、その多くがデータ入力にデータグローブ等の特殊なハードウエアを用いている。

これに対し本システムでは、ＶＳＡの第１段階として、ユーザの手の挙動をカラー画像入力し、これに人間型ソフトウェアロボット（エージェント）を反応させることとした。具体的には、手の位置に加え指の数に応じた１から５までの５つのコマンドを入力とし、エージェントはユーザの手の位置を、顔を振ることによりトラッキングさせると同時に、提示された指の本数により随時表情を変化させた。またユーザが指を立てない場合、システムに入力を行わずカメラの前で手を移動させることも可能とした。

データグローブ等のような特殊なハードウエアを利用しない理由は、ヒューマンインターフェイス研究の目的を考える際、コマンド入力が簡易であることが必須であり、その点ハンドサインは誰にでも容易に利用可能であることによる。

以下にハンドサイン利用の場合の長所と短所の詳細を、データグローブ利用の場合と比較して列挙する。

［ 長所 ］

①．入力のための特別な装置を必要としない。

②．右手用，左手用などの区別がない。

③．手の大きさ（個人差）に影響を受けない。

④．手の２次元平面内での移動距離が容易に計測可能。

⑤．処理開始時の初期設定などの必要がない。

［ 短所 ］

①．本ハンドサインの認識アルゴリズムのみでは、環境（照明の位置など）

に影響を受ける。

②. 掌または手の甲を、常にカメラに向けている必要がある。

③. 各指の関節毎の曲げ角は計測不可能。

④. 指の識別が困難。

⑤. 現段階では背景を調整する必要がある。本システムでは背景を白色とした。

⑥. 手の３次元空間内での移動距離は計測困難。

［その他］

この他、本システムで用いたハンドサインの読みとりアルゴリズムには以下のような特長がある。

①. 簡易な処理アルゴリズムであるため、高速な処理が可能。

②. 手の回転角に影響を受けない。つまり、掌か手の甲がカメラに向いていれば、どの向きでもデータ入力が可能。

③. 手袋などを装着する必要はなく、また長袖でも半袖でも読みとり可能。（ただし長袖の場合、袖のしわの状態、柄によっては認識率が低下する）

④. 必要ならば、一度に両手を用いることも可能。

（２）ハンドサインの認識アルゴリズム

本節では、ハンドサインの認識（提示された手の位置と指の数のカウント）アルゴリズムを示す。図１５に本アルゴリズムのフローチャートを示す。なお下記の画像処理過程は、すべてＶＩＴ上で行っている。

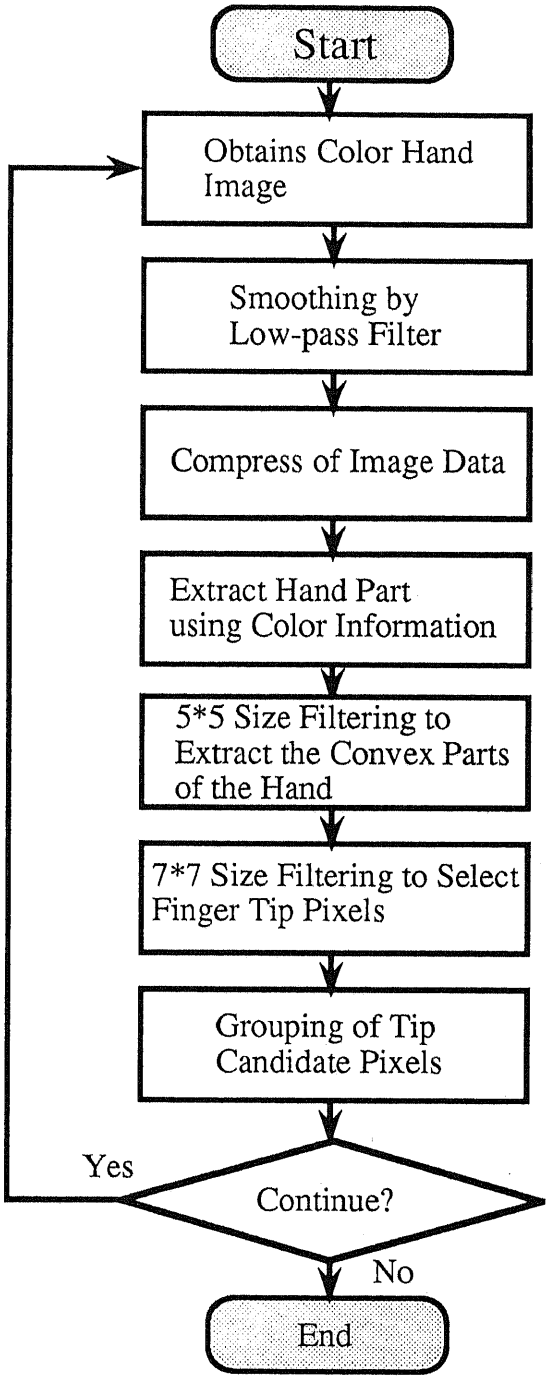


図１５．ハンドサイン認識アルゴリズムのフローチャート

①. 手の画像をカラー入力する(512×400pixel)。

②. ローパスフィルタによる画像の平滑化を行う。

③. 画像のx,y方向それぞれ１０ピクセル毎にデータをサンプリングすることにより、画像の圧縮を行う。

④. 手の色ベクトル空間内に存在するサンプルデータを抽出する。

⑤. ④で抽出されたサンプルデータに対し、５×５サイズのフィルタをかけ、フィルタリングの対象となる２５のデータのうち、④で抽出されたデータを１０以上、１６以下含むデータを抽出する。これは主として、近傍のデータ列が凸をなすデータ群の抽出を意味する（図１６）。これらの値は実験を基に決定した。以下の各値においても同様である。

⑥. ⑤で抽出されたデータに対し、７×７サイズのフィルタをかけ、フィルタリングの対象となる４９のデータのうち、④で抽出されたデータを２１以上３５以下含むデータを抽出する（図１７）。これは⑤で抽出されたデータ群のうち、掌または手の甲から遠い部分、すなわち指の先端部のデータ群のみを抽出することに等しい。

⑦. ⑥で抽出されたデータに対し３×３サイズのフィルタをかけ、近傍に存在するデータのグルーピングを行う。これは⑥までの過程で抽出された、同一の指に対する複数のデータ群を一つにまとめることを意味する。

以上のアルゴリズムにより、指の先端のみを抽出することができる。システムはこの抽出されたデータの数を数えることにより、カメラの前に提示された指の数を認識することができる。また手の位置は、抽出されたデータ（群）の重心を求めることにより算出している。

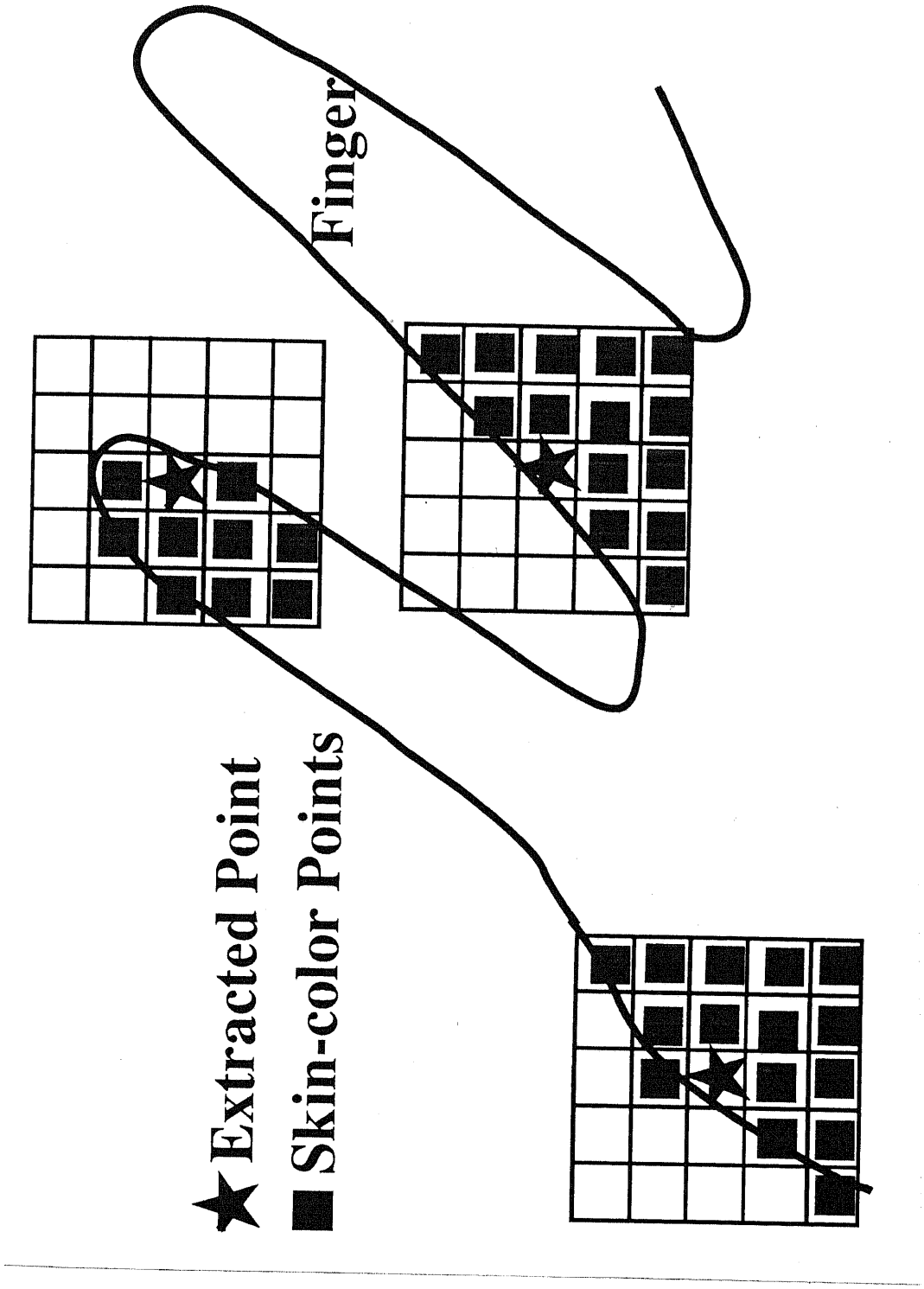


図１６．５×５サイズフィルタリング

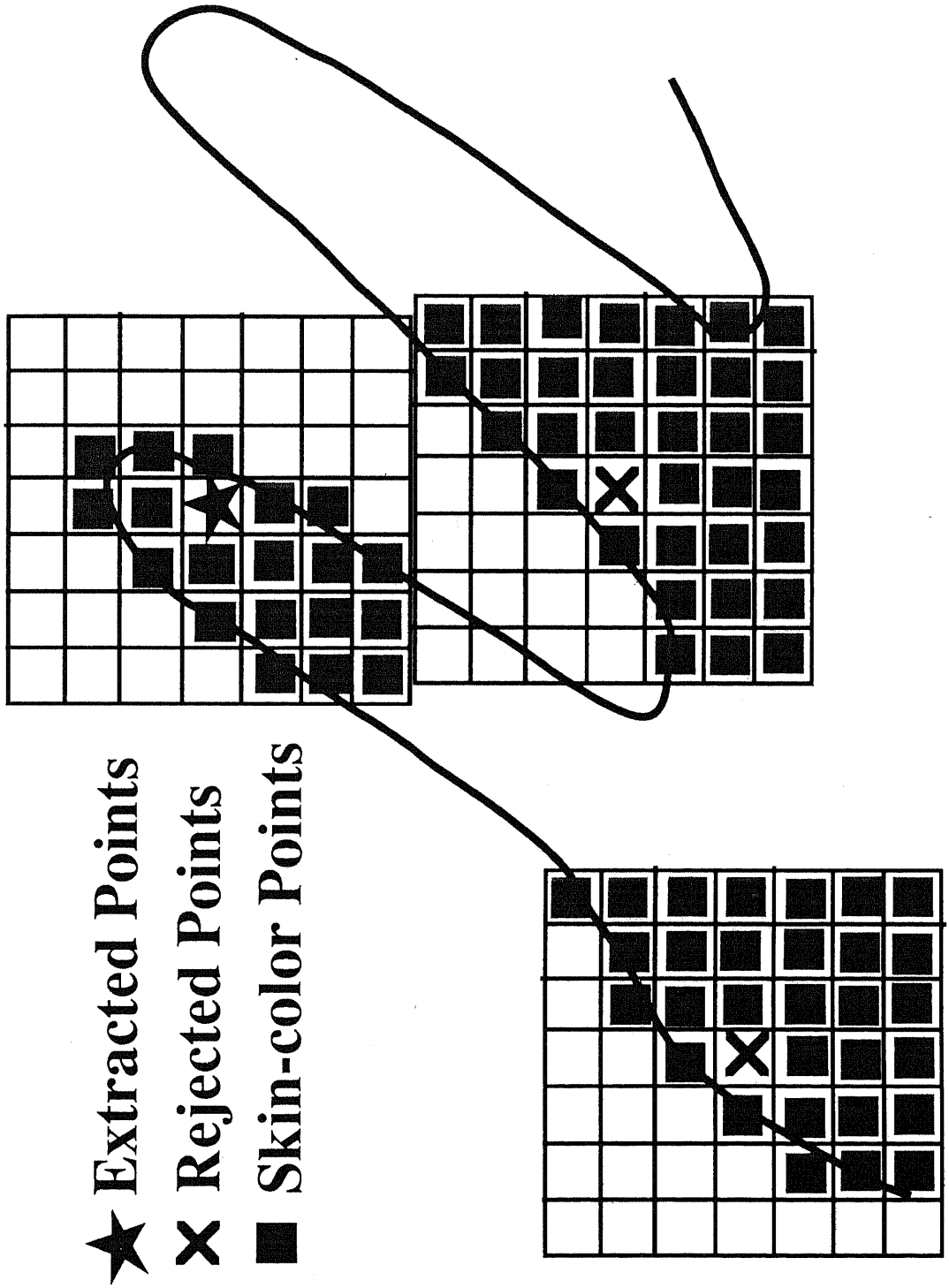


図１７．７×７サイズフィルタリング

５．１．３　画像特徴の統合による対象物の実時間認識・抽出法

（１）概要

ハンドサインの認識に続き、複数の人物が随時出入りする動画像中からシステムを利用しようとする人物の顔のみを抽出する画像処理アルゴリズムと、そのＴＮ－ＶＩＴ上への実装手法について述べる。

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１８）。

①．ボトムアップ処理

画像全体から並列に画像特徴を抽出して統合し、入力画像から人物像の候補を粗く高速に抽出する処理過程。このような画像特徴の並列抽出モジュールは、人間の視覚系においてもその存在が確認されている[44-48]。

②．トップダウン処理

抽出された人物像の候補の頭の部分にウィンドウを設定し、人物像であるかどうかの確認と頭の向きと口の開閉状態を計測する処理過程。

また図１９に、実際のシステム上でのＶＩＴとトランスピュータのコンフィギュレーションを示す。

（２）画像からの人物像の抽出（ボトムアップ処理）

［Ｉ］　画像特徴抽出モジュール

本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、

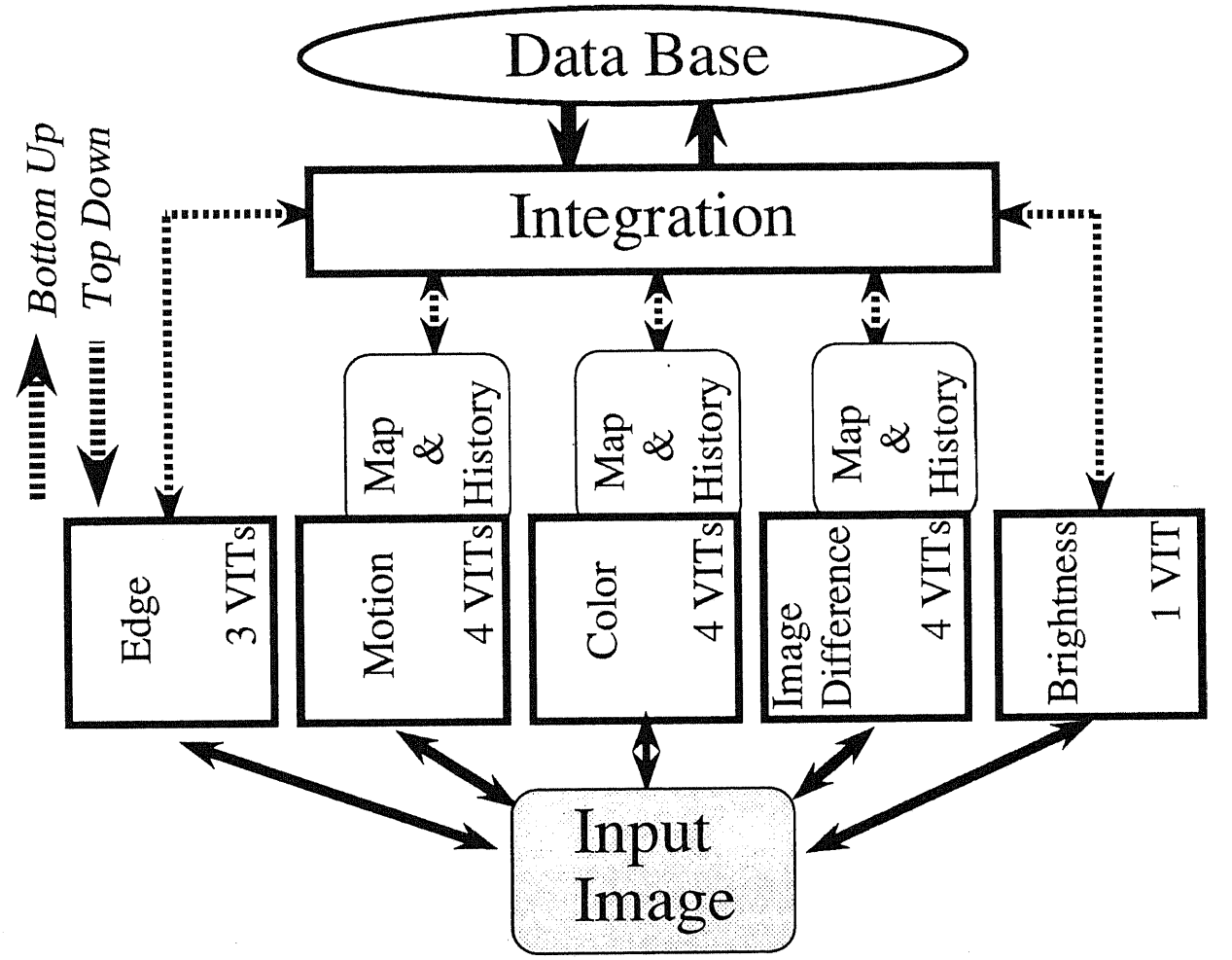


図１８．特徴情報統合手法の基本的アプローチ

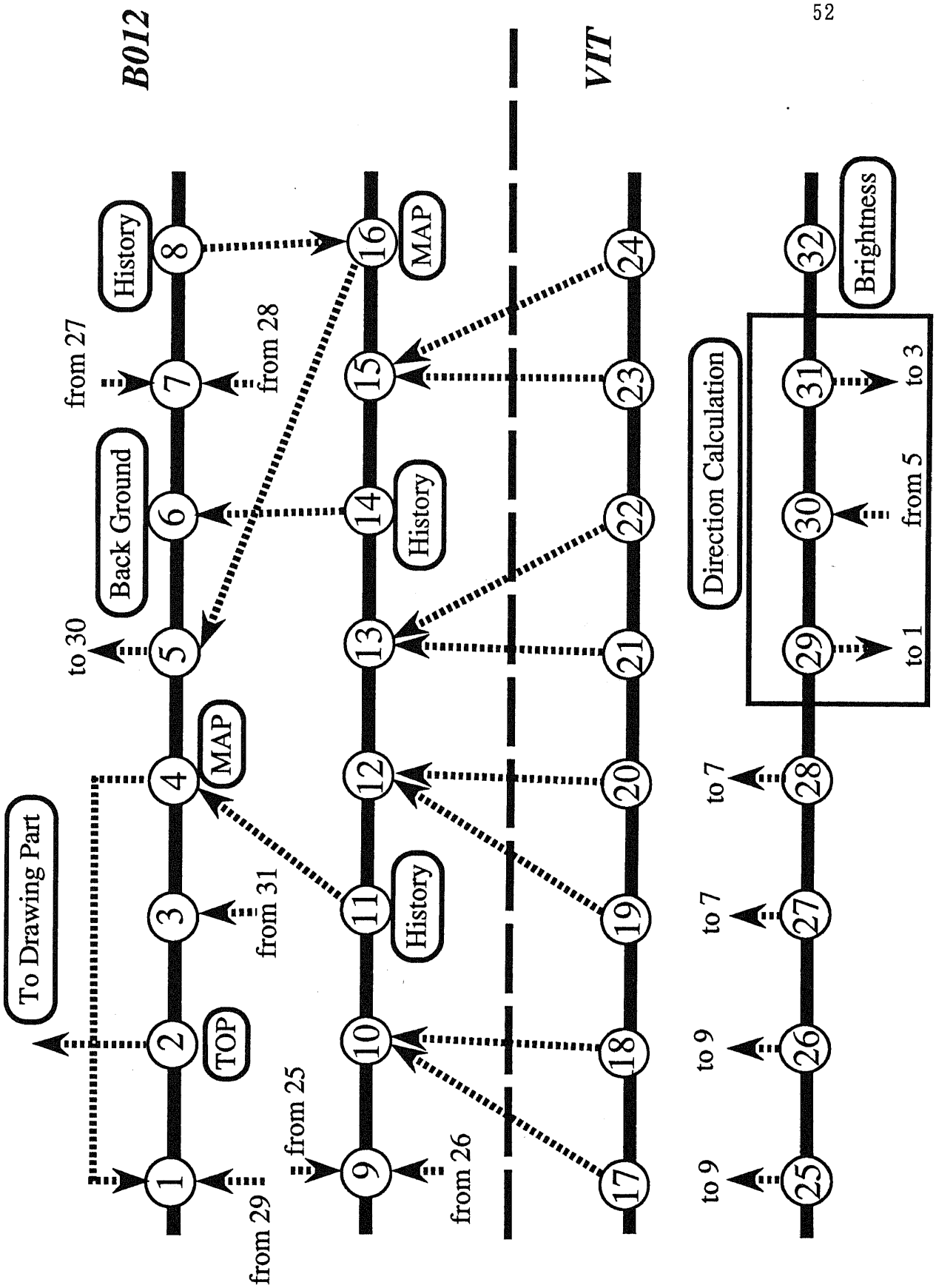


図１９．ＶＩＴとトランスピュータのコンフィギュレーション

①. 画像全体の明るさ抽出モジュール,

②. 動き情報抽出モジュール,

③. 色情報抽出モジュール,

④. 設定画像との差分抽出モジュール,

⑤. エッジ情報抽出モジュール,

の５つの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（［Ⅱ］節参照）を有し、特徴情報の統合過程で活用している。

以下に、システムに実装した各特徴抽出モジュールの有する機能について示す。各モジュール名の次に示した台数は、モジュールの構成に使用したVITの数である。

①. 明るさ抽出モジュール　　（VIT 1台）

本モジュールでは、画像上の平均の明るさを抽出する。厳密にはある入力画像Fn上の画素の濃淡値をFn(x,y)とするとき、その画像の平均の明るさBright_nは式(1)により算出している。

$$Bright\_n = \left(\sum_{x=1}^{m} \sum_{y=1}^{m} Fn(x,y) \right) / m^2 \qquad \ldots.(1)$$

現行では均等に配置した画像上400点（20×20点）を対象としている。Bright_nが大きく変動した場合、Bright_nを他の特徴抽出モジュールに伝達し、各種しきい値の再設定を促す。平均の明るさの算出から他のモジュールへのデータの伝達に要する時間は、100msec以下である。

②. 動き抽出モジュール　　（VIT 4台）

本モジュールはフレーム間差分により画像上の動きのある領域（２次元８方向）を抽出する。以下に抽出過程を述べる。まず連続する２フレームFn,Fn+1上の濃淡値をFn(x,y)及びFn+1(x,y)とする。ここでしきい値Threshold_Motionを定め、

もし

$$(Fn(x,y)-Fn+1(x,y))^2 \geqq Threshold\_Motion \quad ....(2)$$

ならば

$$Diff\_n(x+a,y+b)=\{(Fn(x,y)-Fn+1(x+a,y+b))^2\} \quad ....(3)$$

とする。ここで、(a = -1,0,1, b = -1,0,1)

このときDiff_nが最少値をとる方向を、動きの方向ベクトルとする。現行では画像上に均等に配置した2500点（50×50点）を計算対象とした。①のモジュールよりBright_nの入力があった場合、先のThreshold_Motionを参照テーブルに基づき調整する。参照テーブルは実験により設定した。

③. 色情報抽出モジュール　　（VIT 4台）

本モジュールは、R,G,Bカラー入力画像から各画素におけるR:G:Bの比を基に色情報を抽出する。現行では画像上に均等に配置した2500点（50×50点）を対象とした。①のモジュールよりBright_nの入力があった場合には、実験により設定した参照テーブルに基づき、認識対象色（ここでは髪と肌の色）の抽出時のR,G,Bの比を調整する。

図２０～図２３に明るさの変化に対する、５人の人物の髪と肌の色のRGBの構成比の推移の実測結果例を示す。本モジュールで用いた参照テーブルは、５人のデータを実測し、それらを平均して作成した。

また本モジュールでは、抽出領域に対し形状に基づいた重み付けも可能である（［Ⅲ］節参照）。

④. 設定画像との差分抽出モジュール　　（VIT 4台）

本モジュールでは、ある設定画像Setと入力画像Fnとの差分を抽出する。すなわち、Set,Fn上の濃淡値をSet(x,y)及びFn(x,y)とするとき、しきい値Threshold_Differenceを定め、

$$(Set(x,y)-Fn(x,y))^2 \geqq Threshold\_Difference \qquad ...(4)$$

を満たす画素を抽出する。現行では画像上に均等に配置した2500点（50×50点）を対象としている。①のモジュールよりBright_nの入力があった場合には、実験により設定した参照テーブルに基づき、しきい値Threshold_Differeceを自動的に調整する。またヒストリ機能により設定画像を局所的に更新する他、抽出領域に対し形状に基づいた重み付けを行う（［Ⅲ］節参照）。

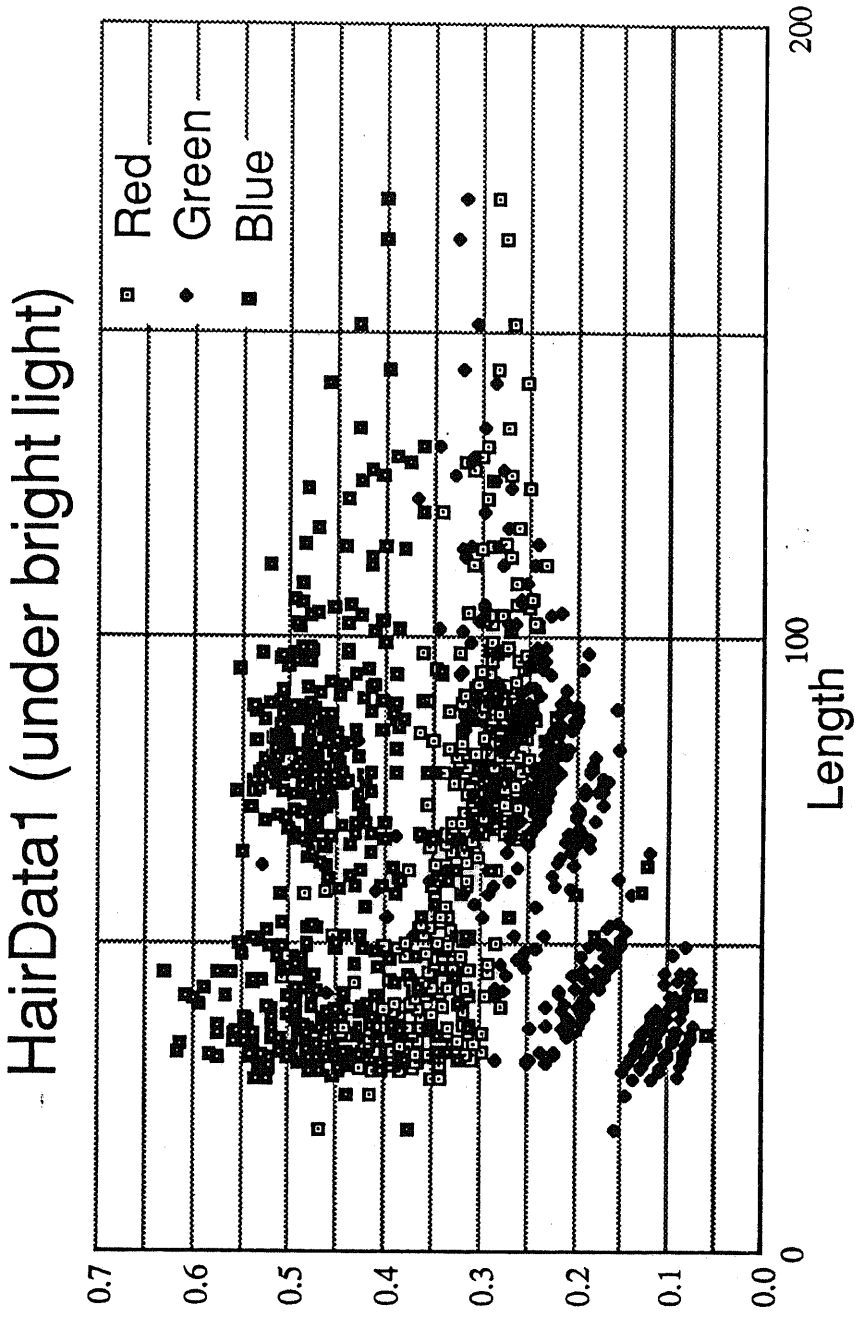


図２０．髪の色のRGBの構成比（明るいライト下）

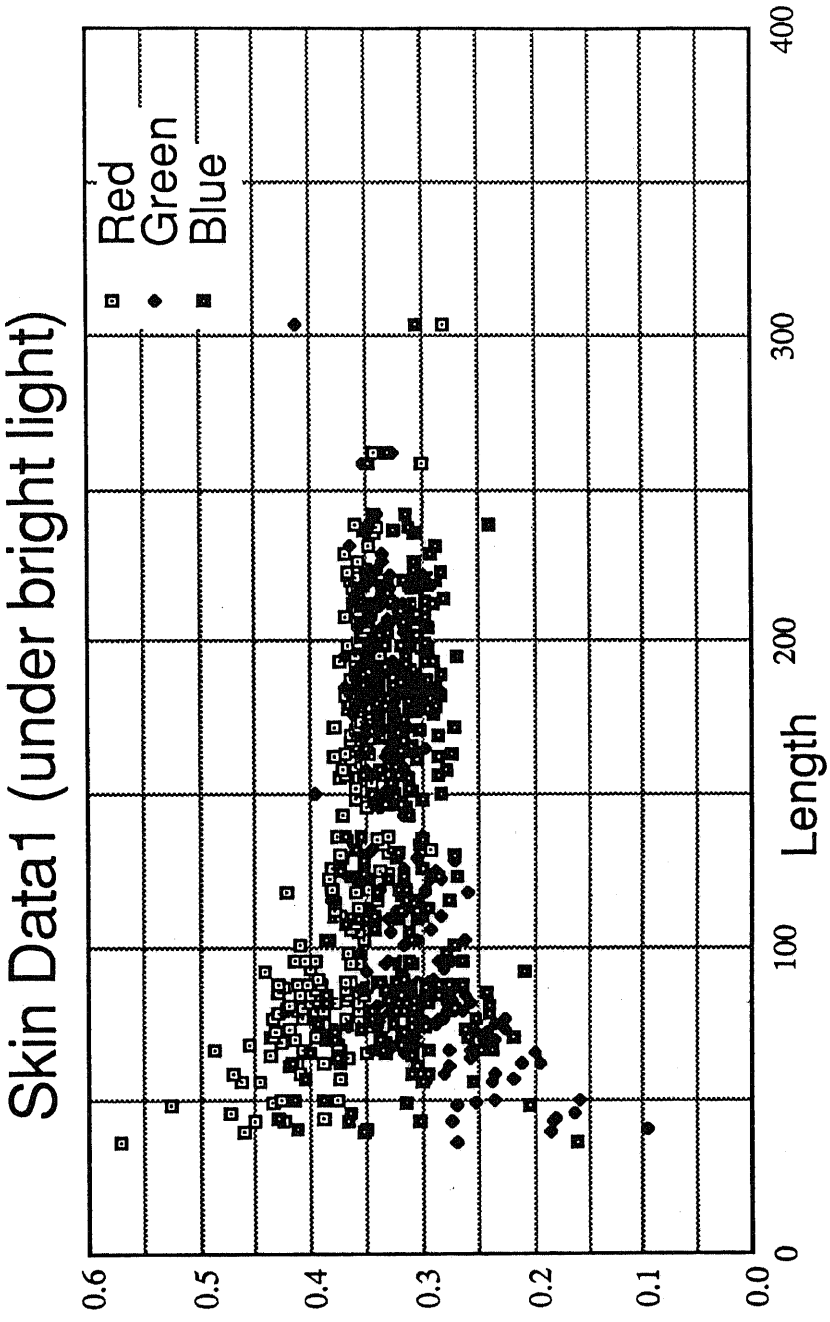


図２１．肌の色のRGBの構成比（明るいライト下）

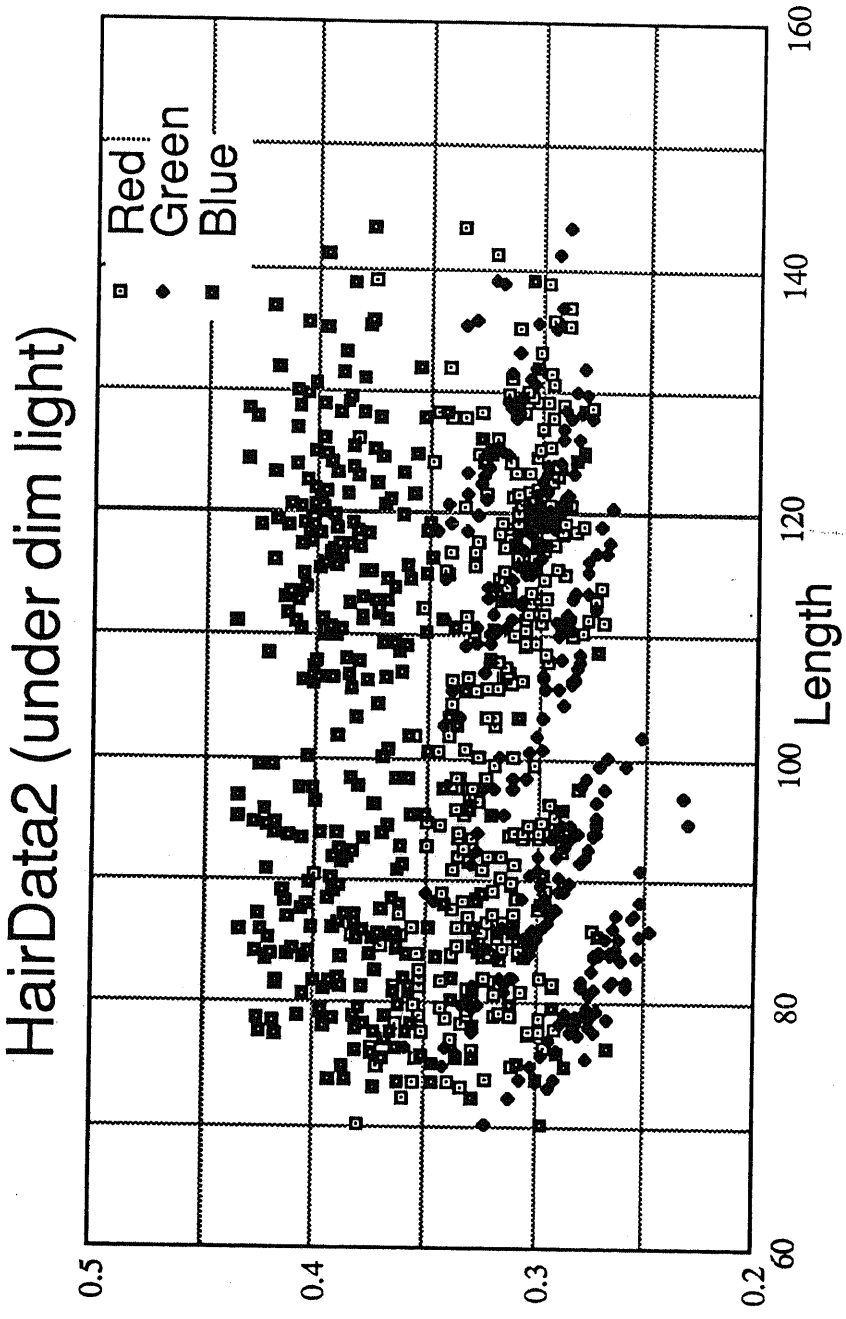


図２２．髪の色のRGBの構成比（薄暗いライト下）

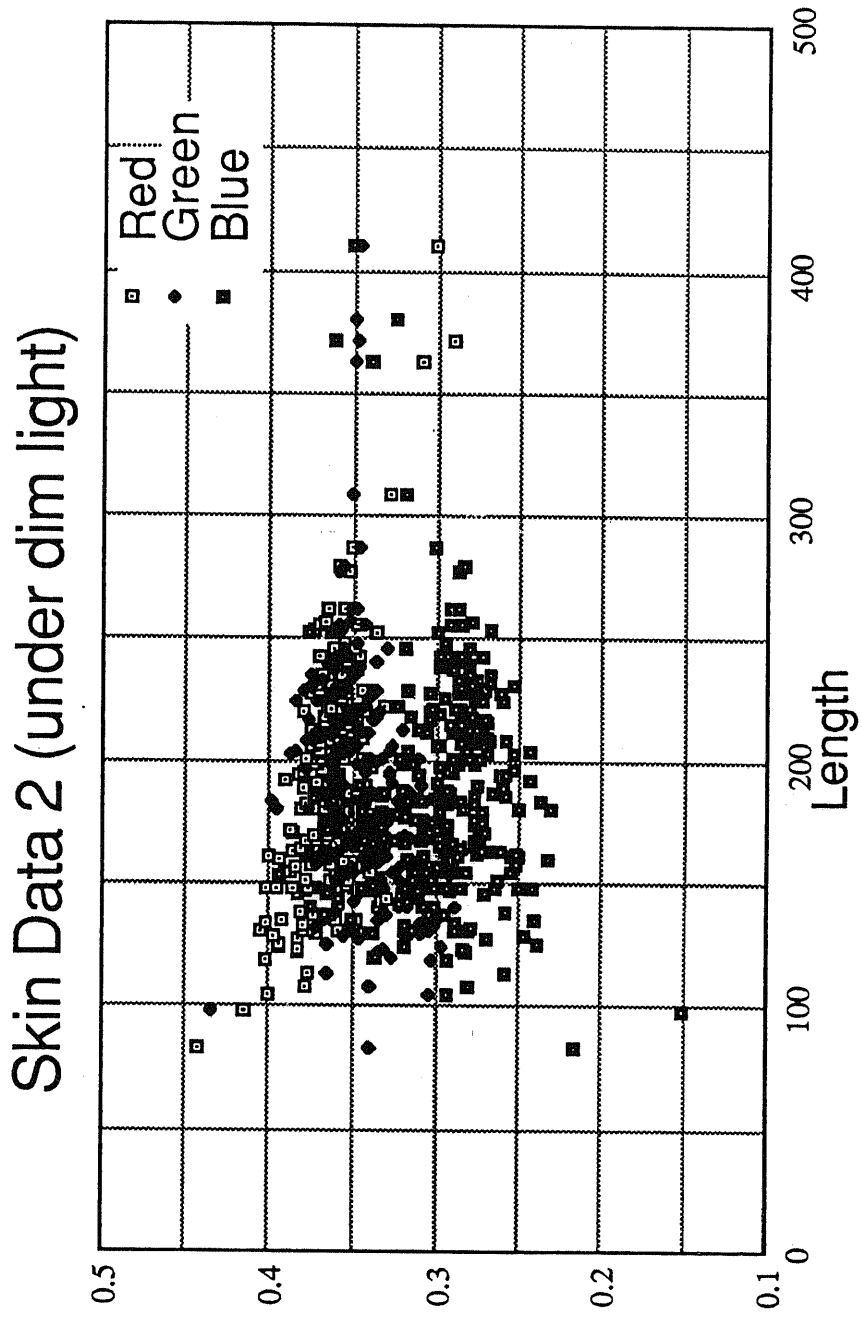


図２３．肌の色のRGBの構成比（薄暗いライト下）

⑤. エッジ抽出モジュール　　（VIT 3台）

本モジュールでは、トップダウン処理時の設定ウィンドウ内に$\nabla^2 G$（ラプラシアン・ガウシアン）フィルタをかけ、そのフイルタをかけた画像のゼロ交差の位置を見つけることによりエッジを抽出している。

ここで、ラプラシアン

$$\nabla^2 = \partial^2/\partial x^2 + \partial^2/\partial y^2 \qquad \ldots(5)$$

であり、ガウシアンは

$$\nabla^2 G = ((-1)/\pi\sigma^4)(1-\gamma^2 2\sigma^2) \times \exp(-\gamma^2/2\sigma^2) \qquad \ldots(6)$$

である。また入力画像をFn(x,y)とするとき、

$$(\nabla^2 G) * Fn = \nabla^2 (G * Fn) \qquad \ldots(7)$$

（ここで＊は Convolution を表す。）

であるので、実際の処理は入力画像にガウシアンをかけてぼかした後に、ラプラシアンをかけて強度変化を抽出している。最終的なエッジは、この強度変化が抽出された画像において、

$$\nabla^2 (G * Fn) = 0 \qquad \ldots(8)$$

となる場所としている。

ところで、本システムでは式⑥におけるσの値を順次変更することにより、エッジに関する多重解像度システムを実現している。実際には、

$$\sigma = 2, 4, 8 \qquad \ldots(9)$$

としている。

これにより、「認識対象物の大きさに基づく設定ウィンドウのサイズの差異を正規化し、処理スピードを安定化させる処理」を実現している。

このような機能は、人間の視覚におけるズーム機能と言われるものに相当する。

また①のモジュールよりBright_nの入力があった場合には、②～④のモジュールと同様、実験により設定した参照テーブルに基づき、エッジ算出時のしきい値を自動的に調整する。

［Ⅱ］　マップとヒストリ

各特徴抽出モジュールにおけるマップとヒストリの機能について述べる。

［マップ］

各特徴モジュールの出力をその位置と共に記述したもの。

［ヒストリ］

上記のマップ上のデータの時間的変化を記述したもの。すなわち、各特徴抽出モジュールからの出力の履歴情報が蓄積される。この情報は主として［Ⅳ］節に述べるMaplの作成時に利用される。

［Ⅲ］　画像データの重みの算出

①．重みの算出法

色情報及び設定画像との差分の抽出モジュールでは、抽出された各データ毎に重みを算出し、しきい値以上の重みを持つデータのみ出力して通信のオーバヘッドを少なくしている。本システムにおける重みの算出には、処理の高速化のために簡易な算出法を考案して利用した。

以下にその過程を Occam 風[78]に記述する。

まずそれぞれの属するデータ領域内(n×m)でのx+,y+方向（ＴＶ型ラスタスキャン）の街区画距離[37] Data_x+(x,y) 及び Data_y+(x,y)を求める。

```
For x = 1 to n
  For y = 1 to m
    If
      Data_x+(x,y) <> 0
        Data_x+(x,y) = Data_x+(x-1,y) + 1
      TRUE
        SKIP
    If
      Data_y+(x,y) <> 0
        Data_y+(x,y) = Data_x+(x,y-1) + 1
      TRUE
        SKIP
    y = y+1
  x = x+1  ■
```

同様にx-,y-方向（逆ＴＶ型ラスタスキャン）の街区画距離Dat_x-(x,y)及びDat_y-(x,y)を求める。このようにして求めた４つのデータのうち、最少値をとるものを抽出しData_1(x,y)とする。

次に、各データの８近傍のデータ数 Data_2(x,y) を求める。

$$Data\_2(x,y) = \sum_{x=1,y=1}^{n\ m} \sum Data(x,y) - 1 \qquad \ldots(10)$$

このとき、この各データの重みData_W(x,y)は、

$$Data\_W(x,y) = Data\_1(x,y) \times Data\_2(x,y) \qquad \ldots(11)$$

と算出する。

本手法は、１回のＴＶ型ラスタスキャンと１回の逆ＴＶ型ラスタスキャンで算出でき、高速な処理に適している。

また８方向ではなく４方向の街区間距離のみを用い、処理を簡略化して計算量の削減を図ったが、Data_2を掛け合わせることにより近傍のデータの密度も考慮してこれを補っている。

図２４にデータの重みの算出例を示す。斜線を施した部分を抽出されたデータ領域とするときの各データの重みの数値を示している。本例では３人の頭部がやや重なっている画像を想定しているが、重みの値は木の年輪のように徐々に大きくなり、この程度の重なりは分離可能であり、対象領

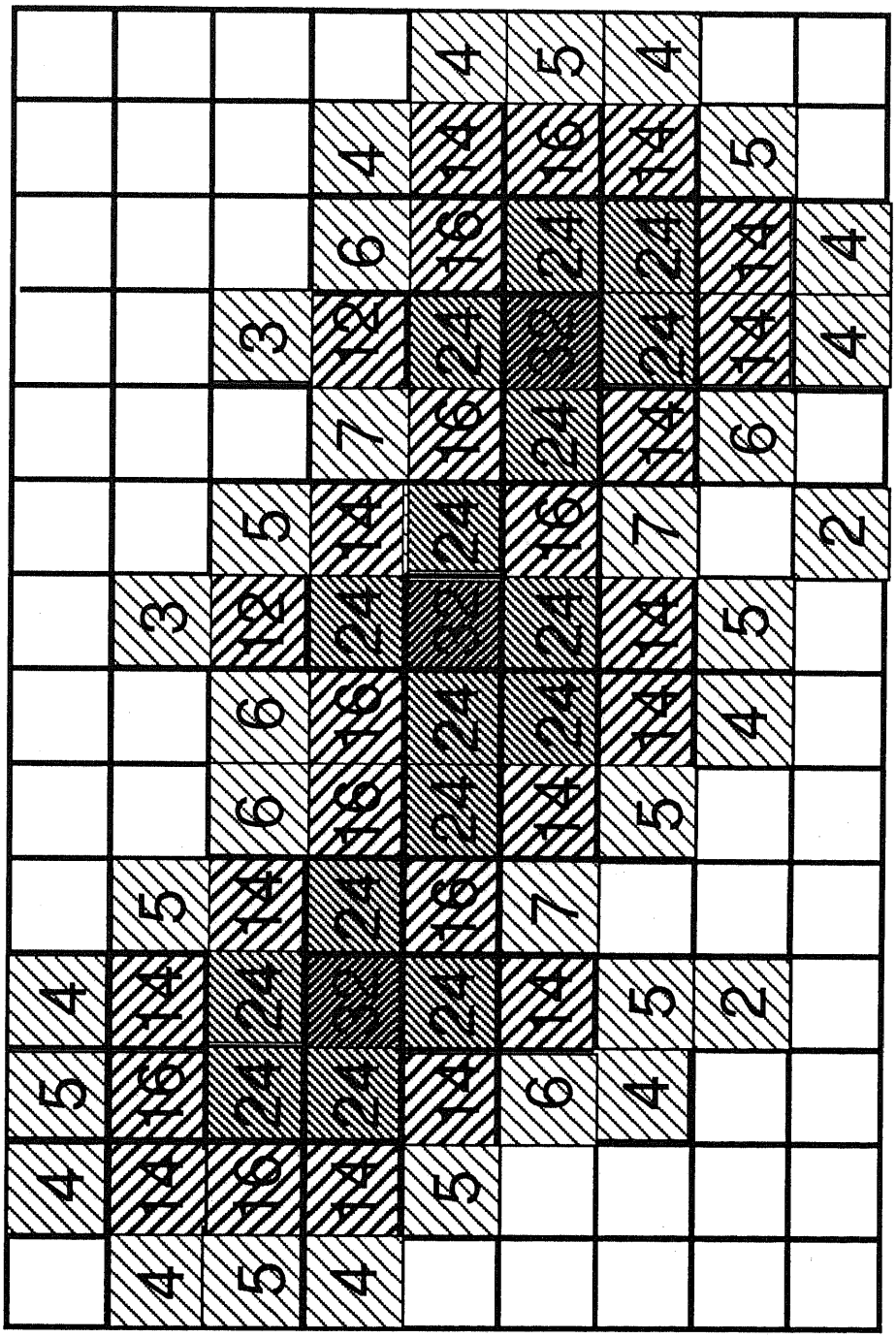


図２４．重みの算出例

域の大まかな形状情報は保存されている。

②. ヒストリの利用

動き，色，設定画像との差分，の各特徴抽出モジュールからの出力は、それぞれのヒストリにおいて重みの調整が行われる。本稿で述べる処理では、定常的に抽出され続けるデータの重みを小さくするよう設定している。

設定画像との差分Weight_diff(x,y)の抽出モジュールでは、ヒストリを用いて設定画像の更新も行っている。例えばこのモジュールからは、画像中の椅子の位置が少し変わっただけでも差分が抽出されてしまう。そこでヒストリの値History_diff(x,y)がしきい値Threshold_hist_diffを越えた領域については、入力画像のデータを設定画像データ中に取り込み、データの更新を行うようにしている。すなわち、式(4)

$$(Set(x,y)-Fn(x,y))^2 \geqq Threshold\_Difference$$

を満たす時、

$$Weight\_diff(x,y) = (Set(x,y)-Fn(x,y))^2 \quad ...(12)$$

$$History\_diff(x,y) \leftarrow History\_diff(x,y) + 1 \quad ...(13)$$

ここで、

$$History\_diff(x,y) \geqq Threshold\_hist\_diff \quad ...(14)$$

ならば

$$Set(x,y) = Fn(x,y) \quad \ldots(15)$$

$$History\_diff(x,y) = 0 \quad \ldots(16)$$

としている。

同様に色情報の抽出モジュールでは、画像中に人物のポスターなどの虚人物像が存在する場合その位置には常に肌と髪の色が抽出されるが、ヒストリの値が大きくなるにつれて重みを小さくし人物像の候補としての優先度が低くなるようにしている。

また画像中に風にそよぐ木の葉などが存在する場合などは、その部分より頻繁に動きの情報が抽出されるが、ヒストリによりその部分の重みは小さくしている。

［Ⅳ］　特徴情報の統合

特徴情報の統合過程を図２５に示す。

ここでヒストリを経た各モジュールからの出力を以下のように定義する。また以下の各統合過程で Occam風の記述を用いたが、これらは概念及び計算式の概要のみを示すためのものであり、実際のプログラムを示したものではない。

画像全体の明るさ抽出モジュール　→　Bright_n

動き情報抽出モジュール　→　MH2(x,y)

色情報抽出モジュール　→　MH3(x,y)

設定画像との差分抽出モジュール　→　MH4(x,y)

エッジ情報抽出モジュール　→　MH5(x,y)

①. 色情報と設定画像との差分情報の統合

まず、ヒストリを経た色情報と設定画像との差分の統合が行われる。この結果、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Map1に記述される。

```
For x = 1 to m
  SEQ
    For y = 1 to n
      SEQ
        Map1(x,y) := MH3(x,y) * MH4(x,y)
        y := y + 1
    x := x + 1
```

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Map2上にヒストリを経た動き情報が書き込まれる。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。

すなわち、Map2にはMap1上の人物画像の候補の移動先と他の動きのデータを加えたデータが、それぞれ区別されて書き込まれる。

```
For x = 1 to m
  SEQ
    For y = 1 to n
      SEQ
        Map2(x,y) := Map1(x,y) + Relation(x,y)
        y := y + 1
    x := x + 1
```

上式において、Relation(x,y)には、人物画像の候補と動きデータの関連が記述されている。この関係は主としてデータ間の距離に基づき決定している。図２６に動きの情報の抽出結果例を示す。

③．明るさ情報の統合

明るさの情報 Bright_n 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。

図２７に明るさ情報の抽出処理例と、Map1の情報の出力例を示す。

④．エッジ情報の統合

エッジの情報はトップダウン処理過程において用いられ統合される。これについては次節以降に述べる。

図２８に人物画像の候補領域の周囲に設定された、ウィンドウ内のエッジの抽出例を示す。

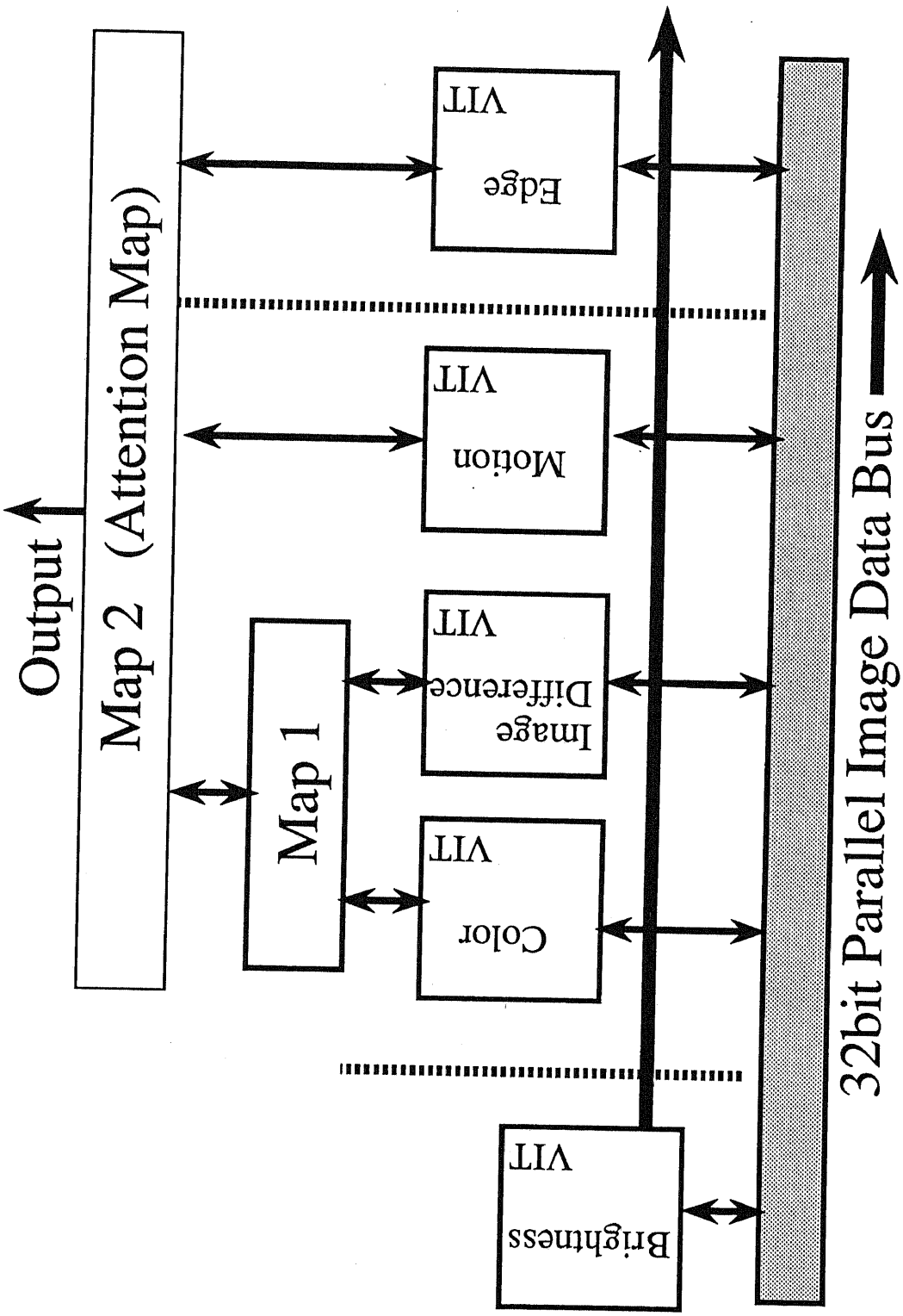


図２５．特徴情報の統合過程

図２６．動き情報の抽出例

図２７．明るさ情報とMaplの記述の出力例

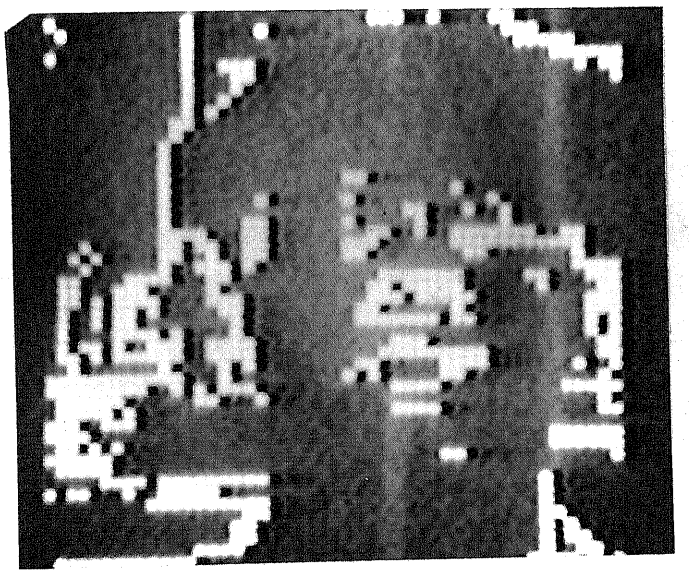


図２８．ウィンドウ内のエッジの抽出例

（３）　人物像の頭の向きと口の開閉の計測（トップダウン処理）

図２９に、本システムにおけるトップダウン処理の概念を示す。本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、顔の向き及び口の開閉の判定を行う処理として実行される。

具体的には、まずMap1中の顔の候補の位置・大きさ・優先順位のデータを得て、Map2に記述された最も優先順位の高い候補の移動後の位置にウィンドウを設定する。これにより、トップダウン処理を起動する時点での顔の位置にウィンドウを設定することができる。

次にウィンドウ内のエッジを抽出し、同一位置のMap1上のデータを参照する。このときシステムは、

①．髪と肌の大まかな形状，

②．髪と肌の位置関係，

③．肌色と抽出されたエッジ部が重なった量，

をデータベースと照合する。これにより、顔かどうかの判定と顔の向きの検出を行っている。さらに顔と判定された場合には、

④．口に相当する部分に赤みがかった黒（口の中の色）が検出されていないか，

を調べ、検出が確認された場合は「口が開いている」と判定している。

このように、本システムにおける「動画像からの顔の切り出しとその向きの検出」のためのデータベースとルールは極めて簡易なもので済み、本提案手法の特徴の一つとなっている。また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）は完全に並列に実行されるという特徴も有する。

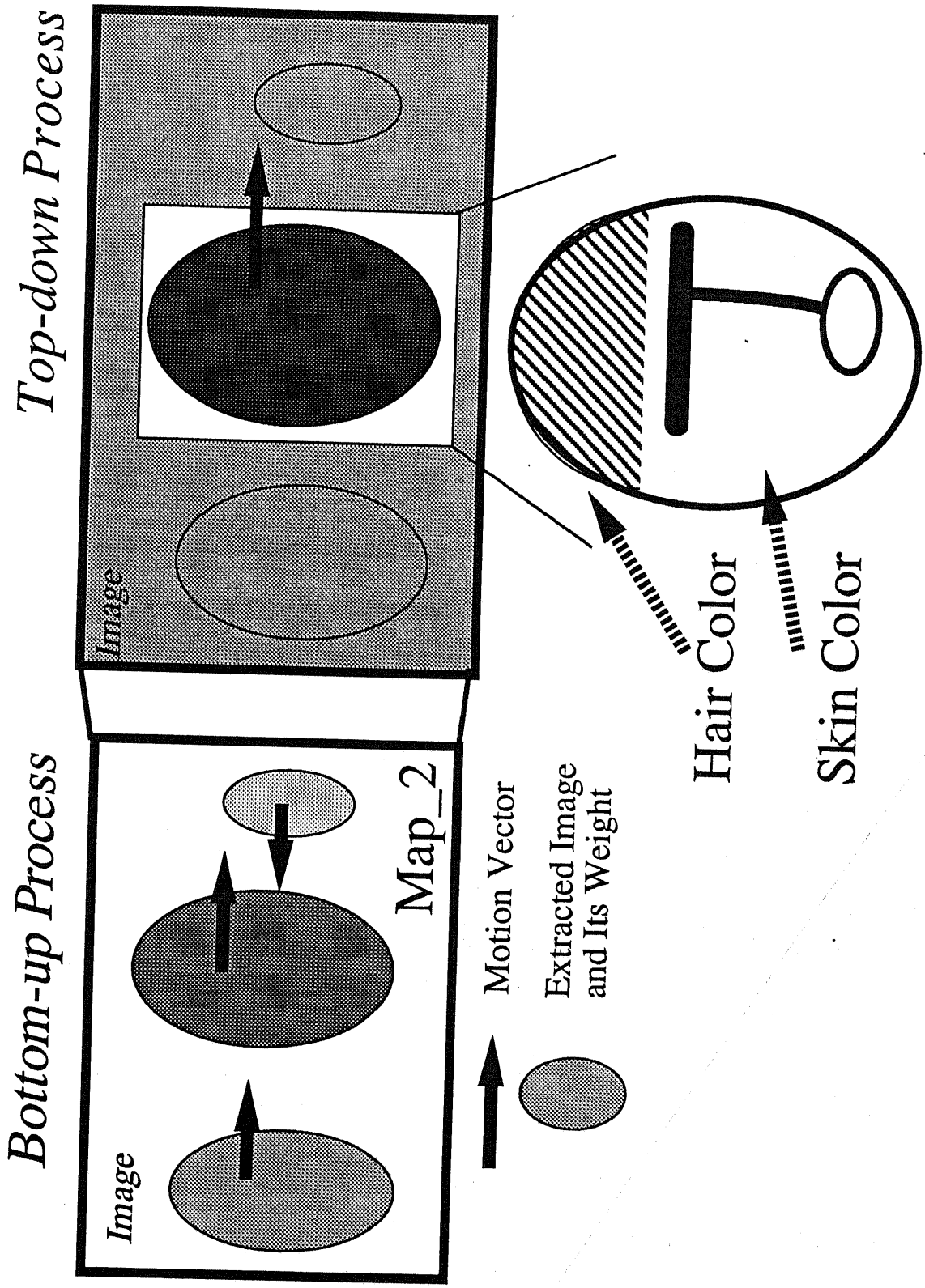


図２９．トップダウン処理

（４）　認識対象に対する制約

本認識手法は、現段階では認識対象に関し以下に挙げる限定条件を要する。

１）髪と肌の色は、標準的な日本人の髪と肌の色またはそれに準じる色とする。

２）口ひげ程度ならば処理は可能であるが、あご全体をおおうひげを持つ人は対象としない。

３）前髪の存在が必要。

４）鉛直方向に対し、±30°程度以上の顔の傾きの検出は対象としない。

ここで、本システムでは色の抽出パラメータは可変であり設定の変更は容易なため、１）は大きな制約とはならない。２）～４）は、データベース中の顔の記述とマッチしないことによる制約である。

（５）従来の研究との比較

［Ｉ］「顔」画像の計測・認識

従来の顔を対象とした画像認識に関する研究は多数報告されている[50, 51, 54]が、ここでは間瀬らによる研究事例と比較する。間瀬らは顔の向きの計測手法、及び計測結果に同期して動くレンダリング顔画像の合成システムに関して報告しており、本研究におけるシステムの構築の目標と類似点が多い[52]。間瀬らの手法は、頭及び顔領域の面積と重心の情報を基に頭の向きを検出するが、①背景は固定，②照明条件は一定とする，③画像中の認識対象人物は常に肩から上の画像が得られるものとする，④髪と肌の輝度にコントラストを必要とする，等の幾つかの限定条件が必要であった。

これに対し本手法では以下のような特長を有する。

①. 背景・照明条件に制限はなく、通常の室内環境下で実行可能。

②. 複数の人物像に対して処理が可能。

③. 入力人物像の画像上におけるサイズに制限はない。

④. また前髪の存在と両目・鼻・口の存在が確認できれば、女性特有の多様な髪型に影響を受けにくいほか、通常の眼鏡ならば処理は可能。

角らは、モデルに基づく濃淡顔画像の解析を並行にトップダウン処理をかけて行う手法を提案している[53]が、人物の識別が主目的であり、本提案手法とは処理の目的が異なる。

［Ⅱ］従来の画像処理アルゴリズムとの比較

本システムにおけるボトムアップ処理では、各特徴抽出モジュールは常に情報を抽出し続け、どのモジュールに変化が検出されてもその変化は実時間で処理に反映される特徴を有する。即ち従来の逐次的・直列的に処理を進める画像処理体系とは異なり、並列処理の利点を活かし、環境の変化を常にモニタリングして多様な環境の変化に対しある程度の頑健性を有する処理体系となっている。このような思想は、ロボティクスの分野においても適用された例がある[49]。

（６）本手法のまとめ

以上に述べた本提案手法の特徴の主なものをまとめると、以下のようになる。

＜ボトムアップ処理＞

①. 直列的処理体系を持たず、並列型の処理体系を有する

→ 多種の画像特徴の変化を常にモニタリングし、実時間で対応可

能

②. 多種情報の統合による画像処理により上位レベルでは、

→ 必要な情報を必要に応じて活用

膨大な画像情報から有意な情報のみを抽出し高速な処理を実現

③. 重みを利用する認識手法をとり、複雑な処理中枢を持たない

④. 重みによる情報の意味付けと活用

→ 上位へ伝えるのは絞られたデータのみ

上位へ行く程扱うデータ量が少なく、高速な処理が可能

⑤. ヒストリは常時更新されているため、システムは徐々に環境に“慣れて”処理がより速くなる

＜ボトムアップ処理とトップダウン処理の融合過程＞

ボトムアップ処理 → 画面全体を荒く見る,

大まかな処理対象の位置の抽出

人間の視覚情報処理における周辺視

トップダウン処理 → じっくり見る,

必要な情報のみを細かく処理

人間の視覚情報処理における中心視

のシミュレーションであると言える。

5．2 その他の周辺装置によるコマンドの入力

①. キーボードによる入力

現在のコンピュータにおける最も一般的な入力装置であり、ＶＳＡにおいても利用するが、コマンド入力などの簡単な処理においては、可能な限りこれを用いない方針である。

利用する際は、メイルの記述などのシステムがあらかじめコマンドなどとして想定できない入力を行うためにのみ用いる予定である。

②．マウスによる入力

キーボードに次いで一般的な入力装置であるが、ＶＳＡにおいてはこの装置も可能な限り用いない方針とする。

利用する際は、簡易な入力装置である利点を活かし、手書き文字や図形の入力デバイスとして利用し、コマンド入力などの簡単な操作には極力利用しない。現在のＶＳＡでは手書き文字の認識を行う予定はないが、書かれた文字を図形データとして保存・際利用する等は、容易に実現可能であろう。

５．３　５章のまとめ

本章では、ユーザからシステムへのデータ入力として、画像処理によるハンドサインの認識、及び動画像からのユーザの認識アルゴリズムについて述べた。また、その他の周辺装置のＶＳＡにおける利用形態についても簡単に触れた。

ハンドサインの認識アルゴリズムは従来様々なものが提案されてきた[40-43]が、本論文で提案した手法は実時間処理を優先度の第一とした。このようなデータグローブなどの特殊なハードウェアを用いない非接触の手法は、今後ますます研究・利用されていくものと思われる。

また動画像からのユーザの認識アルゴリズムは、同じく実時間処理を優先度の第一とし、かつＶＩＴの特長を利用する処理形態とした。本手法は

入力画像に対する制約が少なくて済み、アルゴリズム的に環境の変化に実時間で対応が可能である他、一度に複数の人物をトラッキングすることが可能という特徴も有する。

さらに、画像特徴の統合に基づく画像認識手法は、人間の視覚系における特徴抽出モジュールと、それらからの情報の統合の初期シミュレーション（単眼）となっている。これについては、第７章においてさらに触れる。

ヒューマンインタフェースにおける画像認識技術の利用は、現在盛んになりつつある。その理由の主なものとして、非接触の計測が可能であることからユーザの負担が少なくて済むことが挙げられる。ヒューマンインタフェースにおける画像認識技術の利用分野のうち、主なものから幾つかをまとめると下図のようになる。図中のLevelとは要求される技術レベルを示し、右に行くほど要求技術レベルが高いことを示している。ただしここに示したのは一般的な画像を対象とした極めて大まかな分類であり、入力画像の状態や認識対象により難易度は大きく左右される。

Level 1 人がいるかどうかの認識，
単数の人の動きの認識・追跡，
Level 2 複数の人物の動きの認識・追跡，
手振り、身ぶりの認識，
顔の位置と向きの認識，
Level 3 表情の認識，
視線の検出，
指文字、手話の認識
Level 4 微妙な表情の変化の認識，
人物の識別，

←易 難→

現在はLevel 2の確実な達成と、Level 3への挑戦が各研究機関で進められているという段階であろう。画像認識技術を利用したヒューマンインタフェースが十分にその役割を果たすためには、Level 3 以上程度の能力が必要であるものと考える。今後の研究の成果が期待される。

第６章　動画像のユーザへの実時間出力法

６．１　人間型エージェントの動画像合成

６．１．１　人物頭部ワイヤフレームモデル

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。

図３０に３次元人物ワイヤーフレームモデルを示し、図３１に顔画像のモデルをレンダリングした画像を示す。ポリゴンは３角形状のものを用いているが、これはいかなる形状でも３角形で近似的に表現でき、一般性を失わないと考えたためである。

なお本ワイヤーフレームモデルは、東京大学工学部・原島（博）研究室から提供頂いたものをベースとしている。

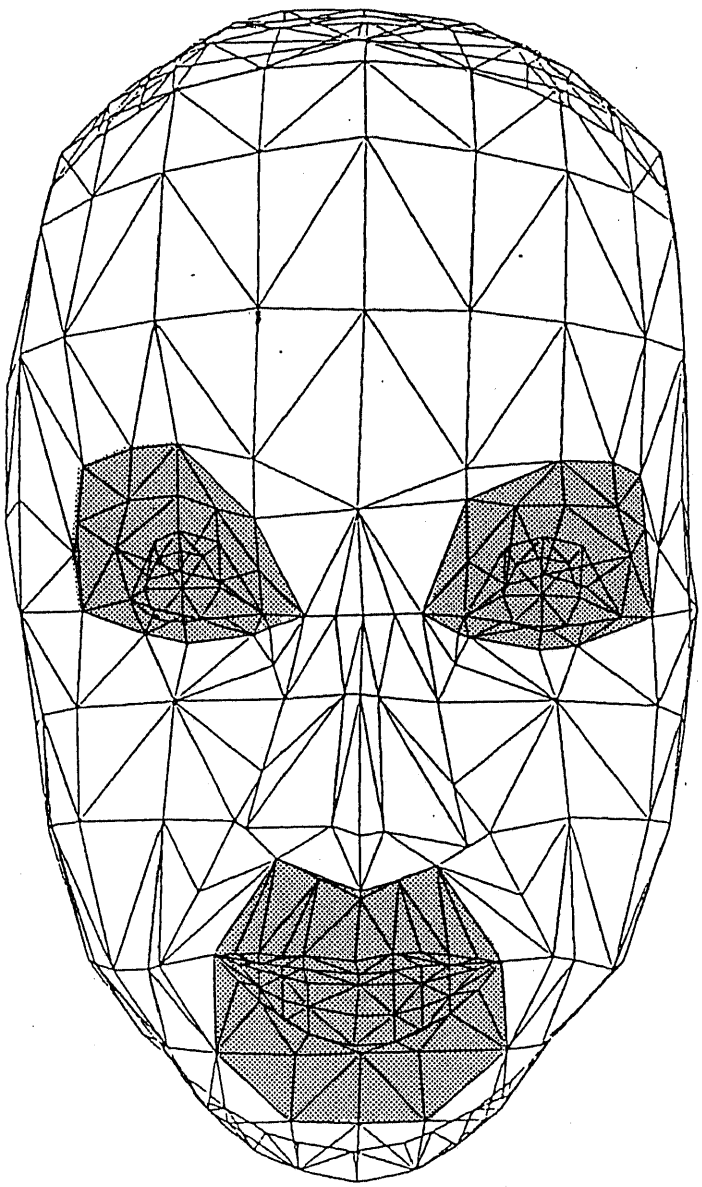


図３０．３次元人物ワイヤーフレームモデル

図３１．ワイヤーフレームモデルのレンダリング画像

6．1．2　テクスチャマッピング

テクスチャマッピングの基本的手法については、既に多く報告されており[55]、重複するので本論文では述べない。

合成の対象となる人物のテクスチャデータは、ＣＣＤカメラ・Ａ／Ｄボードを通じ、ＶＩＴに入力している。このテクスチャデータはハードディスクに保存される。人物のテクスチャの入力時のサイズは512×512ピクセルであるが、実際に合成する際は、これを圧縮することによりスムージングした、256×256ピクセルの画像を用いる。このテクスチャデータのサイズは、512×512ピクセルまで任意に変更が可能である。

6．1．3　人物モデルの表情合成

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、人物モデルに任意の表情をさせることができる。表情の設定には下記の各パラメータを要するが、本システムでは②～③の各パラメータはあらかじめ設定した値を用い、ユーザからの入力コマンドは①の部分の指定のみを行っている。

①描画・表示部分

左瞼・右瞼・左右の瞳・上下唇及び顎（口の周辺領域）から、描画・表示領域を１つまたは複数設定する。

②バーテックスの移動距離

合成画像のバーテックスの初期位置からの移動距離を設定する。ここで設定した値は、それぞれ顔の各部分において挙動の有無を示し得る程度の

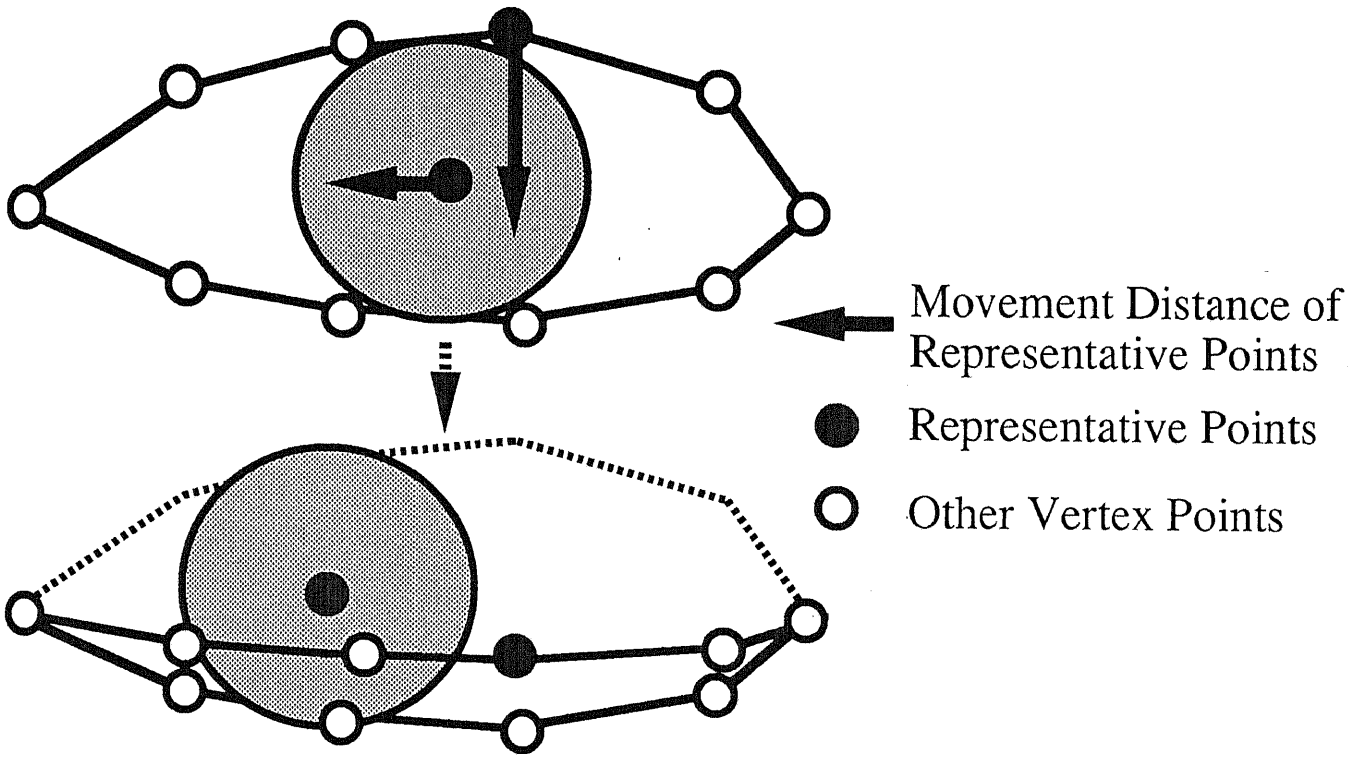


図３２．バーテックスの移動

ものとした。また設定する移動距離は、代表点の移動距離であり、関連する他のバーテックスの移動距離は自動的に変換して算出し、設定される（図３２）。

③描画ステップ

バーテックスの初期位置から移動位置までの、描画ステップ数を設定する。本システムでは、描画ステップを３に設定している。例えば瞼は３ステップ（３コマ）で完全に閉じる（開く）ことになる。

ハンドサインにより表情をコントロールする試みを行ったが、その際の指の数と顔の反応部位との対応を以下に示す。ここでコマンド入力のある場合は、どのコマンドにおいても人物モデルは手の位置をトラッキングするものとする。また、瞳は顔と連動し、手を”見つめる”動作をする。

１本：顔の特定部位の指定はない。

２本：右目でウインクをする。

３本：瞬きをさせる

４本：口をあける

５本：すべての部位を動かす。

６．１．４　バーテックスの座標値の算出と座標系の設定

本システムでは、人物像頭部の首振り動作及び表情の合成のために、系全体を記述する絶対座標系と、頭部内部における相対座標系の２つの座標系を設定した（図３３）。バーテックスの座標値の算出式及びアルゴリズムを以下に示す。

＜Ｓｔｅｐ１＞

まず、相対座標系内において入力コマンドに応じ、顔の各部分のバーテックス座標値の算出を行う。各ＶＩＴはバーテックスの最大移動距離及び描画ステップ数に基づき、逐次描画するバーテックスの位置を線形補完により算出している。

例えば、式(17)においてＶｉ，Ｖｆ，ｋをそれぞれ初期値、バーテックス移動距離、描画ステップ数とすると、バーテックスの描画位置Ｖｏは以下のようにして計算される。

$$Vo = k \cdot Vi + (1-k) \cdot Vf \qquad \cdots(17)$$

$$(0 \leqq k \leqq 1)$$

これによれば、描画ステップ数ｋを適宜定めることにより、大きく開いた口や、少し開いた口などを描画し、表示することができる。

＜Ｓｔｅｐ２＞

Ｓｔｅｐ１で求められた頭全体の座標値を、入力された手の位置を基に絶対座標系に変換する。描画は絶対座標系でのバーテックスの座標値に対して行う。

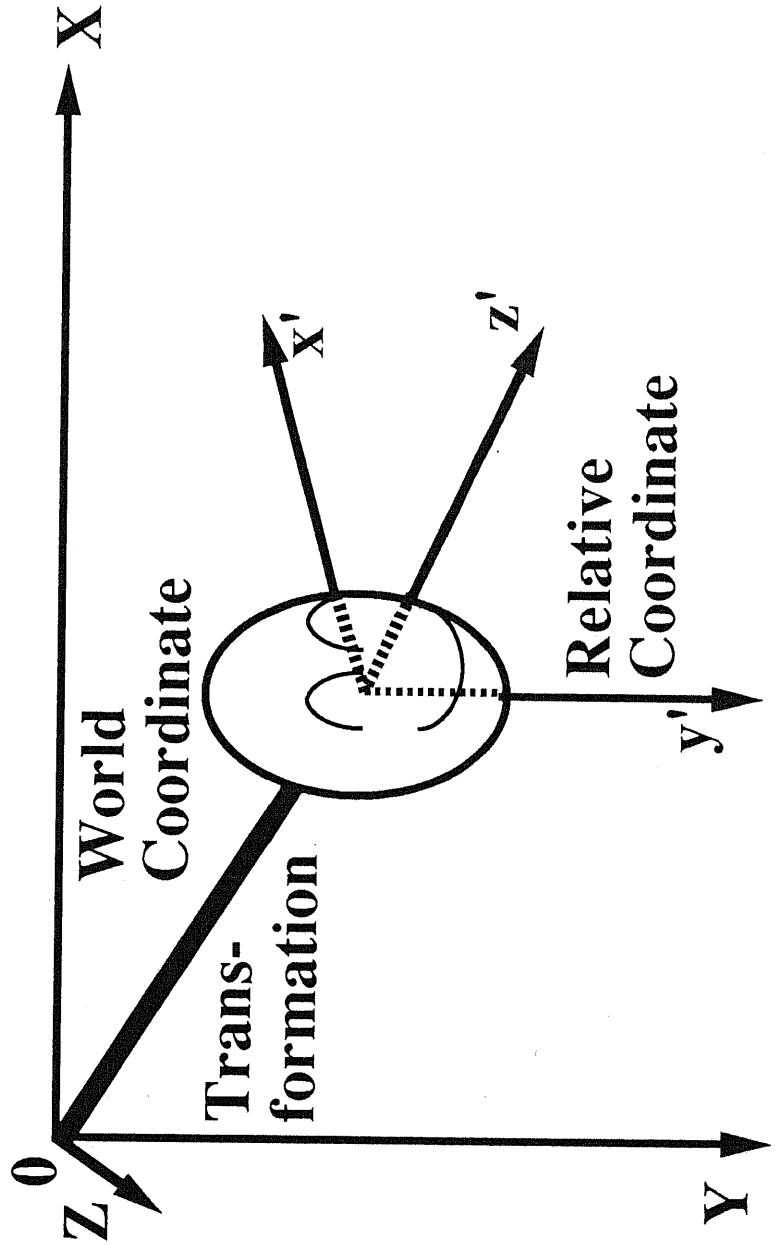

X
World
Coordinate
X'
Z'
Relative
Coordinate
y'
Trans-
formation
0
Z
Y

図３３．座標系の設定

6．1．5　動画の合成

動画の合成には下記の2種類の描画方式を用いた。

（1）時間分割型パイプライン方式

各ＶＩＴはフレームメモリを1つずつ有するため、描画中に画像を表示することができない。即ち描画中に表示すれば、描画過程そのものが表示されてしまう。そこで本手法では、描画用のＶＩＴのうちの1台を順次画像の表示用に割り当てている。

具体的には、画像表示中のＶＩＴはそれ自身の持つ画像の表示のみを行い、次のＶＩＴが描画を終えると次のＶＩＴに対し画像を表示するように命令する。表示を終えたＶＩＴは次の表示のために描画を始める。この過程を順次繰り返す。

各ＶＩＴにおける画像の表示時間は、描画スピードに合わせ、任意に設定することができる。これは毎秒の表示コマ数の設定を意味する（図３４）。

以下に本手法を採用した場合の、ホストコンピュータ（ホストとする）からＴＮ－ＶＩＴにプログラムをダウンロードしてから画像合成に至るまでのアルゴリズムを示す。

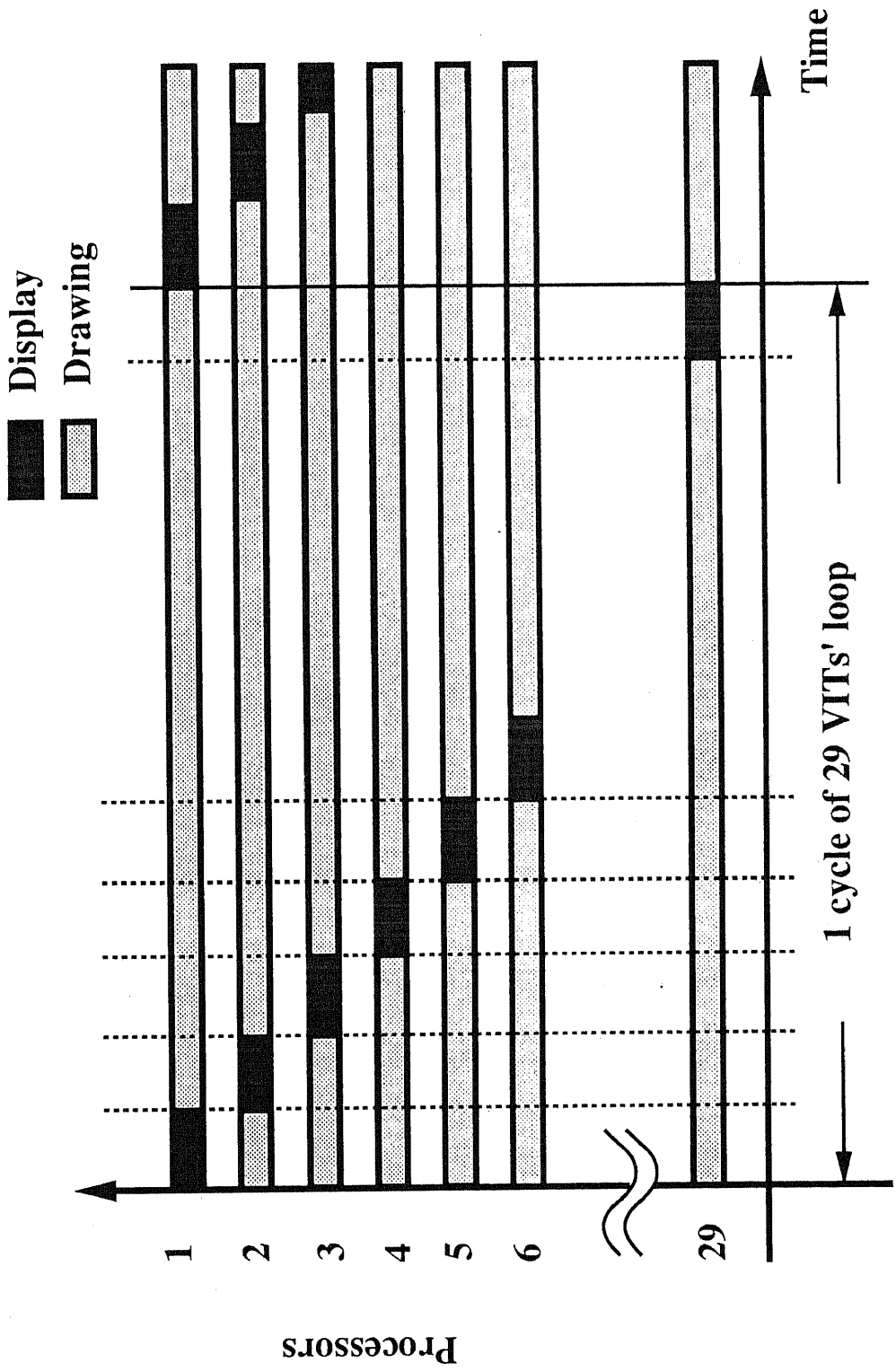


図３４．時間分割型パイプライン方式による動画の合成

①. 各ＶＩＴに番号を割付け、ＶＩＴの構成を図３５に示す構成とする。

②. ホストより1～14番のＶＩＴに人物画像のテクスチュアのデータを転送する。

③. ホストより各ＶＩＴにバーテックスの初期位置及びポリゴンのデータを転送する。

④. 1～14番の各ＶＩＴがバックグラウンド（背景）を描画する。その際、口の中のデータ（歯を含む）は原画像にないので、コンピュータグラフィクスにより背景に書き込む。各ＶＩＴは、処理の間を通じてバックグラウンドを表示し続け、描画領域の描画はバックグラウンドに直接書き込む。

⑤. 処理開始のコマンドをホストよりＴＮ－ＶＩＴに送る。

⑥. 画像認識用のＶＩＴが入力画像の認識結果を描画用のＶＩＴに転送、描画用のＶＩＴは人物モデルの頭部を描画し、これをバックグラウンドと合成した画像を表示する。

⑦. 以後1～14番のＶＩＴは、画像認識用ＶＩＴより解析結果を入力しながら順次描画を行う。

⑧. ホストからの終了コマンドにより処理を終了する。

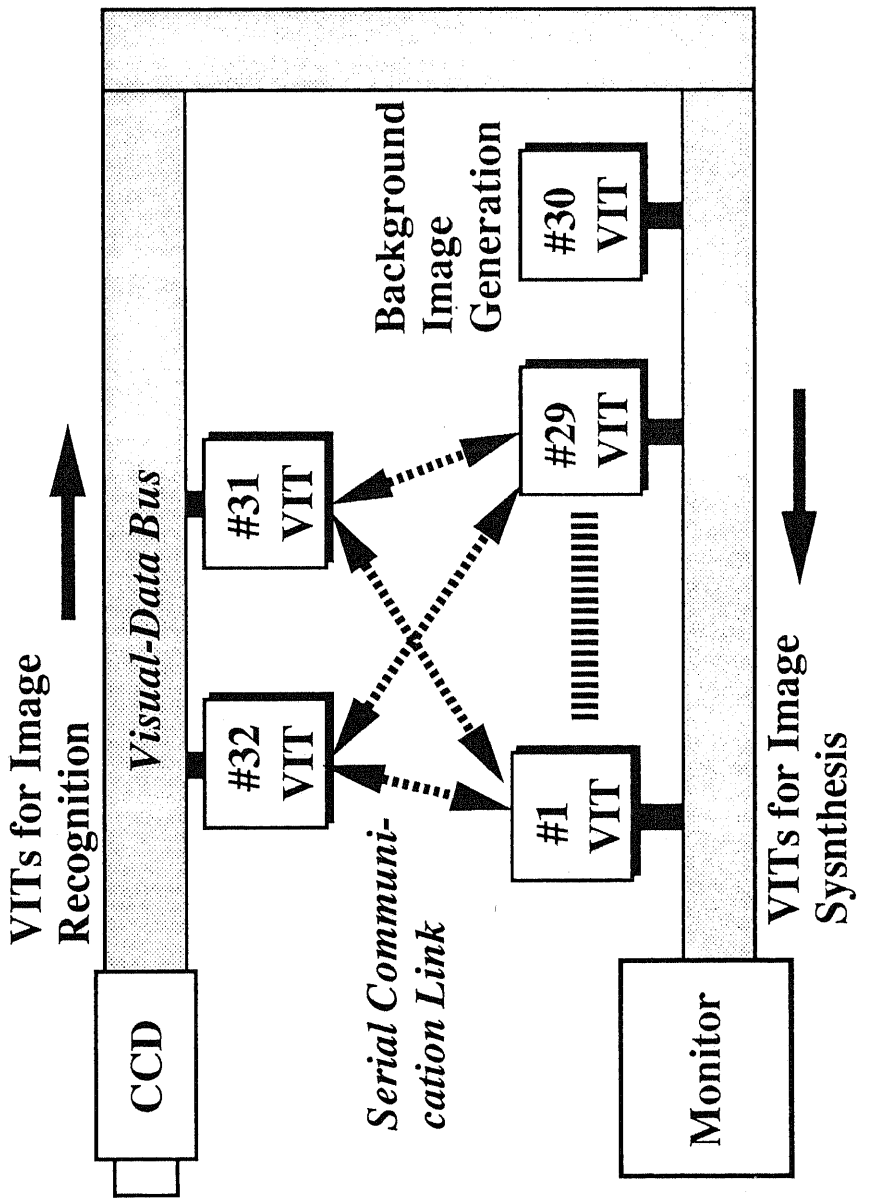


図３５．描画用ＶＩＴの構成

（２）複数台同時描画型スイッチング方式

画像合成用ＶＩＴは、ホストコンピュータからプログラムのダウンロードを受けた後、ディスクよりテクスチュアデータを入力される。描画用ＶＩＴは１７台設定している。描画された画像は、1/30秒毎にモニタに転送され、表示される。図３６にこれら１７台のＶＩＴの接続状況を示す。１７台のうちの１台は、背景の表示用である。

また図３７には、複数のＶＩＴで描画をさせた場合の、人物頭部全体の描画に要する時間の推移状況を示す。これにより、画像の入力から合成に至る処理の実行過程において、合成速度と毎秒の表示フレーム数の兼ね合い、最も効果的な台数効果の観点から、頭部の並列描画には８台を用いるのが適切と判断した。

顔画像の並列描画アルゴリズムを以下に示す。

①．顔画像は、基本的に８台のＶＩＴにより描画する。これらの８台は、それぞれバーテックスの位置の計算を行った後、516個のポリゴンを各８分の１ずつ描画する。具体的には、あるＶＩＴがNo.1のポリゴンの描画を行ったとすれば、以降は、

Ｎｏ．（ｉ＊８＋１）；　ｉ＝０　ｔｏ　（５１６／８）

...(18)

のポリゴンの描画を行う。

これにより、プロセッサ間の負荷の分散をほぼ一定にしている。

②．①の描画は、２Ｍｂｙｔｅのプログラム領域を半分に分割した領域に対して行い、描画が終了した時点で１Ｍｂｙｔｅの画像メモリにコピーしている。ＶＩＴは１フレーム分の画像メモリのみを有するため、通常の描画中にはその描画過程が表示されてしまう。

そこでこのような処理を施すことにより、描画結果表示中でも描画を

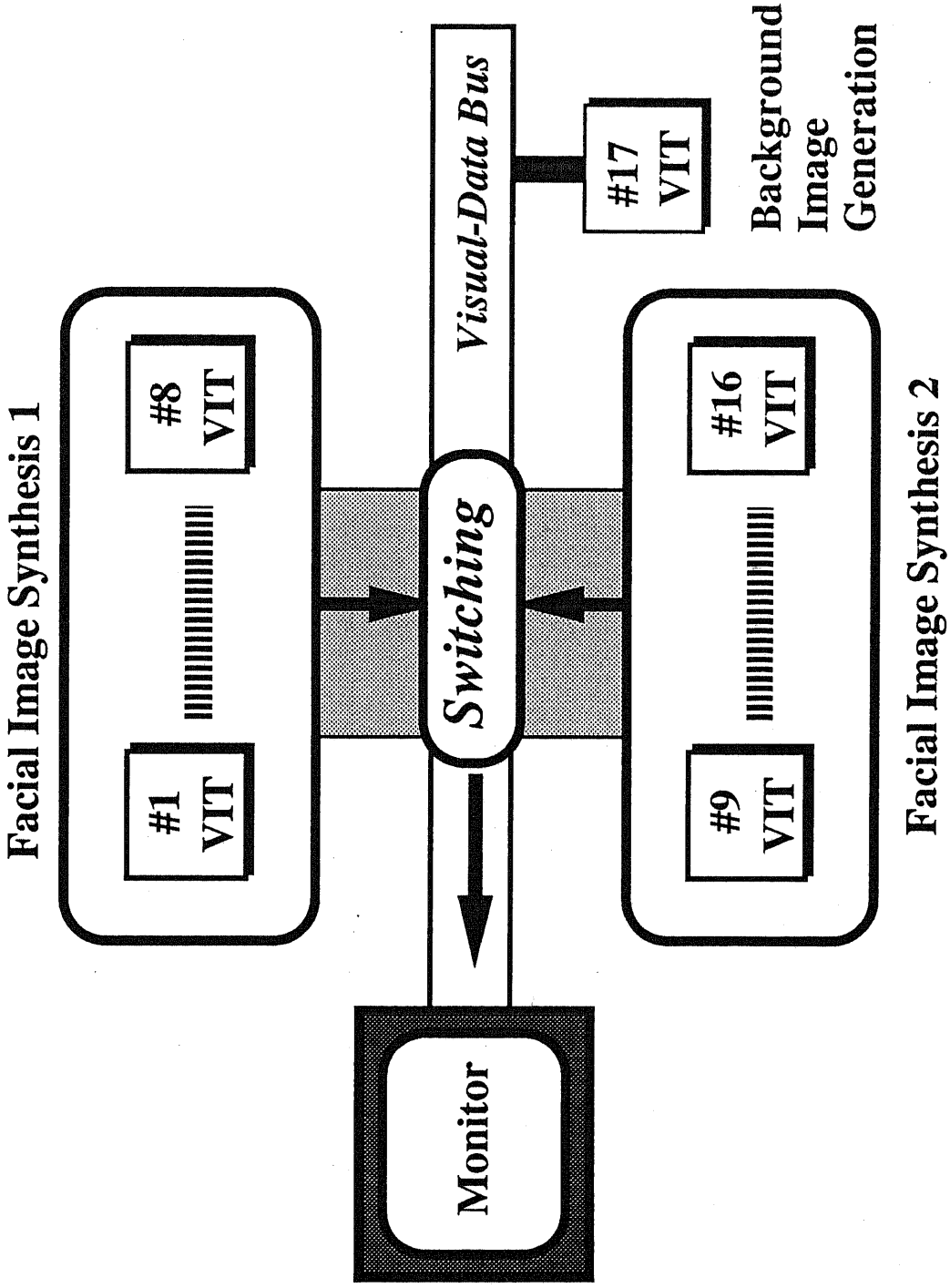


図３６．画像合成用ＶＩＴの接続状況

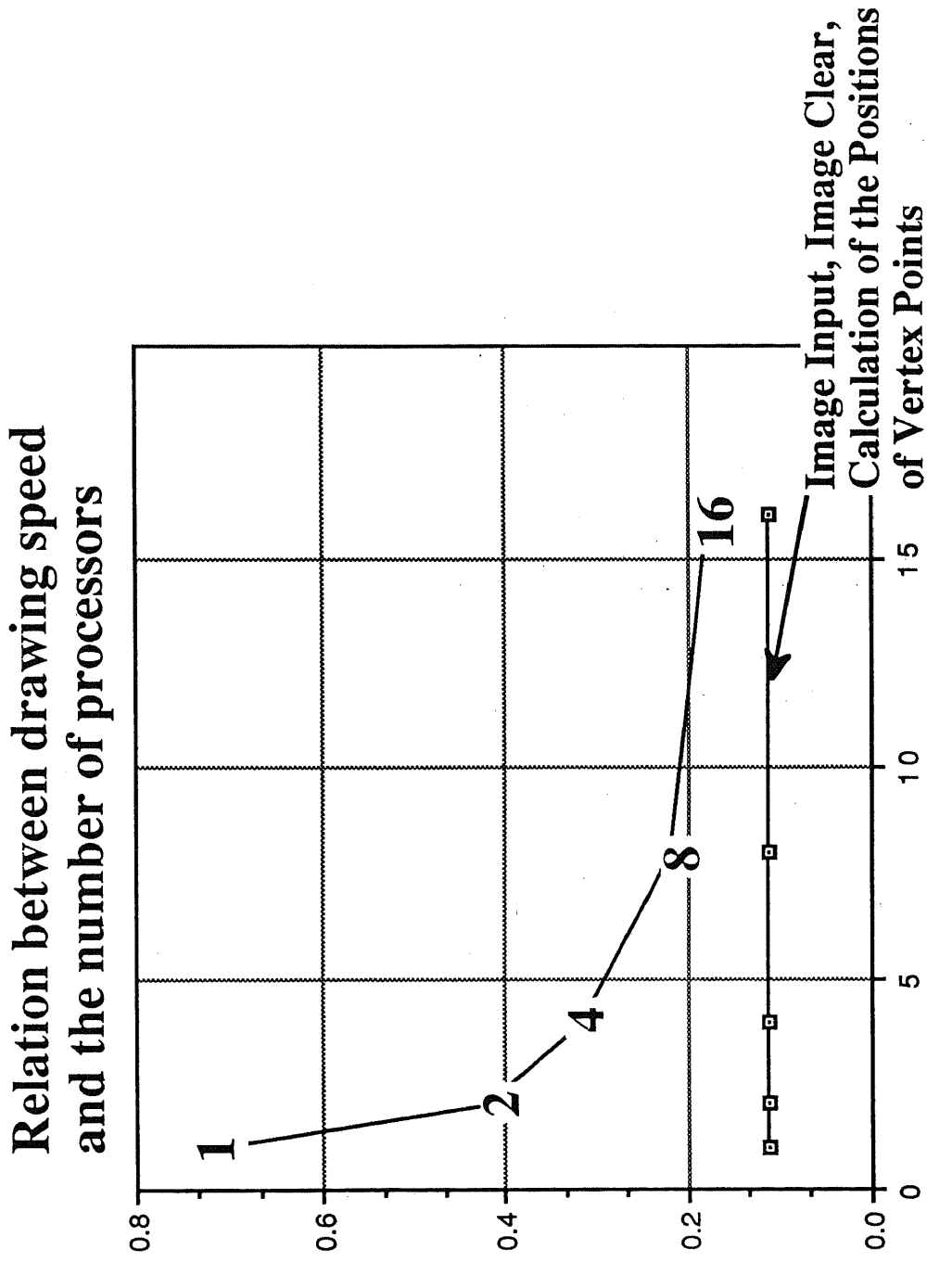


図３７．人物頭部全体の描画に要する時間とプロセッサ数の関係

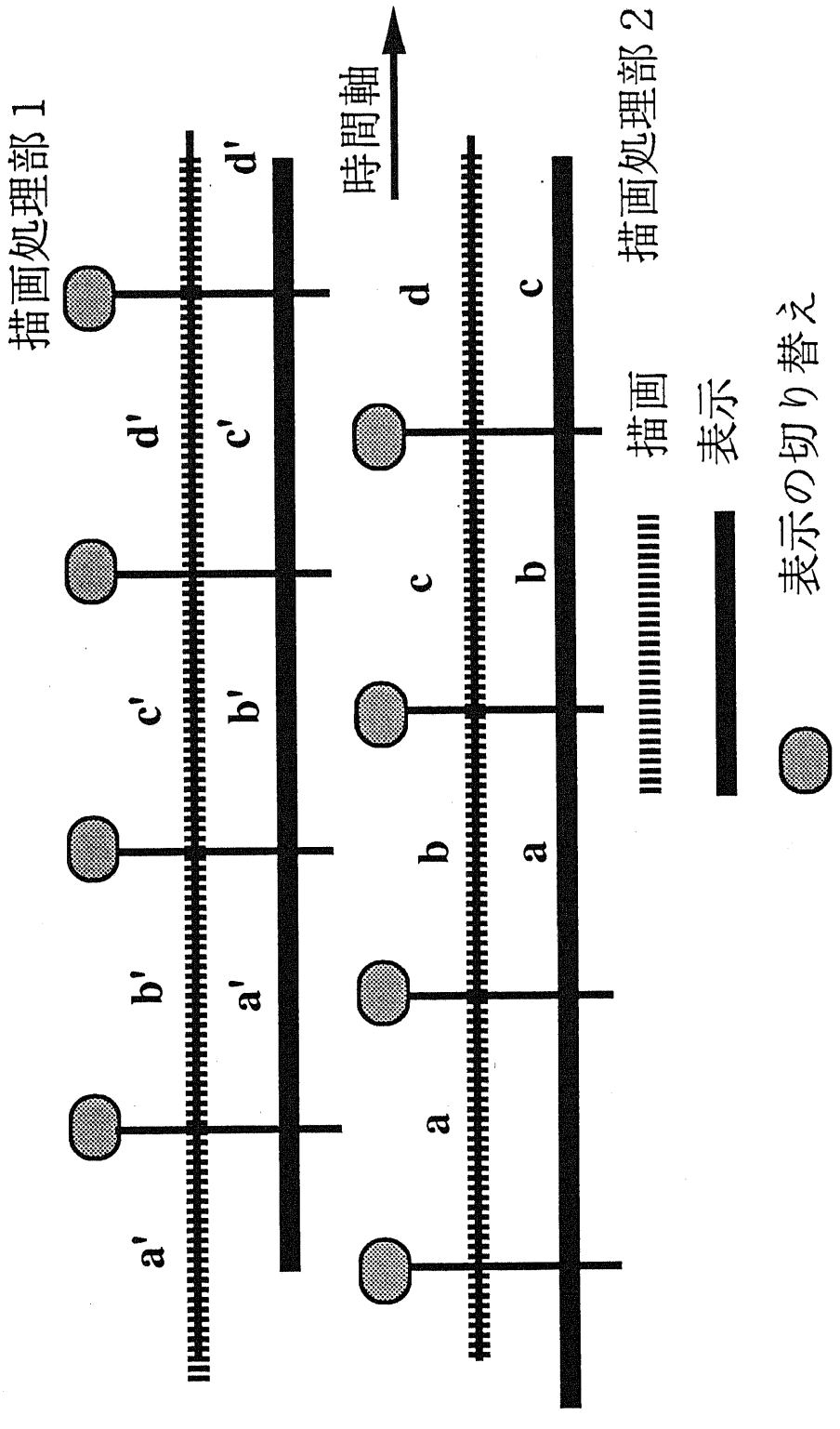


図３８．描画と表示のタイミング

可能とし、より効率的な処理方式としている。

各８台のＶＩＴによる、描画と表示のタイミングの関係を図３８に示す。

６．２　金魚型エージェントの動画像合成

本研究では、人物モデルの他にも金魚のモデルを用い人間との簡単なインタラクションも試作した。以下では、１６台のＶＩＴを用い指サイン認識をコマンド入力とした場合の金魚画像の合成について述べる。

６．２．１　金魚型エージェントモデル

金魚型ワイヤーフレームモデルは、約200個のバーテックス（頂点）よりなる約381個のポリゴン（微小３角平面）から構成されている。図３９に３次元金魚ワイヤーフレームモデルを示す。

合成時に用いるテクスチャのデータは、先の場合と同じくＣＣＤカメラ・Ａ／Ｄボードを通じ、ＶＩＴに入力している。このテクスチャデータはハードディスクに保存される。テクスチャの入力時のサイズや利用方法も先と全く等しい。

しかし金魚画像の合成に用いたテクスチャデータは、金魚のいない金魚鉢の画像と、金魚の存在する金魚鉢の画像の２種を用いた。これは前者は背景に用い、後者は金魚画像の合成用に用いることによる。

６．２．２　金魚モデルの自然な動きの表現

”仮想生物”とのインタラクションは人工現実感技術の一つの形態として関心が寄せられている。本研究においても、指サイン認識の機能をヒュ

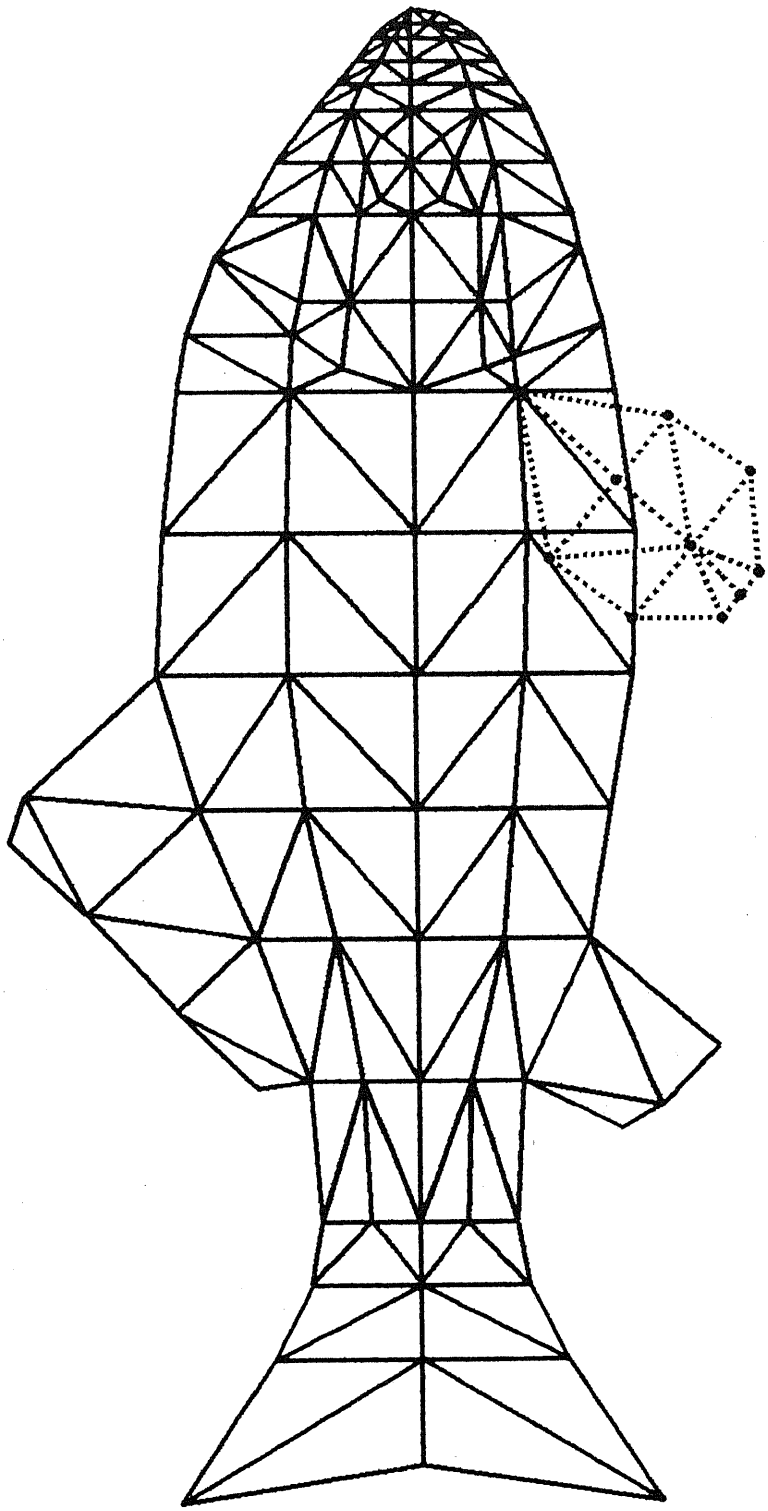


図３９．３次元金魚ワイヤフレームモデル

ーマンインタフェースとして活用する応用システムとして、並列コンピュータ上で仮想金魚との実時間インタラクションを実現した。金魚の合成像は3次元の表面モデル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ので、極めて自然感が高い像が得られている。

金魚のように柔軟な動きをする対象物の場合、モデラの各頂点位置を制御して滑らかな動きを生成するのは大変であるが、本研究では別途、対象物体のほぼ中心付近に節点を持つ仮想の骨を挿入して物体表面の各頂点をこの節点に関連付け、この骨を動作させると関連する頂点も移動するというボーン構造ソリッドモデラ[56]を利用している。

しかし本研究では実時間応答を重視したため、幾分動きのぎこちなさは許容し、これを簡略化した金魚モデルを使用した。

図40は使用した金魚の表面モデルを示している。6個の黒丸で示された頂点は、位置（動き）を直接的に指定する代表頂点を表している。体中央と口先の代表頂点で位置と向きを決定し、尾ひれ、背びれ、尻びれの各代表頂点はそれぞれのひれの動きを決定する。他の頂点はこれらの代表頂点の位置から自動的に算出するようにしている。各ひれの代表頂点は中心位置からの最大の振れ幅と、ひれを振る周期（3、5フレームなど）を指定して動作させる。

大小の金魚は関連するコマンド入力がない場合は、水槽内を円運動を主体とする自由な運動を続ける。指サインの認識により呼び寄せのコマンドを受けると、現在位置より指サインの位置へと所定の速度で向かう。動きのゴールは指サインの位置へ口先を向ける画像に対して横向きになる位置（左右の向きは運動の初期の位置から近い方）であり、コマンドの種類に応じて初期位置から位置と姿勢を線形に変えるか、あるいは初期位置からゴール位置へと垂直軸回りの所定半径の円弧状軌跡の上を、姿勢を変えな

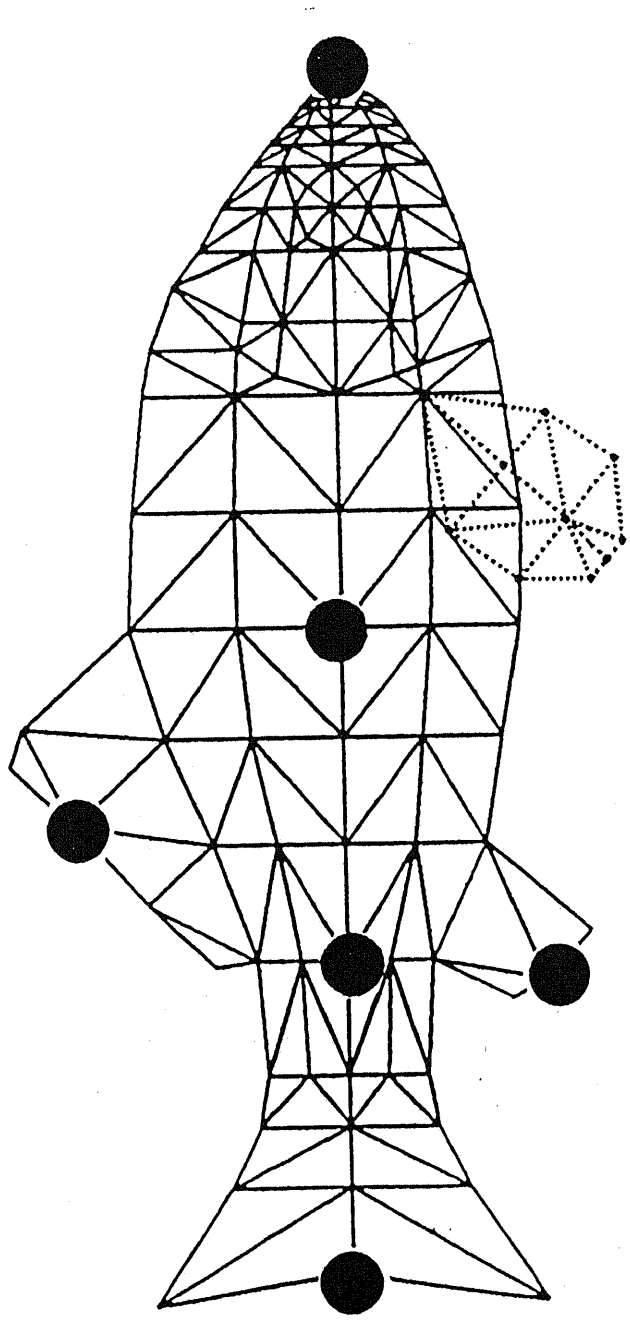


図４０．金魚の動き生成のための代表点

がら所定の速度で移動する。金魚モデルのある時点の位置と姿勢を計算し、各頂点座標も計算した後、その表面に自然の金魚画像から得た表面のテクスチャをマッピングして、金魚の表示像を生成する。

6．2．3　TN－VITの構成

図41は、大小2匹のこのような金魚との実時間インタラクション・システムの実装に使用したTN－VITのハードウェア構成と、処理機能の割り当てを示している。

テクスチャマッピングは自然感の高い合成像を得る非常に有効な方法であるが、計算コストが高い。ここでは図41に示すように、動作する大の金魚の描画に7台のVIT、小の金魚の描画に6台のVIT、そして背景（この場合は水槽内部）の描画に1台のVITを割り当てている。

テクスチャマッピングによる描画は、描画面積が大きい程計算量が大きくなることから、大の金魚の描画に当るVITの台数を多く設定した。（このような負荷の割当ての柔軟性はTN－VITの特徴の一つである。）

6．2．4　指サイン認識と入力コマンド

指サインの認識は生成画像を表示するTVモニタの上に設置したTVカメラからの入力で行う装置構成とし、金魚像との遅滞感を感じさせないインタラクションを実現した。指サインの認識率は白紙を背景にした良い環境下での測定値であり、一般の背景や人間の顔や衣服も含むような実際の環境下では正しく認識される確率は更に低くなる。しかし本システムのような応用では、幾分かの誤認識は大きな障害とはならず、簡単な指サイン認識によるヒューマンインタフェースが有効性を発揮する良い応用事例となっているといえよう。

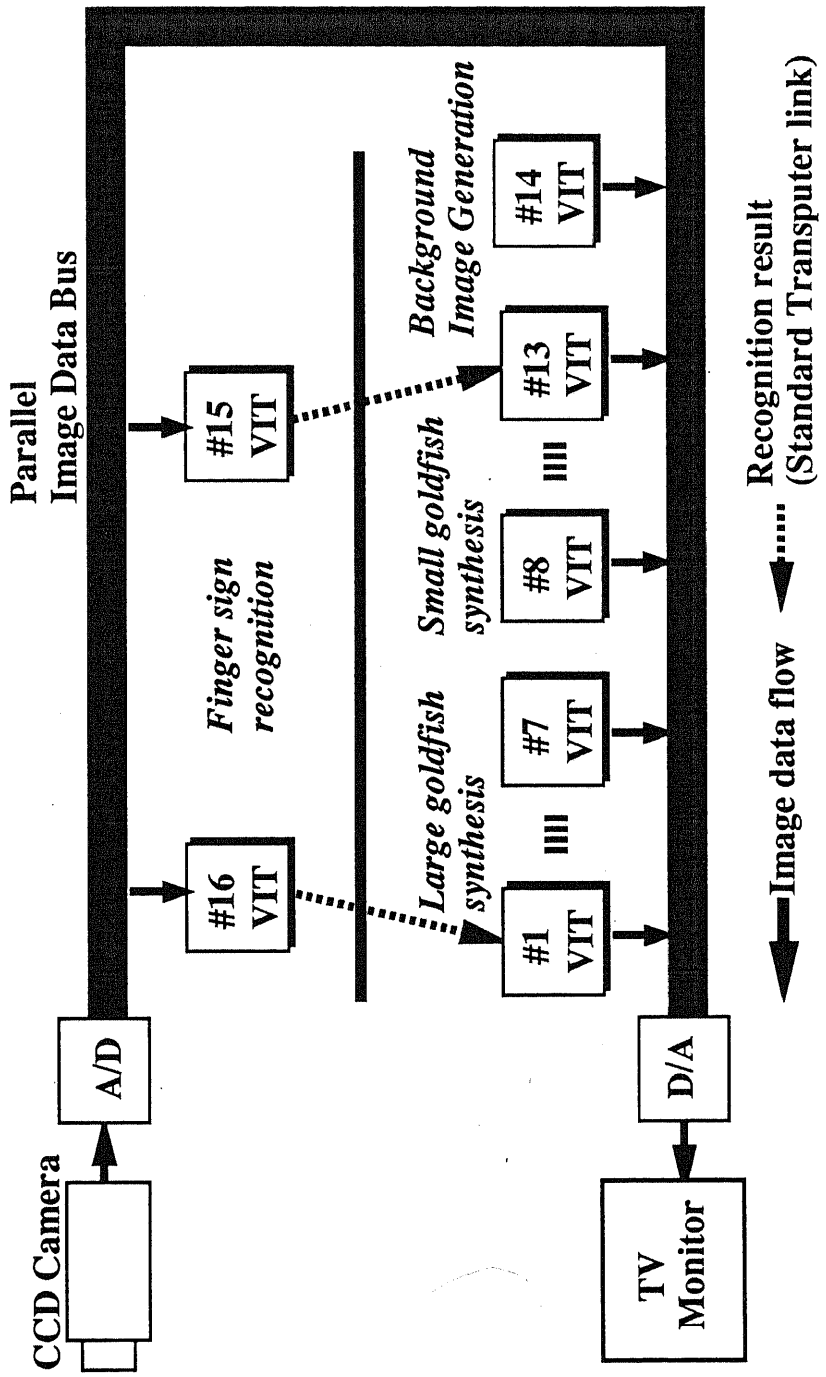


図４１．金魚画像合成のためのＴＮ－ＶＩＴの構成

指サイン認識には大の金魚用のものと、小の金魚用のものとにそれぞれ１台のＶＩＴを割り当てている。これは１台のＶＩＴでも実現可能であるが、１回の指サインの認識時間は0.5秒であることから、２台で0.25秒離れた時間の画像を入力して指サインの種別と位置を認識し、たとえ２匹の金魚を呼び寄せても手を動かしながら行えば、同じ位置へ移動しようとするといったことがないようにするためである。

立てた指の数とコマンドとの対応は次のように設定した。

１本：大の金魚を指の位置へ呼び寄せる。

２本：小の金魚を指の位置へ呼び寄せる。

３本：大・小の金魚とも指の位置へ呼び寄せる。

４本：大の金魚を垂直方向の軸回りに回転させながら呼び寄せる。

５本：大・小の金魚とも垂直方向の軸回りに回転させながら呼び寄せる。

並列処理による動画像の生成には、部分的な画像を分割して生成するなど種々の並列処理モードが可能であるが、ここでは各プロセッサにほぼ同一のプログラムを動作させることが出来る、時間分割型並列処理（異なる時点の画像を異なるプロセッサで処理・生成し、出力を順次切り替えることによって動画を生成する）を採用した。これは、

①．並列処理プログラム開発を容易にする、

②．各プロセッサの負荷が不均衡になることを避けて高い処理効率を達成できる、

③．プロセッサの台数を増やすことで処理速度を向上させるのが比較的容易という拡張性に優れている、

等の利点を有している。

６．２．５　描画プロセス

図４１に示す並列プロセッサによる描画プロセスを、大の金魚について以下に記す。大の金魚の描画には図４１に示されるように＃１から＃７の７台のＶＩＴを割当て、この７台が仮想的にリング状に接続されているとして動画を生成する。モニタに表示されているあるＶＩＴのローカルメモリ中の画像フレームメモリに描画すると、描画過程が表示されてしまうので、ある時点で表示を行っている１台のＶＩＴでは描画は行わない。表示中のＶＩＴの次のＶＩＴが描画を完了すると、次のＶＩＴに画像表示を移行させる。

これらのＶＩＴ間の通信はトランスピュータの標準通信リンクを介して行い、このような表示画像の移行を順次繰り返す。固定的な時間間隔で周期的に表示画像を切り替えていくのと比較すると、描画が完了してから表示までの待ち時間をほぼ０にすることができ、使用したプロセッサ台数の計算能力を十分に活用した単位時間当りの描画枚数の多い動画を生成できる。

運動する金魚の描画は７台のＶＩＴを一つのユニットとして次のように生成している。＃１のＶＩＴは自身の画像出力が終了すると、指サイン認識用のＶＩＴ（図４１で＃１６）と通信を行い、大の金魚が呼び寄せられているか否か、呼び寄せられている場合はその位置と呼び寄せのタイプの情報を受け取る。＃１のＶＩＴは仮想的リングの直前に当る＃７が現在描画しつつある大の金魚の代表頂点の位置情報（参照代表頂点位置情報とする）を、後述するように自らも計算して知っているとする。

垂直軸回りの円弧状軌道を通る呼び寄せのタイプの場合は、参照代表頂点位置情報から７フレーム進んだ位置と姿勢を計算してこれを運動のゴールとする。呼び寄せでない場合には、水平面内の円運動方向に７フレーム進んだ位置と姿勢を計算して、これを運動のゴールとする。

直線軌道の呼び寄せの場合は、指サイン位置と画面に対して横向きになる姿勢が運動のゴールである。＃１のＶＩＴはこの参照代表頂点位置情報と運動ゴールをトランスピュータの標準通信リンクを介して所定のタイミングで＃２へ伝え、そして＃２は＃３へ、・・・＃６は＃７へと伝達する。

各ＶＩＴは運動の速度を知っているので、受け取った参照代表頂点位置情報を起点として、＃２なら２フレーム分、＃３なら３フレーム分と、順次運動により進んだ代表頂点位置を計算し、続いて他の頂点位置も算出する。＃１のＶＩＴは描画用の１フレーム分進んだ運動位置と共に、７フレーム分進んだ位置も計算し、これを次の参照代表頂点位置情報とする。表面のテクスチャ情報はすべて描画用ＶＩＴが保有しており、金魚モデルの頂点位置が計算されるとテクスチャマッピングを行い、描画を完成させる。

以上のように＃１のＶＩＴは他の＃２～＃７のＶＩＴより幾分多くの処理を担当しているが、これはテクスチャマッピングによる描画時間と比べると極めて小さいため、感知される程の描画速度ばらつきにはならない。

小の金魚の生成も図４１の＃８～＃１３の６台のＶＩＴを使用して、同様に行われる。

以上のような方法により、ほぼ滑らかといえる毎秒１０枚の速度で動く金魚像の合成が得られている。使用するＶＩＴ台数を増やすことで、容易

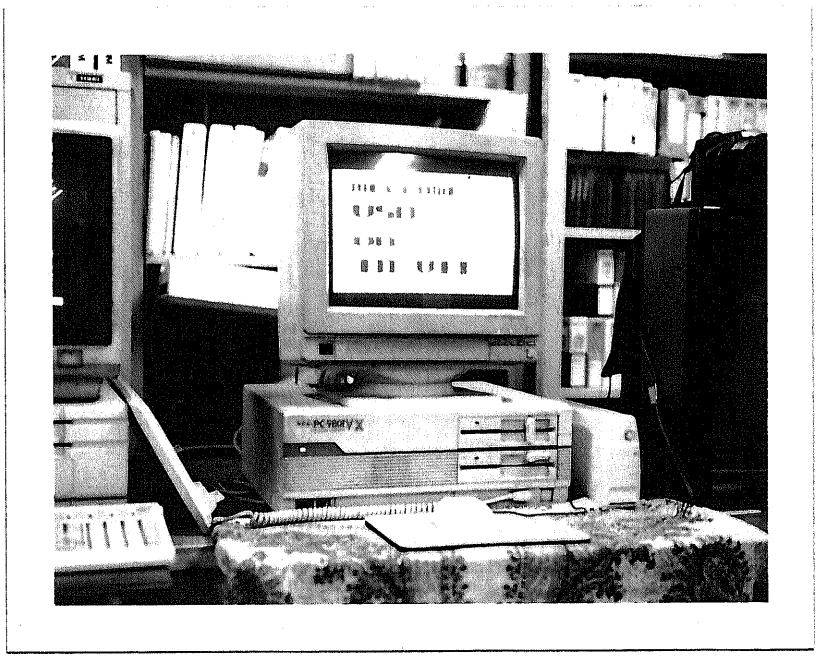


図４２．メッセージ・テキスト・図形データ表示用モニタ

により滑らかな動きを合成することが可能である。

6．3　メッセージ・テキスト・図形データ表示用モニタによる出力

現在、ＶＳＡシステムでは２台のモニタの利用が可能となっている。１台は人間型ソフトウェアロボット（エージェント）の表示用であり、他の１台はシステムからのメッセージの出力や、テキスト・図形データの表示等のために用いている。現在ハードウェア構成としてのシステムへの接続は完了して利用可能な状態にあり、幾つかの利用形態を検討した。しかし、ＶＳＡにおける最も効率的利用の詳細な検討は、今後の課題となっている。

図４２にメッセージ・テキスト・図形データ表示用モニタが稼働している様子を示す。図４２中のマウスも利用可能な状態にあり、簡易なコマンドの入力手段として適宜利用している。

6．4　6章のまとめ

本章では、システムからユーザへの情報の出力について述べた。

中でも人間型及び金魚型エージェントの動画像の合成手法について詳細に述べた。また、動画像以外のデータ表示用モニタについて簡単に触れた。

ワイヤフレームモデルとその上へのテクスチャマッピングによる画像の合成手法は、既に多くの研究で用いられており特に新規性はないが、正確を期するため、あえてＴＮ－ＶＩＴ上でのそれらの実装手法について述べた。

中でも本システムにおける動画の合成アルゴリズムは、ハンドサインや

ユーザの挙動に対するソフトウェアロボットの反応の様子や、動きのリアルさ、反応開始までのタイムラグ、などを決定付ける重要な要因であるので詳細に述べた。

人物画像の合成において、本研究では喜怒哀楽のような感情の表現は対象としなかったが、人間型ソフトウェアロボットの豊かな表情はインタフェースの構成上不可欠であるため、今後必要になるものと考えられる。

第７章　人間型エージェントの自然な挙動の検討

７．１　背景

ＶＳＡではユーザは基本的に人間型ソフトウェアロボット（エージェント）とコミュニケートして処理を進め、ユーザにコンピュータの存在を感じさせないインタフェース環境を提供することを目標としている。その実現のためには、ユーザとエージェントとの間に人間どうしの自然なコミュニケーションをシミュレートすることが強く求められる。すなわちユーザに対し、いかに自然感の高い人間型エージェントを構築するか、言い替えればいかにエージェントが実在の人物であるかのような印象を与えるかが問題であり、重要な課題となっている。

そこで本研究では、自然感の高い人間型エージェントを構築のために、まず３次元ワイヤフレームモデルにテクスチャマッピングを施し、これを実時間で動作させた。これにより通常のレンダリングなどにより合成された画像に比べ、より自然感の高い画像とした。

次にユーザからの入力に対し、いかに人間らしい挙動で反応させるかを

検討・実装した。例えば、人間どうしのコミュニケーションでも重要な役割を果たす「視線」の自然な動きの検討は、有益であると考えられる。

ところで最近の研究成果によれば、人間の無意識の視線の挙動は視覚機能そのものの持つ特性による部分が大きいとされており[3,61-65]、それらの成果を調査し、ＶＳＡに取り入れることを検討した。

本章ではこのような背景に基づき、人間の視覚機能に関する最近の研究成果について述べ、またそれらのＶＳＡにおける活用形態の検討について述べる。

７．２　人間の視覚機能に関する最近の研究

近年の生理学における視覚神経系研究の進展は、各種測定・計測技術の進歩等に支えられ、目ざましいものがある。また心理物理学・認知心理学は、生体の高次視覚機能を徐々に解明しつつある[57,58]。

本節では、生体の視覚に関する最近の研究成果を、生理学（視覚のハードウェア）、及び認知心理学・心理物理学（ソフトウェア）の両面から概説する。

７．２．１　視覚のハードウェア：網膜から脳へ

①．網膜神経細胞：

人間を含む脊椎動物の網膜は、神経回路網の構造が基本的に相似しており、大別して5種の網膜神経細胞から成り立っている。それらは層状に積み重なり、3次元集積回路を形成している[59]。

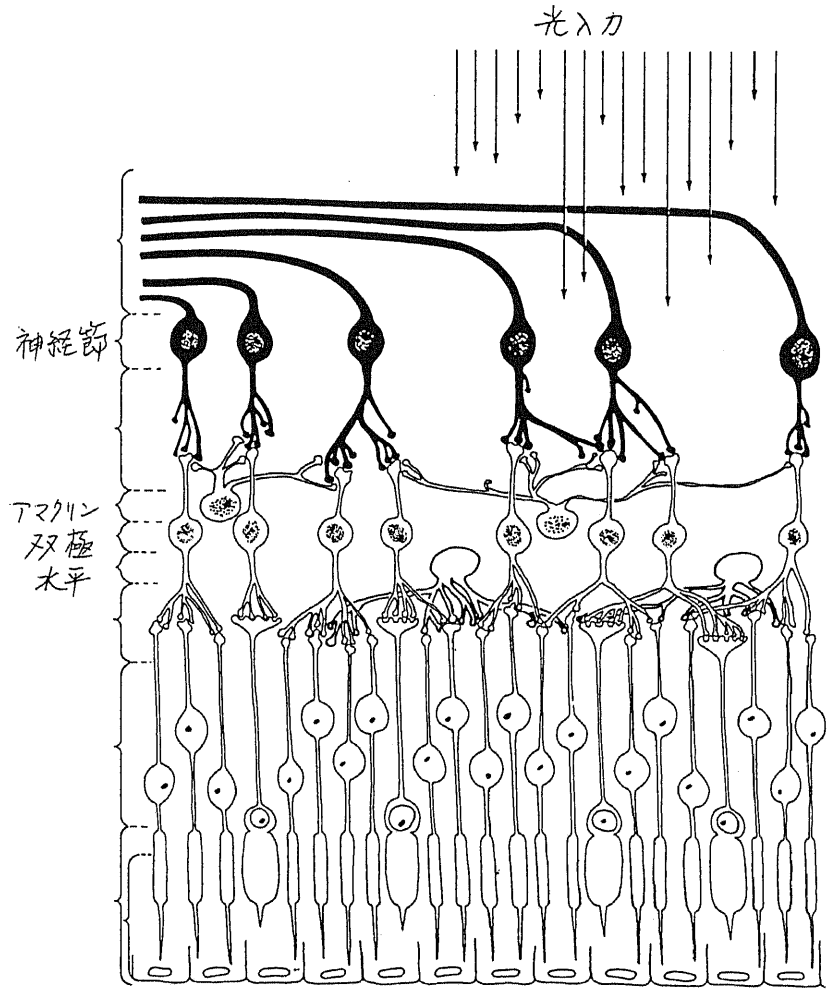


図４３．網膜神経細胞

まず視細胞が光に反応して電気信号を発生する（図４３）[60]。視細胞のうち、錐体は色に反応するが反応速度が遅く、桿体は色に反応しないが光強度に敏感で高速に反応する。この網膜上点領域の光情報は、次に水平細胞に渡される。水平細胞では、文字どおり視細胞の出力の水平処理を行っており、特に錐体からの出力に対しフィードバックをかけ、解像度の改善を行っていると考えられる。双極細胞では、複数の視細胞の出力を統合し、受容野を形成する。受容野は大別してオン中心型とオフ中心型があり、ここではじめて顕著な中心－周辺機構が見られる。受容野はこの後、脳内での視覚情報の表現に至るまで見られるようになる。

双極細胞の出力は、アマクリン細胞を経て神経節細胞に伝達される。ア

マクリン細胞は２次的な水平処理細胞であり、双極細胞の出力の時間的変化のみに反応し、動き検出細胞といわれる。

ここまでの処理は、いわば初期視覚情報処理であるが、全てアナログ処理により行われている。神経節細胞は、以上のように伝えられてきた視覚情報を、スパイク放電の形で視神経繊維に対して出力する。即ちこれは、それまでの視信号のアナログ処理結果を脳の奥に伝えるためにデジタル化する、A/Dコンバータの役割を果たしていると考えられる。

話が前後するが、受容野の感度分布は、一般にガウス関数の差、即ちDOG(Difference of Gaussian)関数で記述できることが知られている。D.Marrらは、ある種の細胞の受容野がラプラシアン・ガウシアンで近似できることを示した。D.Marrは自ら数学者・生理学者でもあったが、このように生理学的研究から明らかにされた情報処理の原理を、工学的に応用されるレベルにまで近づけ、計算論的視覚研究の基礎を築いた。

②. 網膜：

網膜は厚さ200-300ミクロン，面積が数㎠の透明で柔らかい膜状をした神経組織であり、先に述べた網膜神経細胞の総数は数億個と見積もられている。網膜での5種の細胞の接続関係は、網膜上の位置により異なっている。例えば網膜周辺部では、1個の神経節細胞は数千個の桿体から情報を受け取っている。これに対し、網膜の中心の半径200μmの領域は中心窩と呼ばれ（錐体のみが集中している部分。人間が注視する際はここで情報を収集している）、１個の錐体の出力は１個の双極細胞を経由して１個の神経節細胞に直接結合している。このような結合関係の相違は、膨大な視覚情報からの有意な情報の選別に重要な役割を果たしている。最近のCDDは、厚さはほぼ等しいが、集積度は民需製品レベルで数センチ角に数10万のオーダで

あり、集積度だけで3桁及ばない。また網膜では、以上述べたように場所によって５層または３層と異なる３次元構造をなしており、局所的に情報の統合やフィードバックが行われている。さらに、網膜では信号は全てアナログで処理されるが、出力はデジタルである。

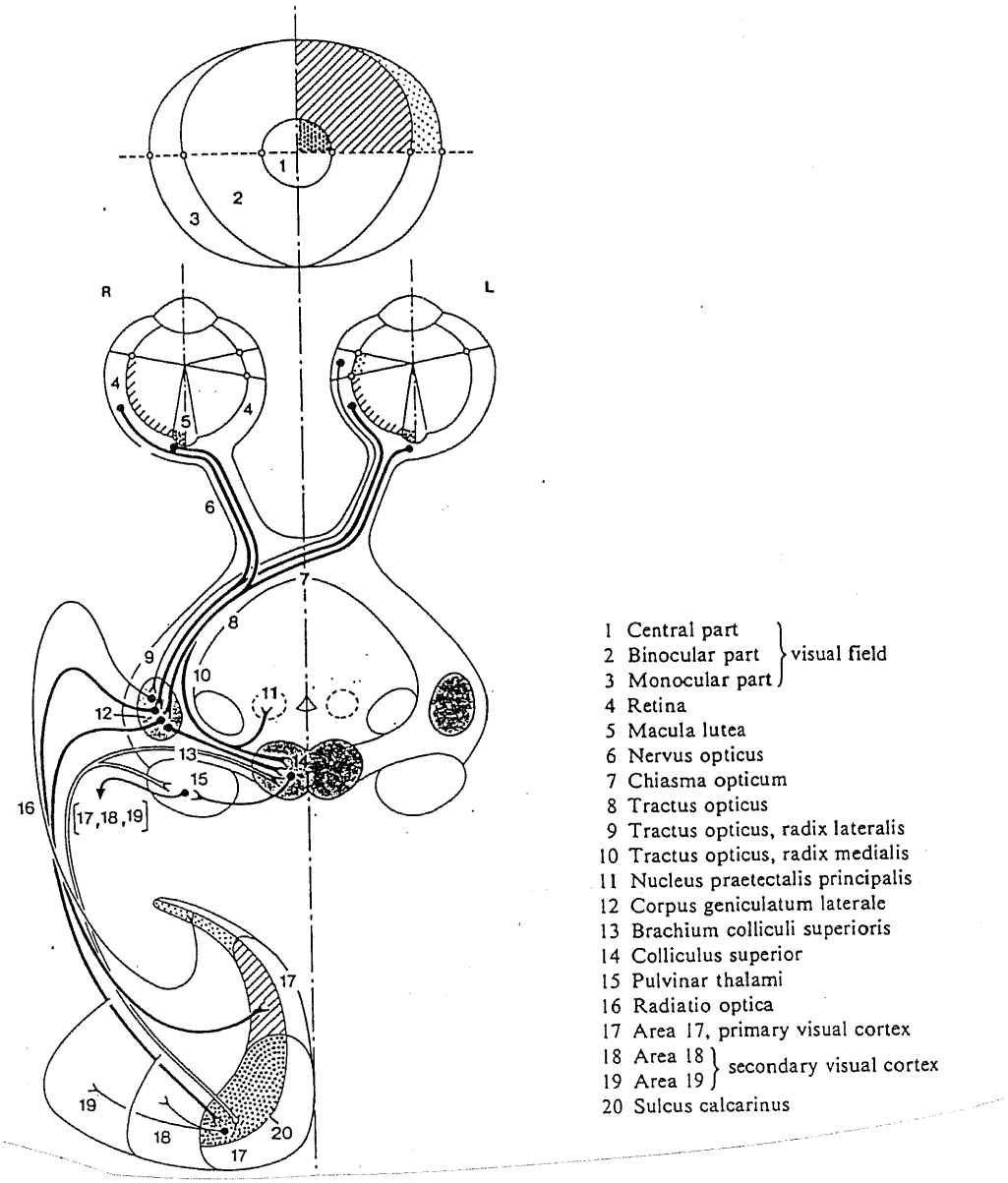


図４４．網膜－外側膝状体－大脳皮質系

③．脳への信号の伝搬：

図４４に、視神経繊維から出力されたスパイク放電が、脳に伝達されまでの経路を示す。図４４において、網膜(4)から出力された電気信号は、視

交叉(7)で半交叉し、外側膝（しつ）状体(12)を経て、大脳皮質系(17,18,19)に伝搬される。ここで一部の出力は、眼球の運動制御、瞳孔の制御のために、別経路(11,14)でフィードバックされる。

これらの信号制御は並列に行われており、例えば「瞬き」の最中は、大脳の知覚感受性全般のゲインを低下させていることが知られている。また「盲点」部の信号の欠落については、脳内で穴埋めを行っているものと考えられる。左右両側の網膜の右側半分（視野の左半分）の情報は、全て右側の後頭葉に伝達され、左側半分は左側の後頭葉に伝達される。

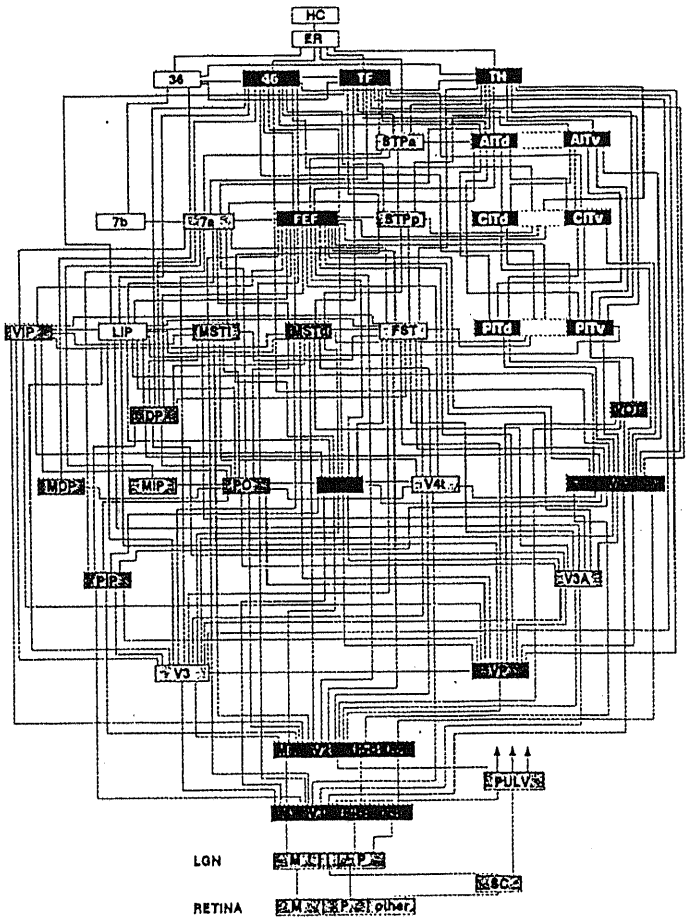
外側膝状体は左右に１つずつ存在するが、両眼の視神経繊維出力を交互に層状に入力しており、単なる中継地点ではなく、両眼立体視のための初期処理や、色と空間情報の処理・またおおまかな形状のコントラストの符号化などに関与しているものと考えられている。

④．大脳視覚領：

視覚領の解明は、主としてマカクサルを用いて進められている。マカクサルの視覚領は、霊長類の中でも、あらゆる点で人間の視覚領の構造に近いとされている。最新の報告によれば、マカクサルの視覚領は32の領域から構成され、それらは305の経路を通じて、階層的に接続されている[31]。３２の領域のうち、直接的に視覚認識にかかわるのは25の領域であり、他の7領域は視覚による運動の制御等に用いられている。

図４５において、下段のRETINAは網膜、LGNは外側膝状体を示している。外側膝状体を出た信号は、まずV1野（視覚1次野，17野とも呼ばれる）に至る。V1野と、V1野を包むV2野では、方位・波長・運動方向・両眼視差に選択的に応じる細胞が存在する。中でもV2野においては、両眼性細胞が極めて多く存在し、立体視に大きく関与していることが示唆されている。ここ

では省略するが、この他にも、これらの出力を統合し、より高次な処理（中期視覚と呼ばれる）を行う各種領域が確認され、それらの機能の解明が急ピッチで進められている。これらは、いわば並列処理マシンの各要素プロセッサの機能と、プロセッサ間の結合状態の解析に相当するといえよう。



A hierarchy of visual areas in the macaque, based on laminar patterns of anatomical connections. About 90% of the known pathways are consistent with this hierarchical scheme; the exceptions may reflect either inaccuracies in the reported anatomical data or genuine deviations from a rigid hierarchical scheme. [Modified, with permission, from (*1*), with subcortical connections based on (*37*)]

図４５．マカクサルの視覚領地図[31]

７．２．２　視覚のソフトウェア

（１）　視覚機能の基本要素

視覚の心理学における研究の歴史は古く、15世紀のLeonardo da Vinciによる線遠近法や対比の研究に端を発するといわれ、その後は生理学・解剖学などの研究成果をも取り入れつつ進展してきた。今日では、人間の視覚系の初期段階（初期視覚）では、明るさ・色・運動・奥行き・テクスチャ

の５つの特徴が独立に抽出され、認識に至る過程で統合（特徴抽出と統合，図４６）されるとする見解がほぼ定説となっている。これは生理学の側面からも確認されつつある。

そして現在、最も研究者の間で熱心に研究・議論されているのは、特徴の詳細な統合のプロセスについてである。２節で見てきたような生理学的知見は、各特徴抽出モジュールの”結線状態”を示してくれるが、現状ではどのようなタイミング・加重で情報を統合しているのかという点で、十分に情報を与えてはくれない。

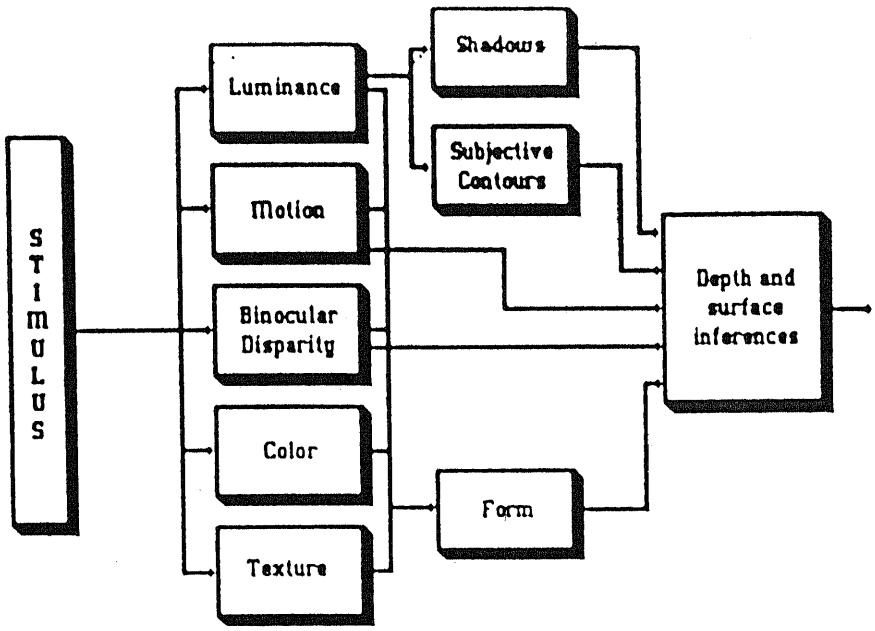


図４６．初期視覚の５つの基本機能と統合

（２）　視覚の特性

①．視覚探索とは：

少々耳慣れない研究分野に、「視覚探索」という分野がある。視覚探索

では、様々な視覚探索課題を被験者に与えることにより、詳細な特徴抽出機能の解明と特徴統合過程を明かにすることを試みている。

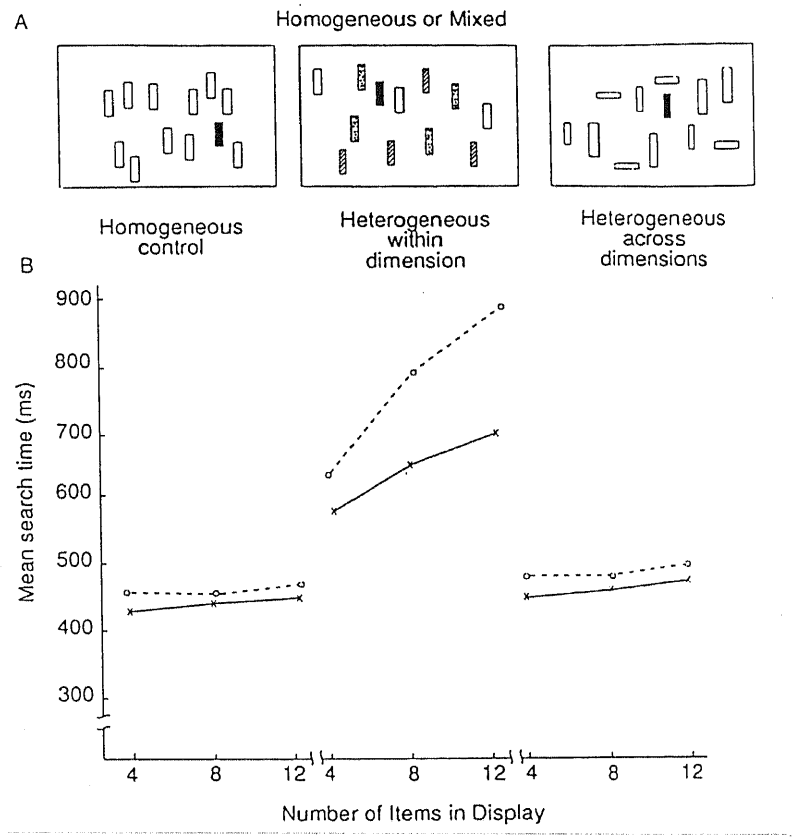


図４７．視覚探索課題と探索時間の例

図４７は、視覚探索課題の一例である。Ａが課題、Ｂは課題を実行した際に被験者が要した時間を示している。Ａの左図では、同一のものに一つだけ課題が存在している。中央は、左図の四角形を様々なテクスチャで塗り、右図は対象の形と向きを様々に変えてある。Ｂの結果を見ると、２つの特徴に差異を有する右図の方が、中央より探索時間が短い。図４８はこの現象を明快に説明している。図４８の下図に示されるように、異なる特徴が多い程、対象物の探索が容易になることが示唆される。

図４７中Ａの左図のように、妨害刺激の個数にかかわらず、等しく素早く目標を検出可能な現象を「ポップアウト(Popout)」と呼ぶ。ここでは、探索目標のみ色が異なるので、ポップアウトが起きている。

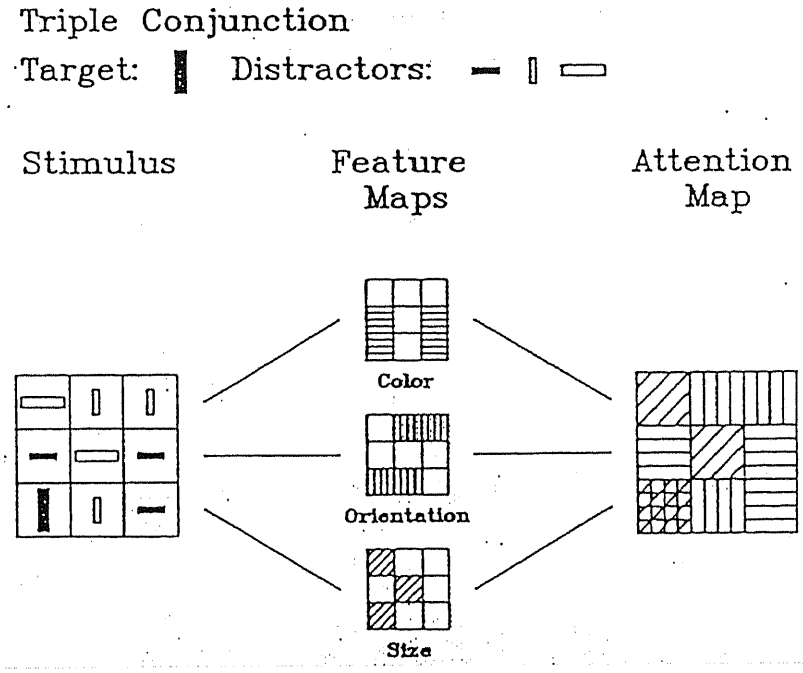


図４８．探索時間の短縮が見られる理由

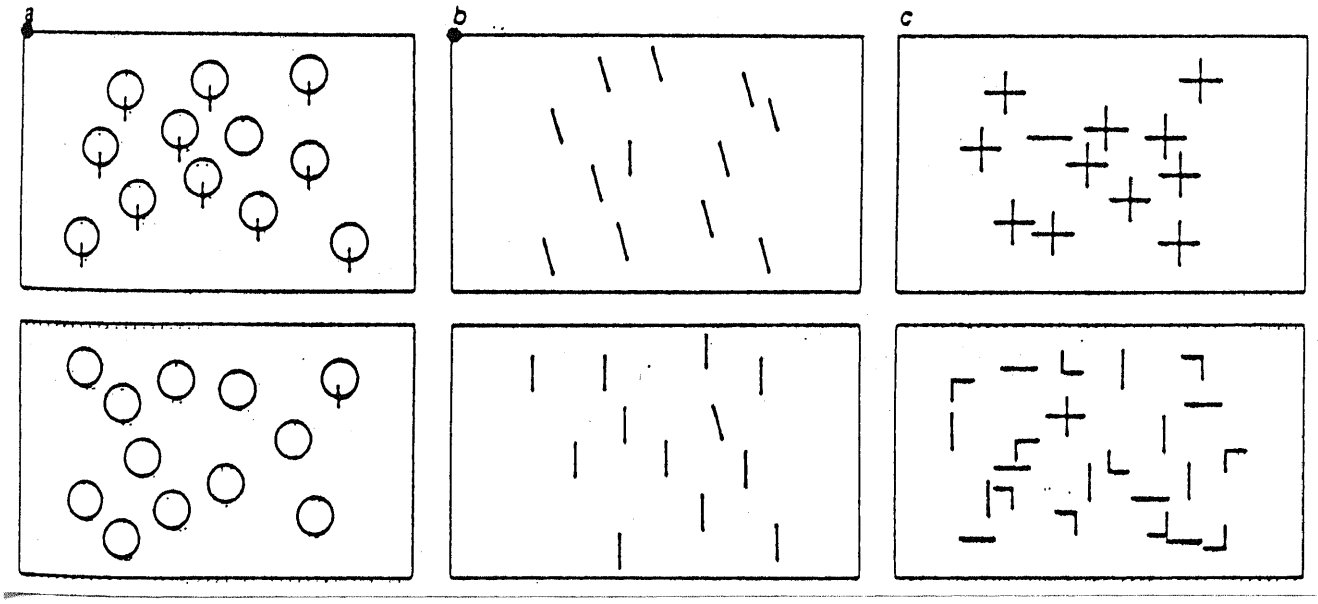


図４９．探索非対象性

②．探索非対象性：

図４９において下段から唯一他と異なる対象を探すのは容易であるが、

上段ではそうは簡単ではない。即ち、上段ではポップアウトは起こらない。

これを「探索非対象性」という。この非対象性は、標準的特徴（縦線，横線，真円など）から逸脱した属性が探索目標に付加されている場合にはポップアウトが起こるが、そうでない場合には起こらないと説明される。人間の形状認識機能の一端を示す例として、興味深い。

③．知覚的群化：

図５０に、知覚の群化の例を示す。しばらく眺めていると、上段の円に２つのグループが存在することがはっきりと「知覚」でき、一度知覚が完了すると２度目からは一目でグループ分けができるようになる。しかし下段では、何度やっても一目でグループ分けすることは困難である。

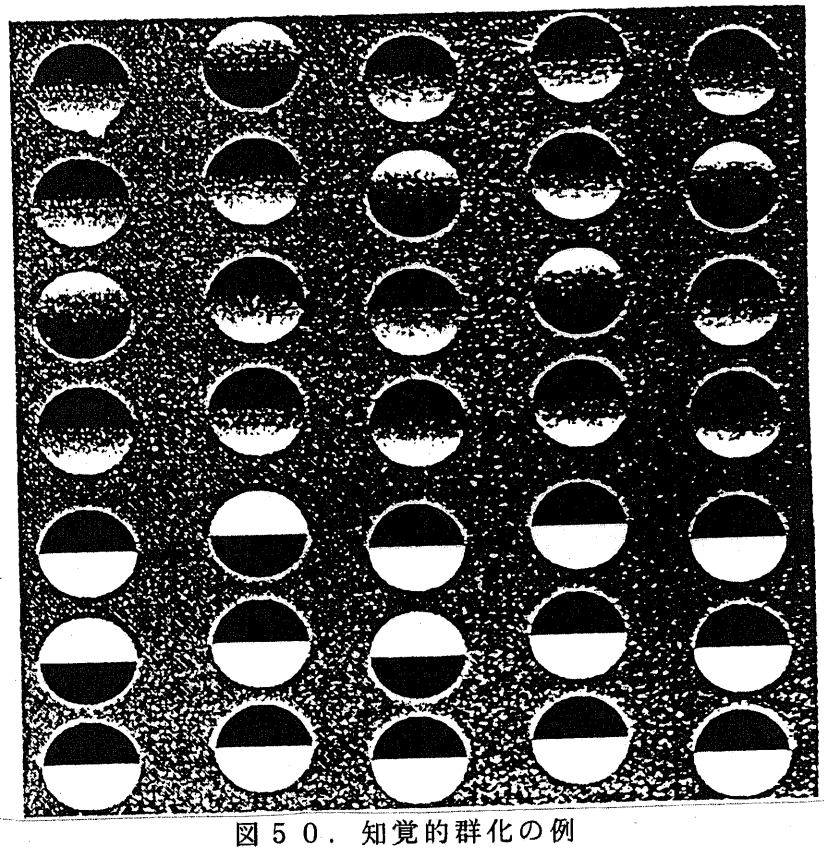
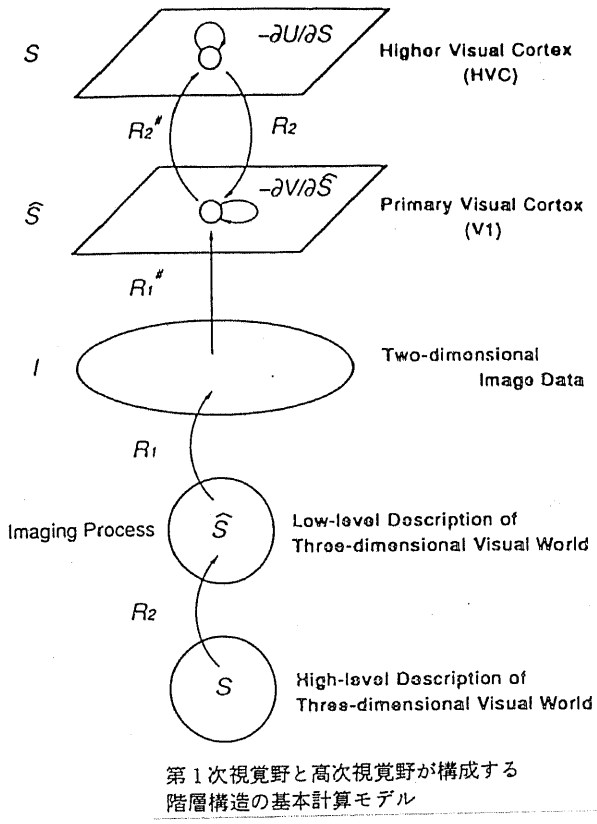


図５０．知覚的群化の例

このように、一目で対象のグループ分けが行われることを、「群化」と呼ぶ。これは、疑似的に陰影をつけて作り出された３次元形状でさえ、群化を促進する特徴であることを示している。他の群化促進の特徴としては、

視覚探索の場合と同様に、明るさ・色・運動方向などが考えられる。

ここで、ポップアウトが特定対象物を一目で探索する能力であるのに対し、群化は多数の対象物を一目でグループ分けする能力であると言える。



第1次視覚野と高次視覚野が構成する
階層構造の基本計算モデル

図５１．川人らの計算論的視覚認識モデル

７．２．３　Marr以後の視覚の計算論モデル

以上で述べてきたような、生理学・心理物理学の研究成果を取り入れつつ、Marr以後の計算理論をさらに進めようとする研究がある。

ATRの川人氏らのグループは、網膜神経細胞に視覚刺激が入力されてから高次認識処理に至るまでの脳の計算理論モデルを提案している（図５１）[66]。氏らの提案した「大脳皮質の層構造を考慮した神経回路モデル」は、現在はダイナミックな初期視覚特徴の統合過程をも記述するものではないが、将来の理論の展開次第では、それらが記述可能となる可能性があると示唆している。

7．2．4　「視覚」に関連する工学的試み

（1）　The MIT Vision Machine[67]

1988年の米国国防総省・防衛高等研究計画局（DARPA)のImage Understanding Workshopにおいて、人間の初期視覚機能をモデルとした、機械の視覚を実現するとする論文が発表された。著者はMIT AI研究所のPoggio教授らのグループである。教授らは、生理学・心理物理学等の知見に基づき、環境の変化に対して頑健性を有し、多目的に使用可能な柔軟な機械の視覚を、コネクションマシン上に実現すると発表した。図５２は当時のProceedingsの表紙に採用された、同論文中の図である。

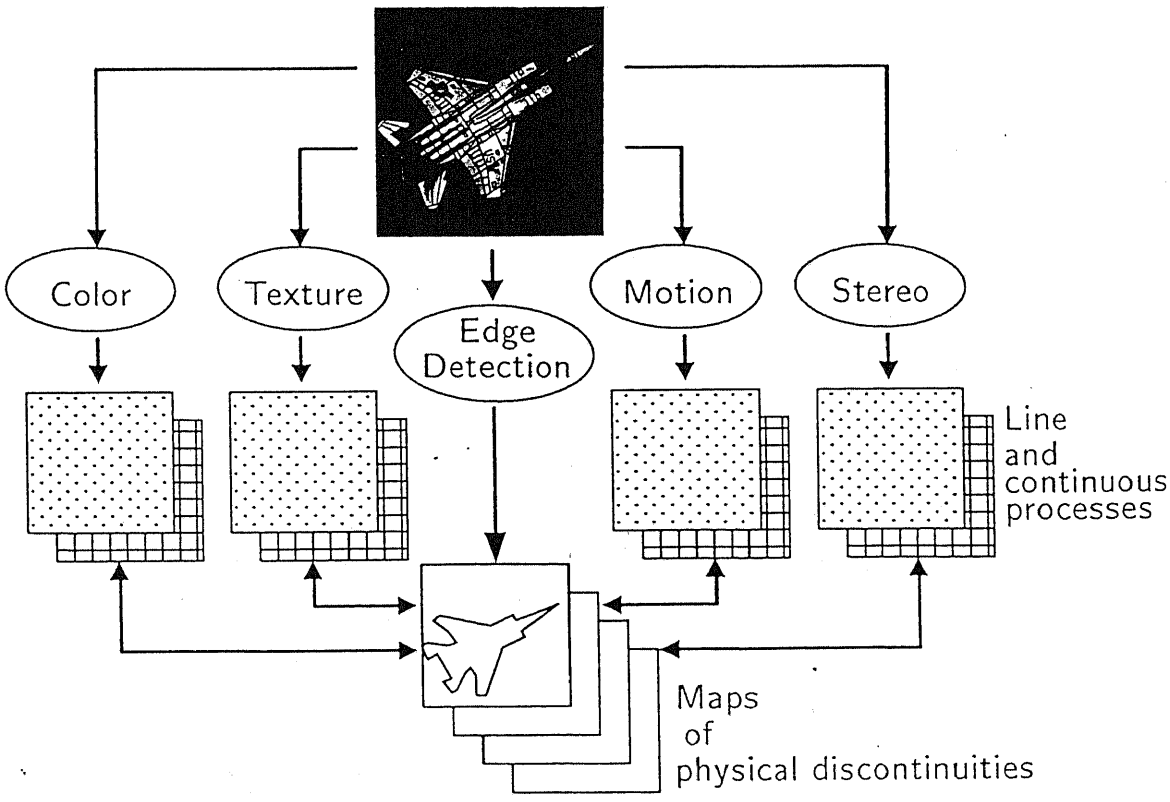


図５２．The MIT Vision Machine

図５３にVision Machine（以下VM）による画像の処理結果を示す。VMにおいて最大の問題は、各特徴抽出モジュールの出力の統合過程にある。VMでは Edge Drivenと呼ばれる手法を用いている。これは各モジュールから

の出力を、Gemen兄弟の結合マルコフ確立場モデル[68-70]における線過程の枠組みを用いて処理した後、エッジ部で統合するという手法である。

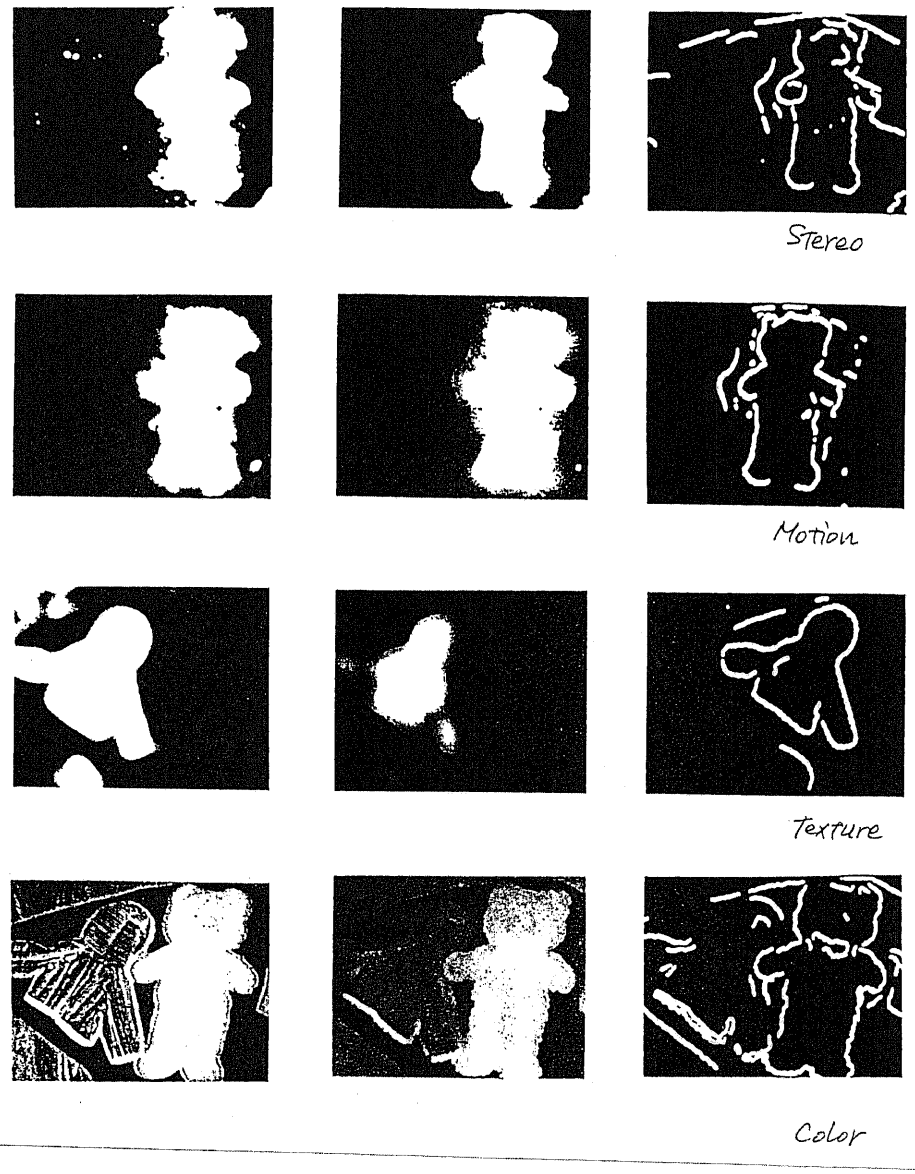


図５３．Vision Machine による画像処理結果

本システムは実際に稼働したが、実時間処理には及ばず、専用のVLSIが必要であると結論している。またこのVLSIは独自に開発・設計し、analogで動作させるとのことである[71,72]。ところで、本論文の他の興味深い点に、本研究への着手を勧めた人物として、Subsumption Architecture(SA)の提唱者である、MITのBrooks教授の名が挙げられている点がある。

（２）　Subsumption Architecture (SA)[73]

①．ＳＡ：

複数の特徴抽出モジュールの出力の統合アルゴリズムに関しては現在様々な研究が行われているが、MIT AI研究所の移動ロボットの研究者Rodney Brooks 教授が提唱する Subsumption Archtecture は、モジュール群の出力の統合を前提としており興味深い。

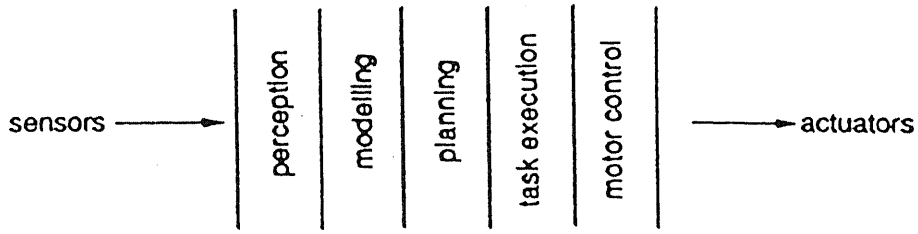
Brooks教授は、従来の移動ロボットの、

環境の認識　　(perseption)
→　モデル化　　(mode-lling)
→　行動の計画　　(planning)
→　タスクの生成・実行　　(task execution),

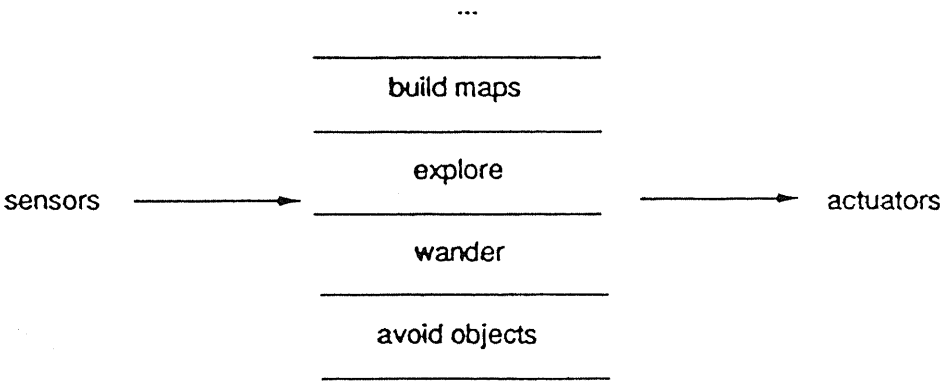
という一連の直列型処理体系に対し、水平型処理体系とでもいうべき処理体系を提案した。これがSabsumption Architecture (SA) である。

"subsume"とは、「他を抑え込む」という意味である。SAでは、従来の直列型の処理方式ではなく、認識用センサ＆行動モデルのペア（エージェントと呼ばれる）を、多数並列に稼働して競合させ、最も他を抑え込んだエージェントの出力をシステムの出力として採用するという手法をとる（図５４）。

例えば移動ロボットにおいて、前進エージェントが他をSubsumeしている場合を想定する。前方に障害物が現れ、それがセンサに探知されると、回避エージェントの重みが増加して他をSubsumeする。回避行動の結果、もはや障害物が探知されなくなると、再び前進エージェントの重みが増加し、他をSubsumeして前進を始める。これはロボットの障害物回避行動に他ならない。



Classical decomposition of an autonomous robot.



Behavior-based decomposition of an autonomous robot.

図５４．Subsumption Architecture

このように、SAでは全体を統治する知能の中枢を設定せず、多数の独立でかつ比較的単純なエージェントの集合体として構成する。これは自律分散処理思想の一つと考えられ、いわゆるSensor Fusion （センサ融合）とは異なるものであり、Sensor Fission （センサ分離：筆者訳）と呼ばれている。

Brooks教授は、やがて高度な知能もこのような枠組みで実現できると考えているようである。事実、伝統的な解析的AI手法を暗に皮肉った論文、"Elephants Don't Play Chess"[74]を発表している。しかしBrooks教授の論文の参考文献には、人工知能の父といわれる Minsky教授の "The Society of Mind"[75]が常に載せられており、極めて興味深い。Minsky教授の発想を、Brooks教授が知能ロボティクスに応用したと見るのが自然かも知れない。

7．3　人間型エージェントの「動き」制御の検討
：「視線」移動のシミュレーション

7．3．1　心理物理学的背景

（1）特徴抽出モジュール

人間の視覚系が抽出する画像情報が、明るさ・色・運動などの多次元情報であることは、D.Marr や P.Cavanaghらの研究により、既に確認済みといってよいが、ここでは特徴抽出モジュールの存在意義について、その計算論的根拠を述べる。

多次元の情報を１つのパラメータ空間で全て表現するとしよう。例えば１０特徴次元で１０レベルの情報を符号化するには10^{10}＝１００億符号単位を必要とする。これを１０のパラメータ空間で表現すれば、１０×１０＝１００符号単位だけで同等の表現が可能となる。即ち独立したモジュールによる特徴抽出は、符号単位の爆発的増加を防ぐために是非とも必要であると考えられる。しかしその反面、結合錯誤が避けられないという本質的欠点を有するが、実生活上ではさほど問題とはならない。

（2）特徴統合理論　－ 基本要素はいかに統合されるか －

では独立に抽出された画像特徴はいかにして統合され、認識が行われるのであろうか。この疑問に答えようとするのが、特徴統合理論[76]に始まる一連の研究である（図５５）。

特徴統合理論によれば、まず入力画像に対し、各モジュールからの出力結果毎に初期特徴のマップ群が形成されるとする。次に入力画像上のある領域（スポットライトと呼ばれる）に注意が向けられると、その領域が示す位置に対応する各マップの情報を統合する。高次の視覚認識は、このス

ポットライトをランダムに移動させることにより行われると考える。

即ち本理論では、人間はスポットライト上で統合される僅かな情報を空間的に収集することにより（中間表現，2.5次元スケッチを構築し）、３次元モデルの脳内表現に至ると説明する。この統合・認識過程は逐次的に行われると考える。

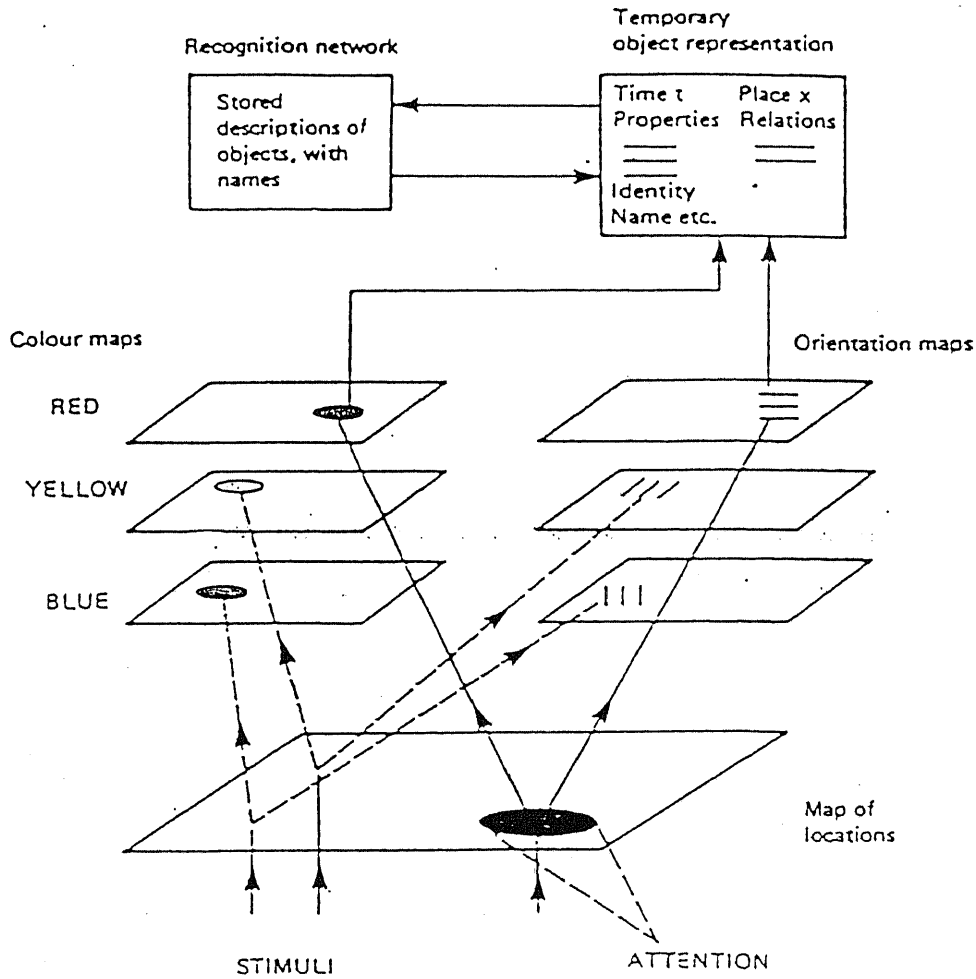


図５５．特徴統合モデル

（３）選択的注意の誘導　－誘導探索モデル－

特徴統合理論の後、この理論の修正モデルがいくつか提案された。中でも代表的なものが誘導探索モデル[77]である。特徴統合理論では、注意の移動はランダムに行われるとしているが、ここではまず各モジュールからの出力に基づき注意のマップ（Attention Map）が形成されるとする。この

マップでは、入力画像において何らかのPrimitiveなレベルでの特徴を有する部位の重みが大きく表現され、注意の移動はこの重みに基づいて行われるとする。

ただし、この注意のマップの生成過程に関し、いかなる特徴が重みとなるのか等について不明な点があり、現在その詳細が研究されている。

7．3．2　本研究で用いた「視線」移動のための視覚モデル

本研究では、以上のような近年の心理物理学を中心とする多数の研究成果をまとめ、図５６に示す視覚モデルを作成した。

また５．１．３節で述べた画像特徴の統合における特徴抽出モジュールは、ステレオ視が不可能である点を除けば、人間の初期視覚認識過程における情報の抽出モジュールとの類似性が高い。この点を利用し、先の画像特徴の統合による人物動画像の実時間認識手法は、本モデルに基づき作成している。すなわち先の人物像の認識手法は、本視覚モデルの並列コンピュータ上への実装例となっている。以下では、図５６に対応する本システムでの処理内容を述べる。

まず特徴抽出モジュールで並列に抽出されたPrimitiveな画像特徴に基づく注意のマップが形成され（図５６中①，ボトムアップによる注意点の選定と呼ぶ）、注意のマップに基づいた順序（②）で特徴抽出モジュールの出力が統合された後（③）、高次視覚系に伝達される（④）。これを受けて、高次視覚系では高次の認識処理が行われ、場合によっては新たな注意点の選定が行われる（⑤，トップダウンによる注意点の選定と呼ぶ）とする。ここで、ボトムアップにより選定された注意点に対しトップダウンによる注意点とが異なる場合、最終的な注意点の決定に際し競合が発生するが、これについては後述する。

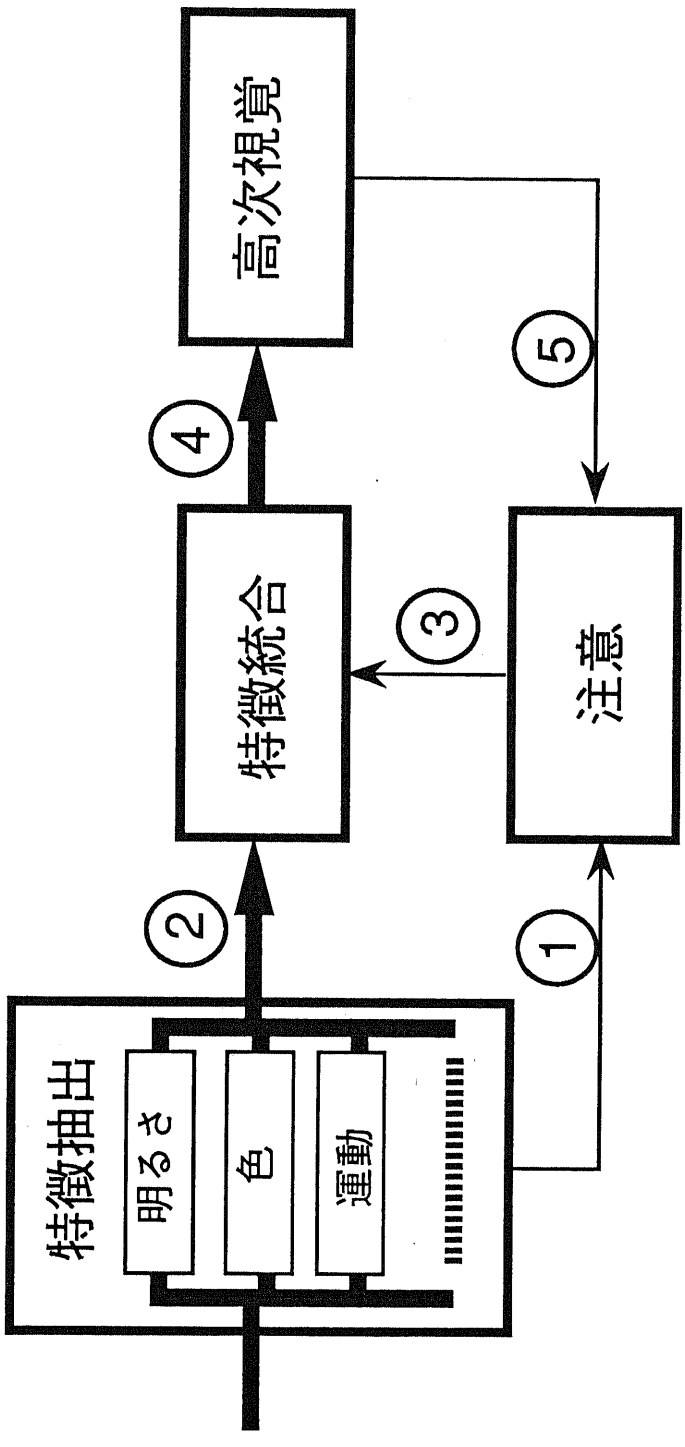


図５６．本研究で用いた「視覚」モデル

7．3．3　特徴抽出モジュール再考

本節では、5．1．3節に述べた各特徴抽出モジュールの有する機能を人間の視覚機能との対比の観点から再考する。

先に述べた5つのモジュールの有する機能はいずれも人間の視覚が有する機能と考えられるが、このうち以下の2つのモジュールに関してはそれぞれMagno型，Parvo型と改称することが可能である。

①．画像全体の明るさ抽出モジュール，

↓

「Magno型」明るさ抽出モジュール

＊本モジュールは低解像度ではあるが高速に明るさの検出を行うため、視覚のMagno系との類似度が高いため。

④．設定画像との差分抽出モジュール，

↓

「Parvo型」明るさ抽出モジュール

＊本モジュールは高解像度であるがタイムスパンの長い明度差情報を抽出するため、視覚のParvo系との類似度が高いため。

7．3．4　注意の競合

本システムでは、トップダウンによる注意点の選定処理と、ボトムアップによる注意点の選定処理は並列に実行することが可能である。これは即ち、注意のマップ①と②が競合することを意味している。

ここで、マップ②に示す注意点の重みを大きくとっているため、通常は

その点をトラッキングするが、画面中に新たに大きな動きなどが計測された場合（新たに人物が画像中に入ってきた場合など）、そちらに一時注意が移動するようにしている。具体的には、マップ①における新たな重みの最大値が、マップ②の重みの最大値の７割を越えた場合としている。

７．３．５　多彩な「視線」の合成

多彩な視線の合成は、各モジュール出力に対する重みを統合時に調節することにより行う。すなわち、各モジュールからの出力を以下のように表すとき、

①．「Magno型」明るさ抽出モジュール　→　Bright_n
②．動き情報抽出モジュール　→　ME2(x,y)
③．色情報抽出モジュール　→　ME3(x,y)
④．「Parvo型」明るさ抽出モジュール　→　ME4(x,y)

視線の方向、つまり画像上の注視点eye(x,y)を以下のように算出している。

```
For x = 1 to m
  SEQ
    For y = 1 to n
      SEQ
        eye(x,y) := max( a × bright_n,
                         b × ME2(x,y),
                         c × ME3(x,y),
```

d × ME4(x,y))

y := y + 1

x := x + 1

(a, b, c, d は統合時の重み係数)

最終的な注視点は、eye(x,y)のデータの大きい順に移動させて合成している。

以下に重み係数の設定値による、合成される視線の様子を示す。

a を大きくしたとき → 大きな明るさの変化があった場合、瞬きする。

b を大きくしたとき → 画像上の動きに対する反応が敏感になる。

c を大きくしたとき → 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。

d を大きくしたとき → 新たに画面に入ってきた対象に対し、敏感に反応する。

ここで注意して頂きたい点は、画像上注視点は人物画像の視線の合成にのみ用いるものであり、システム内部の重みとは必ずしも一致していない点である。

具体的には、人物像が画像上に存在する場合、システム内部のその部分の重みは大きくなるが、背景で動く物体が観測された場合には、重みが大きくなくともそちらに一時的に視線が移動する場合もあるということを意味する。

7．4　7章のまとめ

本章では、ＶＳＡにおける人間型エージェントの自然な挙動の合成について検討した。

従来の顔を用いたヒューマンインタフェースの研究では、主として豊富な表情の合成によるリアリティの追求が進められてきた。そこで本研究では、ユーザの画像入力に実時間で反応し、人間に近い自然な視線の移動をする顔画像の実現を目指した。すなわち、入力に対応し「落ちついた」視線，「きびきびした」視線などを表現することを目指した。

具体的には、システム上に人間の視覚システムとの類似性の高い画像特徴抽出モジュール群を構成し、それらの出力の統合過程において、統合時の重みを調整することにより行った。

これは近年の心理物理学などの研究成果により、人間はその初期視覚認識過程において、誰でもほぼ類似の反応を示すとされることに基づいている。すなわち、人間の視覚認識過程の初期には、視覚のハードウェアのみに依存し意識的なコントロールを受けていない、誰にでも共通の視線の挙動が観察されることを利用している。自然な人間らしい視線とは、ここではこのような視線の挙動を対象とした。

例えば、動き情報に敏感に反応する構成とすれば、画面中に何らかの動きが現れる度にそちらに視線を向けるため、「落ち着かない」視線を表現することができる。その逆にすれば「落ち着いた」視線となる。

他にも、各特徴抽出モジュールからの出力に全体的に反応しないように

すれば、「にぶい，つかれた」エージェントの様子を合成することが可能となる。

第８章 Ｖｉｓｕａｌ Ｓｏｆｔｗａｒｅ Ａｇｅｎｔ（ＶＳＡ）のプロトタイプ

８．１ プロトタイプシステムの構成

プロトタイプシステムは、図５７に示したハードウェア構成により行った。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。マウスのＴＮ－ＶＩＴへの組み込みや、他のＰＣの接続など、周辺機器のプログラミングには、Microsoft C を用いた。

またプロトタイプシステムには、トップエンドにEZEL社のTAICHIが組み込まれており、必要があればローレベルの画像処理を実時間で実行することができる。しかしTAICHIは、カラー動画像の扱いには向かないなどの短所も有する。このため、本プロトタイプシステムにおいては、主として画像の処理・認識もＴＮ－ＶＩＴ上で行った。

8．2　動画像認識結果

8．2．1　実時間動領域検出結果

まず最も簡単なＣＶとＣＧの結合の試みとして、実時間での入力画像上の動領域の抽出と、その結果に反応する動画像の合成を試みた。

動領域の検出は、TAICHI上でフレーム間差分を算出することにより行った。この差分の算出は、TAICHIの画像処理の高速性能を活用し、512×512ピクセルの画像を対象としても、1/30秒で実行が可能である。

入力画像に対する反応形式としては、

①．人物型ＳＲの場合：視線を反応の出た方向に向ける。

②．金魚型ＳＲの場合：２匹の金魚が尻尾を振りながら近寄ってくる。

とした。

ここで人物像の合成には、オブジェクト毎にプロセッサを割り当てる対象分割処理を利用した。オブジェクトは目，口，瞼，背景とし、それぞれにＶＩＴを一台ずつ割り当てた。

金魚像の合成は、大小２匹の金魚を時間分割型の並列処理により行った。

8．2．2　ハンドサインの認識結果

ハンドサインの認識処理の処理速度は、１台のＶＩＴによる処理で約0.5秒／サイクルであった。

認識実験は、長袖・半袖それぞれ100回ずつ行い、正しく認識された回数を調べた。認識率は、半袖の場合で95%以上であったが、長袖の場合は袖の配色により結果が異なった。長袖の場合、袖に肌色を含まない場合認識率は平均約76.8%であり、含む場合は平均約50.6%であった。ただし認識率は袖の配色により大きく左右されるため、本結果は参考データである。

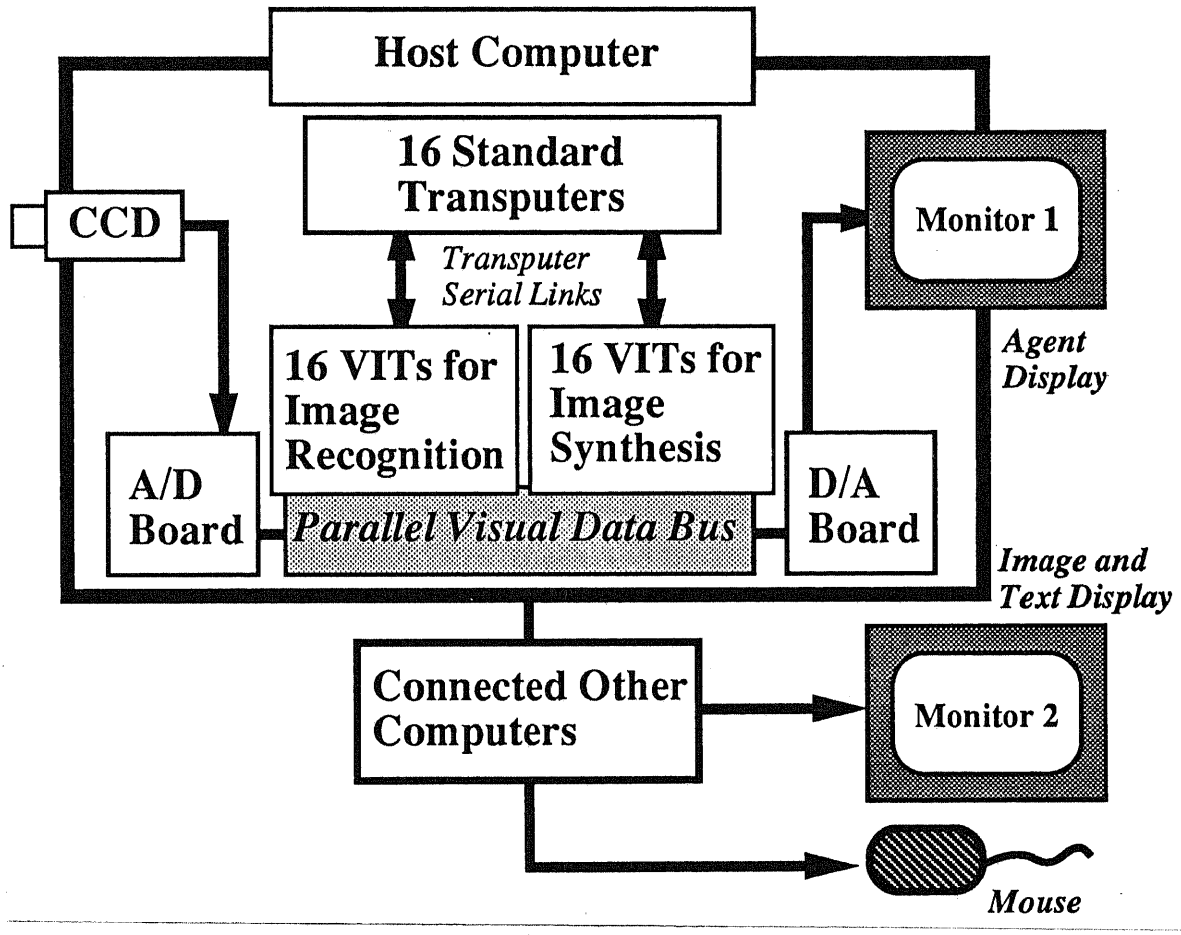


図５７．ＶＳＡのハードウェア構成

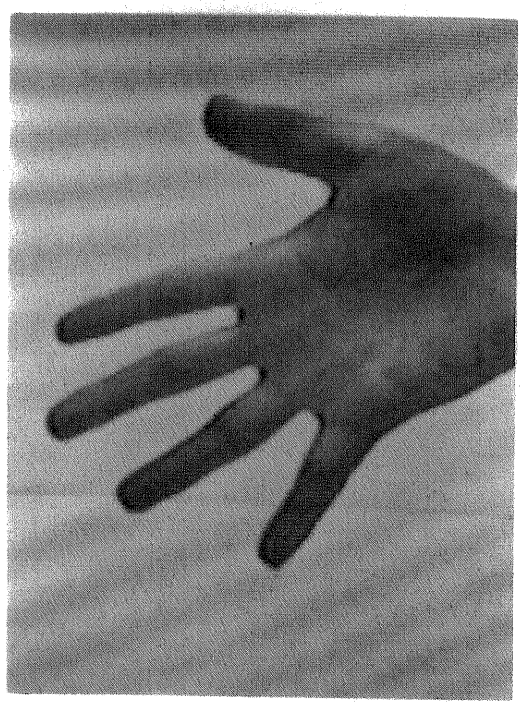


図５８．ハンドサイン１　（原画像）

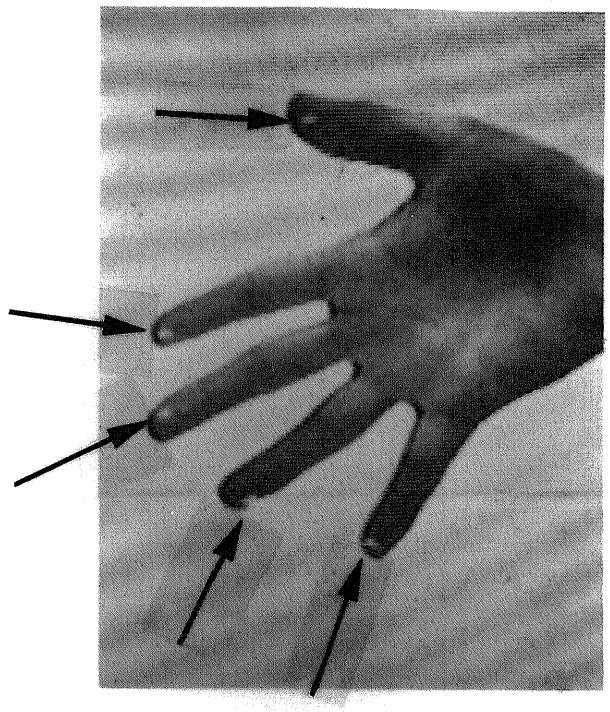


図５９．ハンドサイン１　（指先の検出結果：成功例）

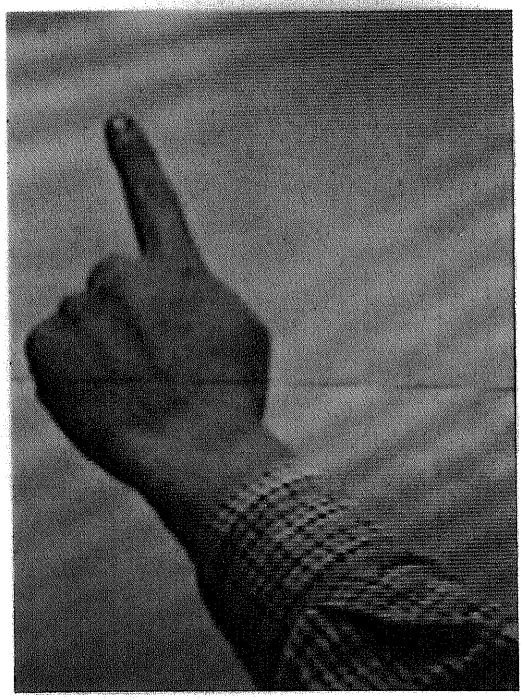


図６０．ハンドサイン２　（原画像）

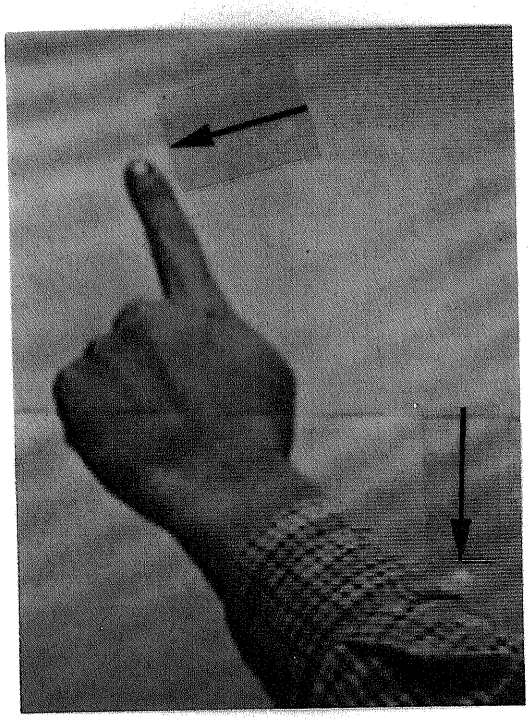


図６１．ハンドサイン２　（指先の検出結果：失敗例）

８．２．３　画像特徴の統合による対象物の認識結果

［Ｉ］．実人物像の抽出実験

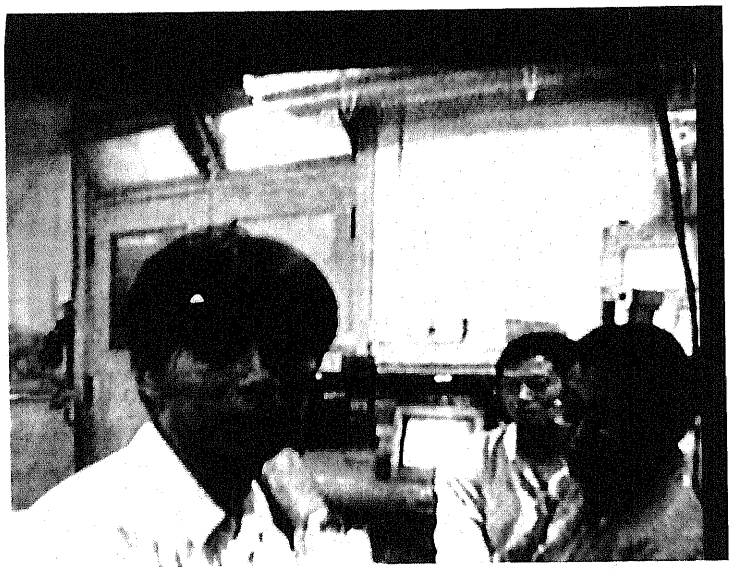
各特徴抽出モジュールにおける処理速度は約90msecであり、最終的な認識結果出力までの所要時間は、約120msecであった。

３～５人程度の人物が随時出入りする画像中より、カメラの方向を向いている人物のみを抽出する実験を行った。その結果、被験者８名（男６名，女２名）に対し、髪を短く刈り込んだ１名のみに対しては認識率が困難であったが、認識可能な人物においては認識率は90%以上であった。また認識処理には約120msec、エージェントの描画終了までには約240msecを要し、描画速度は毎秒約12コマであった。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。

図６２に本実験の実行中の様子を示す。図６２において、システムが抽出した実物像は微小３角形で表示している。

［Ⅱ］．口の開閉状態の認識実験

①の実験で認識可能であった人物の、肩から上の画像が画面全体にわたり得られた場合に対し、口の開閉状態の検出実験を行った。その結果、カメラに対して正面向きを基本とし鉛直軸に対して約±20°の回転角以内であれば、通常の「あ」の発音時程度の口の開閉状態の検出率は90%以上であった。



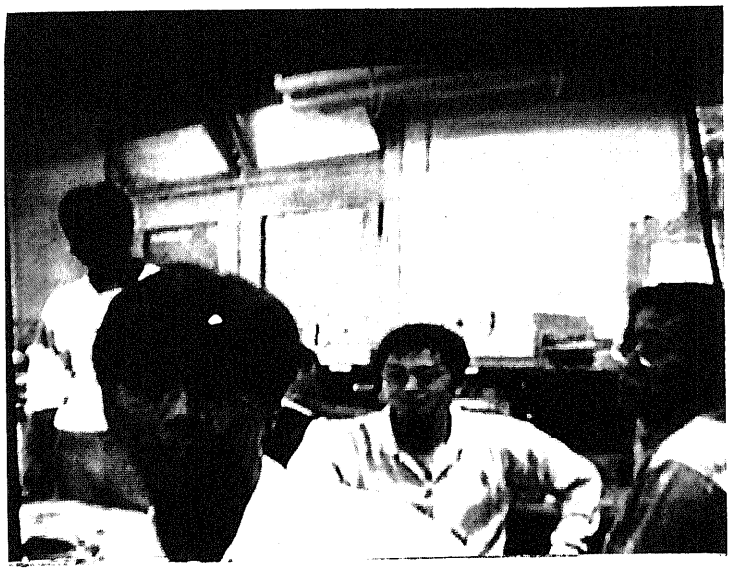


図６２．複数の人物が存在する状況下におけるユーザの抽出結果

［Ⅲ］．虚人物像の識別実験

画像中に人物のポスターなどの虚人物像を混入した実験では、動きの情報が検出されないために実際の人物像との識別が可能であり、識別率は約95%であった。

［Ⅳ］．明るさの変化に対するしきい値の補正実験

照明条件を変化させた実験では、約100msec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。

［Ⅴ］．「視線」の合成実験

画像の認識結果に基づいた、数種の視線の合成を行った。実験結果に対する定量的評価は困難であるため、ここでは結果は概ね良好であったと述べるにとどめておく。

8．3　動画像合成結果

8．3．1　人間型エージェントの合成結果

図６３～図６５に、２体の人間型ソフトウェアロボット（エージェント）の合成結果を示す。

図６３は、エージェントの合成に用いたテクスチャの原画像である。図６４はエージェントの顔の各部分の動作例である。図６５はエージェント頭部の頷きの挙動を示す。

また図６６に、実現されたＶＳＡのプロトタイプシステムを示す。

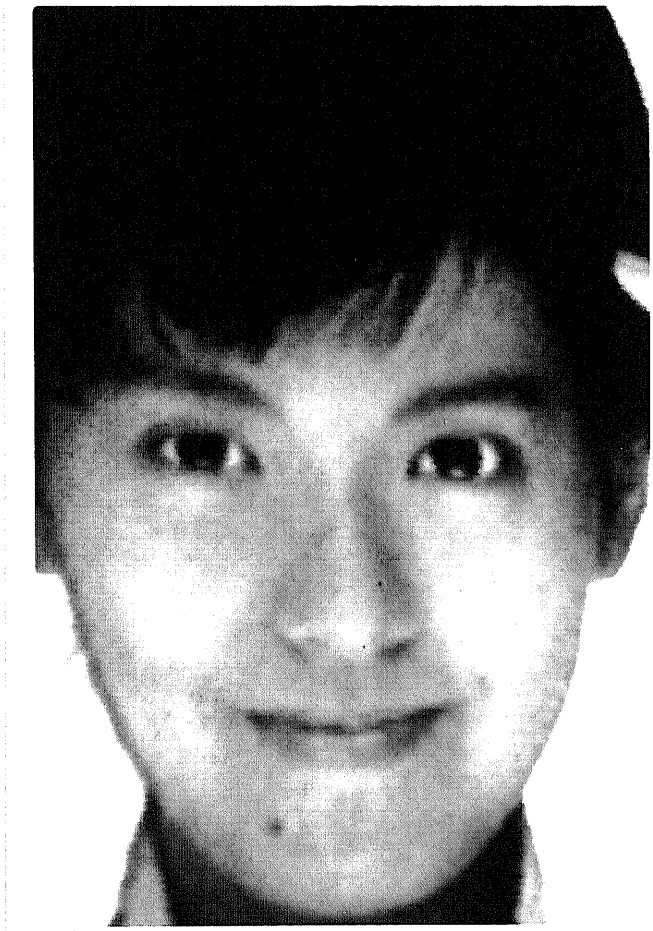


図６３．エージェントの原画像

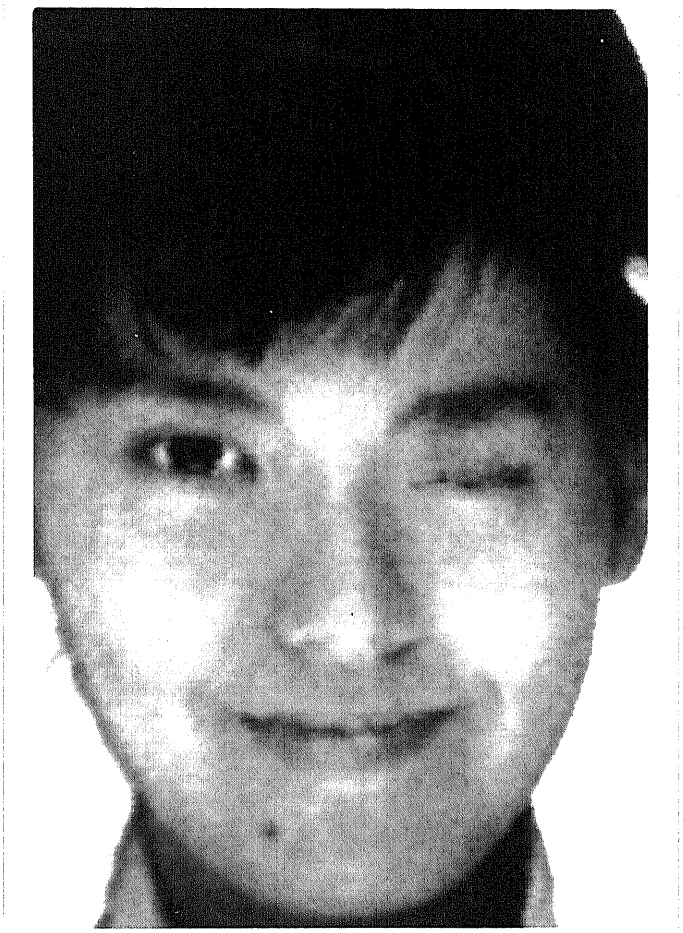


図６４．顔各部の動き

図６５．エージェントの頷き

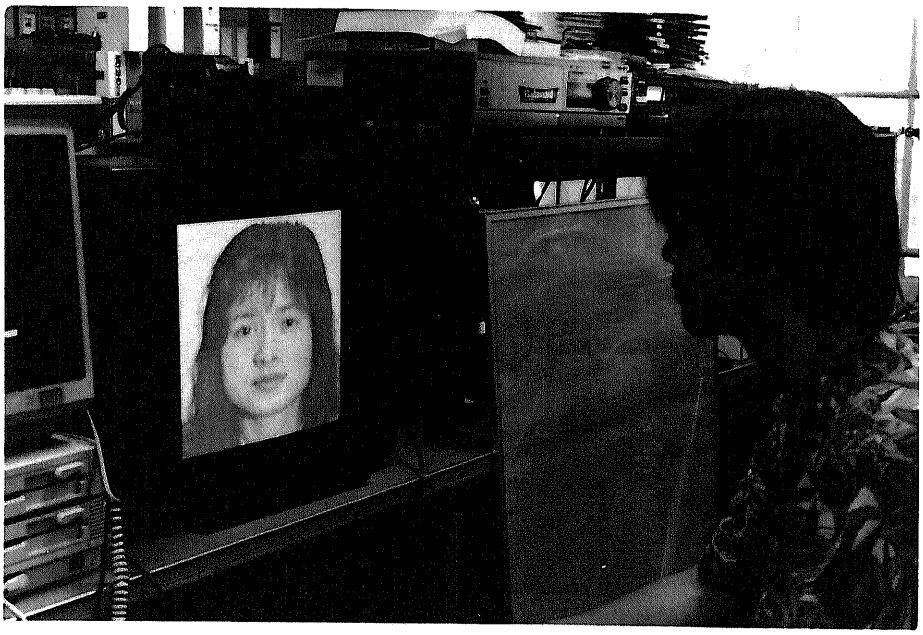


図６６．ＶＳＡのプロトタイプシステム

8．3．2　金魚型エージェントの合成結果

図６７に、本研究で作成した金魚型ソフトウェアロボット（エージェント）の初期のプロトタイプの合成結果を示す。また図６８に、金魚画像とユーザのインタラクションの様子を示す。

図６７．合成された金魚画像

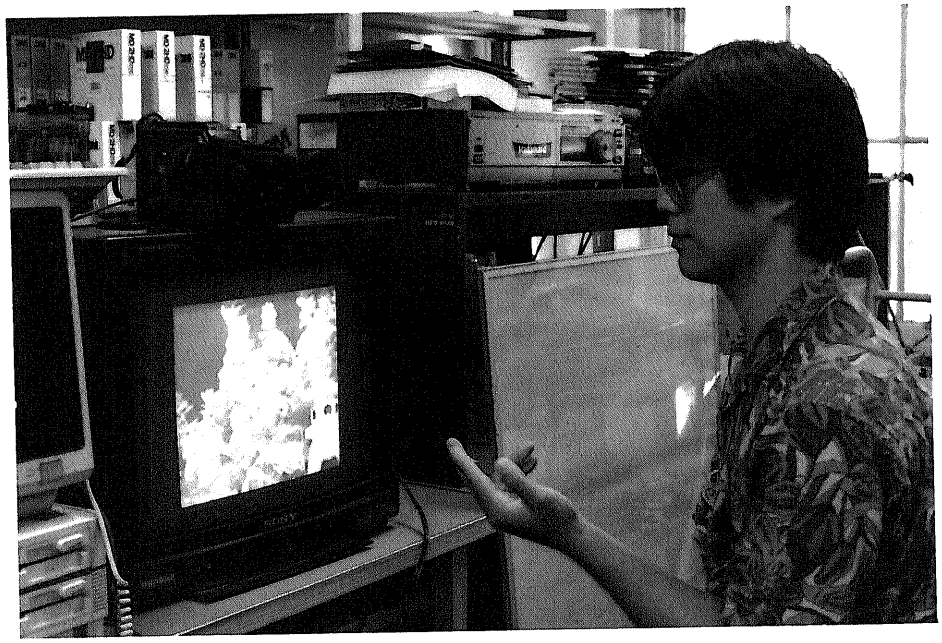


図６８．金魚画像とユーザのインタラクションの様子

8．4　8章のまとめ

本章では、本研究で構築したＶＳＡのプロトタイプの初期版による、動画像認識処理の性能評価、及び合成画像の例を示した。

ハンドサインの認識実験では、実時間処理を重視したアルゴリズムにより簡易で高速な認識処理を実現し、その有効性を確認した。ただし本アルゴリズムでは、条件によっては認識精度の点で改良の余地が残ることも確認された。

ハンドサインの認識は、簡易でしかもハードウェアによる制約（データグローブでは手の大きさに影響を受ける）が少ないため、今後も各方面で研究が進められるものと考える。実際につい最近(1992年秋)、ＡＴＲは手の３次元形状の実時間認識を実現したと発表した。また、通常のインタフェースにおける利用の他にも、様々な形式の手話の認識とその健常者への翻訳などに適用できる可能性があるなど応用範囲も広く、今後の発展に期待されている。

動画像認識手法は、頭の回転方向の大まかな測定が可能であるが、回転角の正確な計測は困難である。しかし人間は興味や関心のある対象に対し、ほぼ正面から対面する傾向があり、現段階の機能のみでも、システム（画像入力用カメラ）に対面している人物像、すなわちユーザのみを実画像からリアルタイムに抽出することは十分可能である。

一方で、現段階では認識対象者の前髪や化粧の有無に影響を受けてしまう短所を有する。すなわち、長髪には影響を受けにくいものの、髪を短く刈り込んでいる人や女性などに対しては認識率が大きく低下してしまう。

これは正面向きの画像において、いわゆる前髪の面積量が不足することによる。これらの欠点は、色情報とは別に、人間の頭特有の動きを調査したり、頭や顔のより詳しい形状情報をシステムに書き加えるなどすれば、改善可能であると思われる。

今後これらの欠点が克服できれば、本稿で述べた手法は他にも、福祉、セキュリティ、市場調査（商品を何人が何秒見ていたかなどを自動計測する）などの分野にも応用が可能であるものと考える。

動画像の合成においては、本研究では基本的にワイヤフレーム上へのテクスチャマッピングの技術と、並列動画像合成の各手法を用いた。これらの手法は以前にも報告された例がある。また本論文においては、喜怒哀楽等、感情を表現する顔画像は合成の対象としなかった。これらについても、既に報告された例がある。これに対し、本研究では画像認識の結果に基づく動画を合成し、人間とのインタラクションを図ること、すなわちソフトウェアロボットの構築を目標とした。

顔画像の合成では、多彩な視線の合成を試みた。合成は、画像特徴並列抽出モジュールの出力の統合時の重みを変えることにより行った。合成結果の評価は課題として残るが、感性情報処理研究のためのツールなどとしての利用が可能であろう。

第９章　結　論

本研究では、次世代のヒューマンインタフェースとしてのＶＳＡ（Visual Software Agent)を提案し、そのうち動画像認識・合成を中心としたプロトタイプを構築した。プロトタイプは、石塚研究室独自の画像バス付きトランスピュータボード：ＶＩＴを中心とする並列コンピューティングシステム：ＴＮ－ＶＩＴ上に構築した。

コンピュータのインタフェースの歴史は、プラグボードにはじまり、パンチ・カードからデスクトップメタフォアに代表されるＧＵＩへと進展し、現在ＭＭＩが花開こうとしている。そしてその次の世代には、人間の秘書のような働きをする、エージェント・メタファーが登場するとされている。

本論文で述べたのは、そのエージェント・メタファーの具体的な形での構築の第一歩であった。

動画像の実時間認識においては、ＶＩＴとＴＮ－ＶＩＴシステムの各種

機能を活用し、ハンドサインの認識アルゴリズムと、画像特徴の統合による並列協調的動画像認識手法を提案した。

中でも後者は、本研究では人物像の認識に用いたが、並列画像認識手法の一つとして一般化・改良することが可能であると考えられる[81,82]。特に画像情報の統合時における結合型マルコフ確立場モデル（Ｃ－ＭＲＦ）の利用は、今後の重要な方向であろう。Ｃ－ＭＲＦは通常多大な処理時間を必要とするが、並列処理との整合性が良く、ＴＮ－ＶＩＴ上での処理においては大きな利点となる。

Ｃ－ＭＲＦにおけるラインプロセスの学習の問題も興味深い。池田・乾らは、ラインプロセスの学習にボルツマン型の学習則[79]と数学的に等価な学習則を考案し、人物濃淡画像のエッジの抽出に適用した事例を報告している[80,84]。

また本研究においては、単眼で全ての認識処理を行ったが、複眼での認識処理の検討も画像認識の観点からは興味深い。

動画としてのエージェント像は、ワイヤフレームモデル上にテクスチャマッピングし、人間型と金魚型の２種を作成して自然な動きの合成を試みた。またそれぞれのエージェントは画像認識結果に反応することとし、ユーザの入力に対する実時間インタラクションを実現した。ただし人間型エージェントにおいて、喜怒哀楽のような感情の表現については既に多数の報告例があるため、ここでは対象としなかった。

また一方で、ＶＳＡを真の人間（ユーザ）中心のインタフェースとするために、人間の諸特性を詳しく調べることは重要であると考える。近年「感性工学」や「感性情報処理の研究」として、文字どおり人間の感性の

世界にメスを入れる試みが盛んに行われて徐々に成果が報告されつつあり、それらの成果も多くの貴重な示唆を与えてくれるであろう。

今後は、本研究において実現できなかった、ＶＳＡに対する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となっている。

現在のところＶＳＡはまだその第一歩を踏みだしたに過ぎない。ここに記したささやかな成果が、今後のＶＳＡの研究に役立つならば幸いである。

参考文献

[1] 文部省科学研究費総合研究（Ａ），「感性情報の抽出・検索・表現に関する総合的研究」，第１年次研究報告，研究代表者・辻三郎（阪大），(1991)

[2] 文部省科学研究費総合研究（Ａ），「感性情報の抽出・検索・表現に関する総合的研究」，第２年次研究報告，研究代表者・辻三郎（阪大），(1992)

[3] 重点領域研究，「感性情報処理の情報学・心理学的研究」，第１回公開シンポジウム資料，(1992)

[4] 石塚満：「ヒューマンコミュニケーションと知識処理」，信学ヒューマンコミュニケーション研資，HC90-5,(1990,4)

[5] 郵政省電気通信局電気通信技術システム課（監修）：ヒューマンインターフェイス－人間中心主義のメディアステーションに向けて－，（財）日本データ通信協会（1990）

[6] 小林幸雄：”マルチメディアヒューマンインターフェース”，テレビジョン学会誌，Vol.45,No.8,pp.945-950,(1991)

[7] 計測自動制御学会，ヒューマンインタフェース・シンポジウム講演論文集の各論文

[8] 「９０年代のユーザ・インタフェースを展望する」，日経エレクトロニクス，No.547，pp.297-313（1992）

[9]「次世代ユーザ・インタフェース」，日経エレクトロニクス，No.559，pp.117-133（1992）

[10] B.Shneiderman著，東・井関訳：「ユーザー・インタフェースの設計」，日経マグロウヒル（1987）

[11] S. Brand: "The MIT Media Labolatory", (1988)

[12] Richard A. Bolt : "THE HUMAN INTERFACE", Wadsworth Inc., California, (1984)

[13] H.Harashima, K.Aizaka and T.Saito: "Model-Based Analysis Synthesis Coding of Videotelephone Images -Cooperation and Basic Study of Intelligent Image Coding-", Trans. IEICE, Vol.E72, No.5, May, (1989)

[14] S.Morishima, K.Aizawa and H.Harashima : "Model based facial image coding controlled by the speech parameter", Picture Coding Symposium '88, 4-3, (Sept, 1988)

[15] P.Ekman and W.V. Friesen : "Manual for the Facial Action Coding System", Consulting Psychologists Press, Palo Alto, Calif., (1977)

[16] 森島繁生，小林誠司，原島博：”マルチプロセッサ構成による知的画像符号化のためのリアルタイム表情合成の試み”，信学論（Ｄ－Ⅱ），J73-D-2, pp.1647-1654, (1990,10)

[17] 坂内正夫，佐藤真一：”画像データベースにおけるモデル形成”，信学誌，J74-DI,（新しいデータベース特集号招待論文），(1991)

[18] S.Satoh, M.Sakauchi, et al.: "Drawing Image Understanding Framework using State Transition Models", Proc., 10th ICPR, pp.491-495, (1990)

[19] 坂内正夫：”マルチメディアシステムにおける情報アクセス”，生産研究，44巻11号，pp.53-57, (1992,11)

[20] Tom Calvert : "Composition of Realistic Animation Sequences for Multiple Human Figures", Making Them Move, (N.I.Badler et al. Ed.), pp.35-50, (1992)

[21] N.Thalmann and D.Thalman : "Human Body Deformations Using Joint-dependent Local Operations and Finite Element Theory", Making Them Move, (N.I.Badler et al. Ed.), pp.243-262, (1992)

[22] 服部桂：「人工現実感の世界」，工業調査会，(1991)

[23] 柿本他：”仮想生物システム”，日本非破壊検査協会，００５・００８合同特別研究委員会資料，pp.1-6，(1990)

[24] 林他：”人工現実感による仮想生物との対話（１）”，第４２回情報処理学会全国大会講演論文集（２），pp.293-294，(1991)

[25] 藤田他：”人工現実感による仮想生物との対話（２）”，第４２回情報処理学会全国大会講演論文集（２），pp.295-296，(1991)

[26] 井上他：”並列型画像処理システムとその応用”，情処論文誌，Vol.32，No.7，pp.924-932，(1991)

[27] 飯島泰裕他：”デジタル映像制作編集システムの概要”，第43回情報処理学会全国大会予稿集，2-411，(1991)

[28] M.Ishikawa, A.Morita and N.Takayanagi : "High Speed Vision System Using Massively Parallel Processing", Proc. IEEE IROS, pp.373-377, (1992)

[29] 大田友一：”画像理解アーキチクチャとその動向”，第２０回画像工学コンファレンス，pp.223-228，(1989)

[30] D.C.Van Essen, C.H.Anderson, D.J.Fellema : Science, Vol.255, pp.419-422, (1992)

[31] INMOS Ltd.:"Transputer Reference Manual", Prentice Hall (1988)

[32] Wiwat Wongwarawipat : "Parallel Image Processing System with Distributed Memory and its Application to Fast Moving Image Analysis", Ph.D Thesis, University of Tokyo, (1990)

[33] 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)

[34] 高木幹雄, 下田陽久監修：「画像解析ハンドブック」, 東京大学出版会, (1991)

[35] 谷内田正彦：「ロボットビジョン」, 昭晃堂, (1990)

[36] 白井良明：「コンピュータビジョン」, 昭晃堂, (1980)

[37] A.Rosenfeld, A.Kak : "Digital Image Processing", (1976)

[38] P.Winston編, 白井良明, 杉原厚吉訳：「コンピュータビジョンの心理」, 産業図書, (1979)

[39] R.A.Bolt:"Put-That-There : Voice and Gesture at the Graphic Interface",Comput.Gra., 14, 3, pp.262-270 (1980)

[40] 石淵耕一, 竹村治雄, 岸野文郎：" 画像処理を用いた実時間手形状認識とマンマシンインターフェイスへの応用", 信学秋全大講演論文集, Vol.1, 1-132, (1991)

[41] 加藤昌央, 吹野美和, 小山隆正, 三輪道雄：" 立体図形における手振りインターフェース", 信学秋全大講演論文集, Vol.1, 1-127, (1991)

[42] S.Tamura and S.Kawasaki : "Recognition System for Sign Language Motion Image", 情報処理学会, CV-44-1, pp.1-8, (1986)

[43] H.Kawai and S.Tamura : "Deaf-and-Mute Sign Language Generation System", Pattern Recognition, Vol.18, Nos 3/4, pp.199-205,

(1985)

[44] David Marr : "Vision: A Computational investigation into the human representation and processing of visual information", W.H.Freeman and Company, San Fransisco (1982)

[45] P.Cavannah : "Reconstructiong the Third Dimension: Interactions between Color, Texture, Motion, Binocular Disparity, and Shape", CVGIP, 37, pp.171-195, (1987)

[46] 清水嘉重郎編著：「生物の目とセンサ」，情報調査会，(1985)

[47] K.T.Spoehr, S.W.Lehmkuhle : "Visual Information Processing", W.H.Freeman, New York, (1982)

[48] 乾敏郎：「視覚情報処理の基礎」，サイエンス社，(1990)

[49] P.Maes & R.Brooks : "Learning to Coordinate Behaviors", Proc., AAAI-90, pp.796-802 (1990)

[50] V．ブルース著，吉川左世紀子訳：「顔の認知と情報処理」，サイエンス社，(1990)

[51] Venu Govindaraju et al : "A Computational Model for Face Location based on Cognitive Principles", Proc. AAAI-92, pp.350-355, (1992)

[52] K.Mase, Y.Suenaga and T.Akimoto : "Head Reader: A head motion understanding system for better man-machine interaction", IEEE Proc. SMC, pp.970-974, (1987)

[53] 角，大田：" 多様な入力を許容する顔画像解析システムの一構成法 "，信学論，D－2，J75-D-2, 2, pp.236-245 (1992)

[54] T.Kurita, N.Otsu and T.Sato : "A Face Recognition Method Using Higher Order Local Autocorrelation and Multivariate Analy-

sis", Proc. ICPR'92, pp.213-216, (1992)

[55] 高木, 坂内編, 安居院, 中嶋共著:「コンピュータグラフィックス」, 昭晃堂, (1992)

[56] 李七雨:" ディフォーマブルモデルを用いた動画像認識・合成に関する研究", 博士論文, 東京大学, (1992)

[57] 横澤一彦:" 一目でわかること", 科学, 岩波, (1992)

[58] 長谷川:「生体に学ぶ視覚情報処理」, 東京大学工学部電子工学科, 大学院論文輪講資料 (1992,6)

[59] 田崎, 大山, 樋渡:「視覚情報処理」, 朝倉書店, (1979)

[60] P.H.Lindsay and D.A.Norman : "Human Information Processing", Academic Press, New York, (1977)

[61] J.J.Gibson : "The Ecological Approach to Visual Perception", Houghton Mifflin Company, Boston (1979)

[62] G.Kanizsa : "Organization in Vision: Essays on Gestalt Perception", Praeger Publishers, (1979)

[63] 伊藤, 佐伯編:「認識し行動する脳－脳科学と認知科学－」, 東京大学出版会, (1988)

[64] D.E.Lumelhart : "Introduction to Human Information Processing ", John Wiley & Sons, Inc., (1977)

[65] R. J. ワット著, 乾敏郎監修:「視覚情報処理モデル入門」, サイエンス社, (1989)

[66] 乾, 川人:" 視覚大脳皮質の計算理論", 信学論, D-2, Vol.J73-D-2, pp.1111-1121, (1990)

[67] T.Poggio et al : "THE MIT VISION MACHINE", Proc. Image Understanding Workshop, pp.177-198, (1988)

[68] S.Geman and D.Geman : "Stochastic relaxation, Gibbs distributions, and the basian restoration of images", IEEE Trans., PAMI, Vol.6, pp.721-741, (1984)

[69] D.Geman : "Stochastic model for boundary detection", Image and vision computing, Vol.5, No.2, pp.61-65, (1987)

[70] H.Derin, H.Elliott, R.Cristi and D.Geman : "Bayes Smoothing Algorithms for Segmentation of Binary Images Modeled by Marcov Random Fields", IEEE Treans. PAMI, Vol.6, No.6, pp.707-720, (1984)

[71] J.L.Wyatt et al : "The MIT Chip Project: Analog VLSI Systems for Fast Image Acquisition and Early Vision Processing", Proc. IEEE R&A, pp.1330-1335, (1991)

[72] M.Ishikawa et al:"High Speed Vision System Using Massively Parallel Processing", Proc., IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems, pp.373-377, (1992)

[73] R.Brooks : "Intelligence without Reason", Proc.IJIKAI-92, pp.569-595, (1992)

[74] R.Brooks : "Elephants Don't Play Chess", Designing Autonomous Agents (P.Maes Ed.), pp.3-15, (1991)

[75] M.Minsky, 安西祐一郎訳：「心の社会」, 産業図書, (1990)

[76] A.Treisman and S.Gormican : "Feature Analysis in Early Vision : Evidence from Search Asymmetries", Psycological Review, 95, 1, pp.15-48 (1988)

[77] K.R.Cave and J.M.Wolfe : "Modeling the Role of Parallel Processing in Visual Search", Cognitive Psycology, 22, pp.225-271

(1990)

[78] D.Pountain and D.May : "A Tutorial Introduction to OCCAM Programming", BSP Professional Books, Great Britain, (1988)

[79] 合原一幸：「ニューラルコンピューター脳と神経に学ぶー」, 東京電機大学出版局, (1988)

[80] 麻生英樹：「ニューラルネットワーク情報処理」, 産業図書, (1988)

[81] R.C.Jain : "DIALOGUE : Ignorance, Myopia, and Naivete in Computer Vision Systems", CVGIP, Vol.53, No.1, Jan., pp.112-117, (1991)

[82] S.M.Culhane and J.K.Tsotsos : "A Prototype for Data-Driven Visual Attension", Proc. ICPR'92, pp.36-40, (1992)

[83] Y.Sato and M.Toyoda : "Active Vision with Two Differentiated Visual Fields", Proc. ICPR'92, pp.31-35, (1992)

[84] S.Hongo, M.Kawato, T.Inui and S.Miyake : "Contour Extraction of Images on Parallel Computer - Local, Parallel and Stochastic Algorithm which Learns Energy Parameters -", IJCNN-89, Vol.1, pp.161-168, (1989)

付　録

－ トランスピュータを介した他のNEC_PCとの接続・通信プログラム －

本プログラムは、ＴＮ－ＶＩＴに対するユーザの画像入力に反応し、ホストコンピュータ用の PC 以外の PC 上にメッセージなどを表示させるプログラムである。

本プログラムには、ＮＥＣ製パーソナルコンピュータに挿入したトランスピュータボード（B012など）上のトランスピュータと本体との、通信制御部が含まれている。従ってこれを書き換えることにより、ＴＮ－ＶＩＴへの任意のNEC-PCの付加が可能となる。

コンパイラには Microsoft C を用いている。

```
/*****************************************************************/
/**                                                             **/
/**  TR - NEC_PC communication program                          **/
/**                                        by Osamu Hasegawa    **/
/**                                        July, 1992           **/
/**                                                             **/
/*****************************************************************/

#define LINT_ARGS       1
#define REV_A_FIX       1
#define TRUE    1
#define FALSE   0
#define BIT_0   1

#include <cfunc.h>
#include <conio.h>
#include <stdio.h>
/*#include <srvconst.h>*/
#define         BASE                    0xd0
#define         READ_OFFSET             0
#define         WRITE_OFFSET            2
#define         IN_STATUS_OFFSET        4
#define         OUT_STATUS_OFFSET       6
#define         RESET_OFFSET            8
#define         ERROR_OFFSET            8
#define         ANALYSE_OFFSET          10

#define         RESET_COUNT             10000
#define         ANALYSE_COUNT           10000
#define         MAX_LINK_TIME           32000

unsigned int
                link_base = BASE,
                link_read = BASE + READ_OFFSET,
                link_write = BASE + WRITE_OFFSET,
                link_in_status = BASE + IN_STATUS_OFFSET,
                link_out_status = BASE + OUT_STATUS_OFFSET,
                link_reset = BASE + RESET_OFFSET,
                link_error = BASE + ERROR_OFFSET,
                link_analyse = BASE + ANALYSE_OFFSET;

void init_root()
{
#ifdef NECserver
                if(link_base == 0)
                link_base = BASE ;
                link_read = link_base + READ_OFFSET;
                link_write = link_base + WRITE_OFFSET;
                link_in_status = link_base + IN_STATUS_OFFSET;
                link_out_status = link_base + OUT_STATUS_OFFSET;
```

```
                link_reset = link_base + RESET_OFFSET;
                link_error = link_base + ERROR_OFFSET;
                link_analyse = link_base + ANALYSE_OFFSET;
#endif
}


int link_in_test()
{
        return( (inp(link_in_status) & BIT_0) != 0);
}

int link_out_test()
{
        return( (inp(link_out_status) & BIT_0) != 0);
}



int byte_from_link()
{
        while(!(inp(link_in_status) & BIT_0));
        return(inp(link_read));
}


void byte_to_link(ch)
int ch;
{
        while(!(inp(link_out_status) & BIT_0));
        outp(link_write, ch);
}


int byte_to_link_t(ch)
{
/*int ch;*/
        register int not_written = TRUE, i;
        for( i=0; ((i < MAX_LINK_TIME) && not_written); i++)
                if(inp(link_out_status) & BIT_0){
                        outp(link_write, ch);
                        not_written = FALSE;
                };
        return(not_written);
}


int error_test()
{
        return(!(inp(link_error) & BIT_0));
```

```
}


int word_from_link()
{
        register int t, ch;
        t = byte_from_link();
        ch = byte_from_link();
        t = t | (ch << 8);
        ch = byte_from_link();
        ch = byte_from_link();
        return(t);
}


void word_to_link(w)
int w;
{
        byte_to_link (w & 0xff);
        byte_to_link (( w >> 8) & 0xff);
        if ( w < 0 )
        {
        byte_to_link (0xff);
        byte_to_link (0xff);
        } else
        {
        byte_to_link (0);
        byte_to_link (0);
        }
}


/*
void slice_to_link(length, slice)
{
int length;
char *slice;

        register int t, real_len = length;
        word_to_link(real_len);
#ifdef REV_A_FIX
        if (real_len == 1) real_len = 2;
#endif
        for(i = 0; i < real_len ; i++)
                byte_to_link(*slice);
}
*/


/*------- Main Program is Below ------------------*/
```

```
char *str1        = "WELCOME";
char *str2        = "VSA";
char *str3        = "on";
char *str4        = "TN-VIT";

/*void main()*/
main()
{

int x1= 1, x2=456, y1=1, y2=512;
int push = 0, count = 0;
int mx, my, px, py;
int     i = 0;
int     y = 0;
int     no = 4;
double  ans = .0;

        csr_off();
        dsp_25();
        key_off();
        g_init();
        g_cls();
        g_screen(3,0,0,1);

/*メニュー表示 */

        g_paint(1,1,5,7);

        g_str(250,10," Ishizuka Laboratory  ",1,2);

    for (i=0;i<7;i++) wsf_l  (60+55*i ,40 ,*(str1+i),55 ,50 ,0,1);
    for (i=0;i<3;i++) wsf_l  (60+70*i ,110,*(str2+i),80 ,70 ,0,1);
    for (i=0;i<2;i++) wsf_s  (60+70*i ,200,*(str3+i),80 ,70 ,0,1);
    for (i=0;i<6;i++) wsf_l  (60+70*i ,290,*(str4+i),80 ,70 ,0,1);
                      wsf_k  (60+70*2 ,290,*(str4+2),80 ,70 ,0,1);

        m_init();
        m_level(x1,x2);
        m_vertical(y1,y2);
        m_xhear();
g_str(220,375," Please push mouse left bottun to continue.",1,3);
        while(1){
                m_stat( &push, &count, &mx, &my );
                if (push) break;
        }

        g_cls();
        s_cl();
```

```
        key_on();
        csr_on();

/* Message Display */

        printf("  I am your personal secretary. ¥n");
        printf("¥n");
        printf("  May I help you? ¥n");
        printf("¥n");
        printf("          Please move to ....¥n");
        printf("        Up Left          ----  New Mail¥n");
        printf("        Up Right         ----  Your Schedule¥n");
        printf("        Down Left        ----  Clock¥n");
        printf("        Down Right       ----  Calculator¥n");

        init_root();


/* Data Read from link */
/*
        while(1)
        {
        px = word_from_link();
        printf("Read_X_data_from_link  %d¥n",px);

        py = word_from_link();
        printf("Read_Y_data_from_link  %d¥n",py);

        if ((px >= 100) | (px <= 30))
                {
                if ((py >= 100) | (py <= 30)) break;
                }
        }

        if ((px <= 30) & (py <= 30))
                no = 1;
        if ((px >= 100) & (py <= 30))
                no = 2;
        if ((px <= 30) & (py >= 100))
                no = 3;
        if ((px >= 100) & (py >= 100))
                no = 4;
*/

/* Process Start */
                no = 4;         /* dummy */

        printf("¥n");
        printf(" Please push mouse left bottun to end.¥n");
```

```
        while(1){
        if (no == 3){
                  for( y = 7; y < 18; y++) {
                        locate (56,y);
                        spc(23);
                  }
                  ans = gcalc(ans,2,448,116,1,7,0,5,1,2);
        printf("Step  1¥n");
                  break;
                  }
        if (no == 4){
                  for( y = 6; y < 13; y++) {
                        locate (42,y);
                        spc(15);
                  }
                  g_watch(400,100,50,1,1);
        printf("step  2¥n");
                  break;
                  }
        m_stat( &push, &count, &mx, &my );
        printf("Step  3¥n");
                if (push)
                {
                printf("Mouse bottun has been pushed.¥n");
                break;
                }
        }

        g_cls();
        s_cl();
        key_on();
        csr_on();
}

/*
                if( ( inp(link_out_status) & 01) != 0)
                {
                outp(link_write, (mx & 0xff));
                outp(link_write, ((mx>>8) & 0xff));
                outp(link_write, 0);
                outp(link_write, 0);
                printf("Output_X_data_to_link  %d¥n",mx);

                outp(link_write, (my & 0xff));
                outp(link_write, ((my>>8) & 0xff));
                outp(link_write, 0);
                outp(link_write, 0);
                printf("Output_Y_data_to_link  %d¥n",my);
                }
*/
```

Figure Captions

発表文献

＜ 本論文に関連するもの ＞

1．邦文・英文論文誌

(1) 長谷川・李・Wongwarawipat・石塚：「手形状の認識に基づき実時間で反応する人物表情の動画像合成」，計測自動制御学会論文誌，Vol.28，No.11，pp.1327-1336，1992

(2) O.Hasegawa, C.W.Lee, W.Wongwarawipat and M.Ishizuka : "Realtime Synthesis of Human-like Agent in Response to User's Moving Image", ICPR Special Issue of "Machine Vision and Applications", Springer International, 掲載予定

(3) C.W.Lee, O.Hasegawa, W.Wongwarawipat and M.Ishizuka:"Realistic Image Synthesis of a Deformable Living Thing Based on Motion Understanding", Journal of Visual Communication and Image Representation, Vol.2, No.4, December, pp.345-354, 1991

(4) 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

(5) 長谷川・藤木・李・Wongwarawipat・石塚：「指サイン認識による自然感を有する仮想金魚像との実時間インタラクション」，電子情報通信学会論文誌，（投稿中）

(6) 長谷川・横澤・藤木・石塚：「自然感の高いビジュアルヒューマンインタフェースの実現のための人物顔画像の実時間並列協調的認識」，電子情報通信学会論文誌，（投稿中）

2．国際会議発表論文

(1) W.Wongwarawipat, C.W.Lee, O.Hasegawa, H.Dohi, M.Ishizuka: "Visual Software Agent Built on Transputer Network with Visual Interfaces", Transputing'91, pp.815-827, IOS Press, U.S.A., 1991

(2) O.Hasegawa, W.Wongwarawipat, C.W.Lee, M.Ishizuka: "Real-time Moving Human Face Synthesis Using a Parallel Computer Network" IEEE, Industrial Engineer's Conference (IECON'91), pp.1380 - 1385, November, Kobe, 1991

(3) M.Ishizuka, O.Hasegawa, W.Wongwarawipat, C.W.Lee, H.Dohi: "Visual Software Agent (VSA) built on Transputer Network with Visual Interface (TN-VIT)", Computer World'91, pp.36-46, Osaka, 1991

(4) O.Hasegawa, C.W.Lee, W.Wongwarawipat, M.Ishizuka: "Realtime synthesis of human-like agent in response to user's moving image", 11 th International Conference on Pattern Recognition, August (ICPR), 1992

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

(6) C.W.Lee, O.Hasegawa, H.Dohi and M.Ishizuka : "Motion Understanding of a Nonrigid Object Using Deformable 3-D Model and Constraints", Int'l Conference on Electronics, Information and Communications (ICEIC'91), China, 1991

3．口頭発表論文

(1) 長谷川・李・Wongwarawipat・石塚：「ＶＳＡのための入力画像に反応する実時間人物表情の合成」，第７回ヒューマン・インタフェースシンポジウム，論文講演，pp.477-484，1991

(2) 李・長谷川・Wongwarawipat・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」，第７回ヒューマン・インタフェースシンポジウム，論文講演，pp.477-484，1991

(3) 長谷川・W.Wongwarawipat・李・石塚：「Visual Software Agent による次世代ヒューマンインタフェースに関する研究」，電子情報通信学会秋季全国大会，6-275，1990

(4) W.Wongwarawipat・李・長谷川・土肥・石塚：「並列トランスピュータ上でのビジュアル・ソフトウェアエージェントの実現」，
電子情報通信学会研究会，HC90-30，pp.43-50，1991

(5) 長谷川・W.Wongwarawipat・李・土肥・石塚：「高次ヒューマンインタフェースとしての並列処理によるビジュアル・ソフトウェア・エージェント」，情報処理学会第４２回全国大会，5分冊，pp.265-266，1991

(6) 長谷川・W.Wongwarawipat・李・土肥・石塚：「高次ヒューマンインタフェースとしてのビジュアル・ソフトウェア・エージェントの試作」，
電子情報通信学会春季全国大会，SA-6-2，1991

(7) 長谷川・李・石塚：「TN-VIT上での金魚画像を用いた動画像並列トラッキングシステム」，電子情報通信学会秋季全国大会，A-125，1991

(8) 長谷川・李・石塚：「人物モデルがユーザの挙動に反応する並列画像処理・合成システム」，情報処理学会第４３回全国大会，2分冊，pp.249-250，1991

(9) 藤木・長谷川・李・石塚：「動画像の実時間認識に基づく金魚画像の合成」，情報処理学会第４３回全国大会，1U-03，1991

(10) 長谷川・横澤・石塚：「環境の変化に対し頑健性を有する実時間動画像並列認識システム」，情報処理学会第４４回全国大会講演論文集，2，pp.229-230，1992

(11) 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

(12) 長谷川・横澤・石塚：「生体の視覚系をモデルとした画像認識システムの試作」，第７回生体・生理シンポジウム講演論文集，pp.59-64，1992

(13) 藤木・長谷川・李・石塚：「指サイン認識に基づき動作するデフォーマブル金魚像の実時間合成」，情報処理学会第４４回全国大会講演論文集，1D-03，1992

(14) 長谷川・藤木・陳・石塚：「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」，情報処理学会第４５回全国大会，発表予定

(15) 藤木・長谷川・石塚：「遺伝的アルゴリズムによる感性的動体画の創作支援」，情報処理学会第４５回全国大会，発表予定

4．その他

(1) O.Hasegawa, C.W.Lee, W.Wongwarawipat and M.Ishizuka : "Visual Software Agent Built on Transputer Network with Visual Interface", 東京大学生産技術研究所電気談話会報告, Vol.41, No.19, 1991

(2) O.Hasegawa, C.W.Lee, W.Wongwarawipat and M.Ishizuka: "Realtime Synthesis of Moving Human-like Agent Which Responds to Finger Signs", 東京大学生産技術研究所電気談話会報告, Vol.42, No.20, 1992

(3) M.Fujiki, O.Hasegawa, C.W.Lee, W.Wongwarawipat and M.Ishizuka : "A Prototype of Goldfish Software Robot with Real-time Response Function by a Parallel Computer", 東京大学生産技術研究所電気談話会報告, Vol.42, No.20, 1992

(4) 長谷川・石塚：「並列トランスピュータ上でのビジュアル・ソフトウェアエージェントの実現」, 富士通・ＫＳＡフォーラム第４回, ビジョン・ロボット分科会資料, 1991

(5) 石塚・土肥・長谷川・藤木：「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）」, Vol.44, No.11, pp.525-533, 1992

＜ その他の発表文献 ＞

1．邦文・英文論文誌

(1) 福田・長谷川：「マイクロマニプレータの制御」（第２報, 画像セグ

メント法による生物の認識と同定），日本機械学会論文誌，Ｃ編，55巻512号，pp.959-967，(1989)

(2) 福田・長谷川・浅間・長嶺・遠藤：「画像処理エキスパートシステムによるマイクロキャリア上動物細胞の視覚認識エキスパートシステム」日本機械学会論文誌，Ｃ編，56巻523号，pp.700-705，(1990)

(3) 福田・長谷川：「曲げ・ねじり・亀裂・欠損のある歪直方体の視覚認識及び把握プランニング用エキスパートシステムの研究」
（第１報，亀裂・欠損のある歪直方体の視覚認識法）
日本機械学会論文誌，Ｃ編，56巻529号，pp.2393-2401，(1990)

(4) 福田・長谷川：「曲げ・ねじり・亀裂・欠損のある歪直方体の視覚認識及び把握プランニング用エキスパートシステムの研究」
（第２報，作業面・把握点決定及び把握経路の算定方法）
日本機械学会論文誌，Ｃ編，57巻536号，pp.244-252，(1990)

(5) 福田・長谷川：「マイクロマニプレータの制御」（第５報，任意方向を向き多セグメントに欠損を有する生物画像の推論認識法）
日本機械学会論文誌，Ｃ編，57巻536号，pp.205-212，(1990)

(6) T.Fukuda, O.Hasegawa :"Creature Recognition and Identification by Image Processing Based on Expert System",JSME International Journal, Series 3, Vol.33, No.2, pp.269-277 (1990)

(7) T.Fukuda, O.Hasegawa :"3-D Image Recognition System for Distorted and Cracked Objects", JSME International Journal, Series 3, 1992年 6月号　掲載予定

２．国際会議発表論文

(1) T.Fukuda and O.Hasegawa:"Expert System Driven Image Processing

for Recognition and Identification of Micro-organisms"
IEEE, Machine Intelligence and Vision, pp.33-38, Tokyo (1989)

(2) T.Fukuda and O.Hasegawa :"Reasoning Method for Visual Identification and Recognition of Living Organisms in Micro-Manipulator Control", The 20th International Symposium on Industrial Robots, pp.735-742, Tokyo (1989)

(3) T.Fukuda and O.Hasegawa : "3-D Image Processing and Grasping Planning Expert System for Distorted Objects", IEEE, Industrial Engineer's Conference, pp.726-731, Philadelphia, (1989)

(4) T.Fukuda and O.Hasegawa : "New Recognition and Identification Method for Micro-organisms by Expert System Driven Image Processing", Korean Automatic Control Conference, pp.1005-1010, (1989)

(5) T.Fukuda and O.Hasegawa : "Creature Recognition and Identification by Image Processing Based on Expert System",
IEEE, International Conference on System, Man and Cybernetics, pp.837-842, (1989)

(6) T.Fukuda, O.Hasegawa, H.Asama, T.Nagamune and I.Endo :
"A Monitoring System for Animal Cell Recognition"
The 1989 UK-Japan Biotechnology Seminar,
- Animal Cell Technology -, Oct., (1989), U.K.

(7) T.Fukuda, O.Hasegawa, H.Asama, T.Nagamune, I.Endo:"A Monitoring System for Animal Cell Cultures", Program of the Second Annual Meeting of Japanese Association for Annual Cell Technology, (Poster Session), Tsukuba, (1989)

(8) T.Fukuda, M.Ishizuka, O.Hasegawa, H.Asama, T.Nagamune, I.Endo: "Vision System for Animal Cell Recognition in a Bio Engineering Process", IEEE, Industrial Engineer's Conference, pp.552-557, (1990)

3．口頭発表論文

(1) 福田・長谷川：「生物の認識と同定に関する研究」
（第1報，画像処理を用いた生物の認識・同定法）
計測自動制御学会・学術講演会，SICE'88，pp.591-592，(1988)

(2) 福田・長谷川：「生物の認識と同定に関する研究」
（第2報，エキスパートシステムを用いた微生物の視覚認識の研究）
日本ロボット学会・第6回学術講演会，pp.561-562，(1988)

(3) 福田・浅間・長棟・遠藤・長谷川：「生物の認識と同定に関する研究」
（第3報，マイクロ・キャリア上の動物細胞の視覚認識法），日本機械学会，ロボティクス・メカトロニクス講演会，pp.22-23，(1989)

(4) 福田・浅間・長棟・遠藤・長谷川：「生物の認識と同定に関する研究」
（第4報，マイクロ・キャリア上の動物細胞の視覚認識・計測エキスパートシステム），計測自動制御学会・学術講演会，pp.515-516，(1988)

(5) 福田・長谷川：「画像処理エキスパートシステムを用いた不定形対象物の視覚認識の研究」，（第1報，歪み・欠損等を有する直方体状対象物の視覚認識法），日本ロボット学会・第7回学術講演会，pp.303-304，(1989)

(6) 福田・細貝・長谷川：「画像処理エキスパートシステムを用いた不定形対象物の視覚認識の研究」（第2報，歪直方体の把握経路の算定）

日本ロボット学会・第8回学術講演会, pp.30-31, (1990)

(7) 福田・浅間・長棟・遠藤・長谷川:「画像処理エキスパートシステムによるマイクロキャリア上の動物細胞の認識・計測法」
化学工学会, 金沢, (1990)

4. その他

(1) 理研シンポジウム

バイオプロセス・シンポジウムにて講演

「動物細胞視覚認識システムの試作」, 1990年 2月 2日,

於:理化学研究所

(2) 解説記事

「バイオ分野における視覚認識技術」

精密工学会・学会誌, 1990年 3月号, pp.17-20

謝　辞

本研究を進めるにあたり、多くの方々の御指導・御協力を頂きました。ここに記し、心より御礼申し上げます。

石塚満教授は、機械工学科出身の筆者に対し研究室への在籍を快諾して下さったばかりでなく、本研究を進めるにあたり、終始多くの御指導を頂きました。また個人的にも、筆者の仲人を御務め下さるなど、暖かい御配慮を頂きました。

Wiwat Wongwarawipat氏は本研究で利用したＶＩＴの設計者であり、ハード・ソフト両面において親身に御指導下さいました。氏の明晰で発想の豊かな頭脳と、ユーモアあふれる御人柄には、深い感銘を受けました。

李七雨氏は画像工学全般に渡る博学であり、幅広く知識を御与え下さり、また研究上の多くの相談に快く応じて下さいました。

横澤一彦客員助教授は、筆者の求めに応じ熱心に御指導下さいました。中でも人間の視覚情報処理過程に関し、情報工学・心理物理学・生理学などの広範囲に渡り文献や研究事例を御紹介下さり、また本研究遂行過程でも多くの有益なコメントを頂きました。

藤木真和君には、本研究の遂行にあたり多くの御協力を頂きました。

阿部明典氏，牧野俊朗氏・徐行倹氏には、入学当初より学業面・生活面などにおいて親切なアドバイス・御配慮を頂きました。

石塚先生の秘書の宇野由美子さん、川瀬清子さんの御二人は、主に事務処理に関し、御親切に多くの協力をして下さいました。

石塚研究室の職員の方々、諸先輩方、並びに現在籍者の皆さんには大変御世話になりました。石塚研究室で学んだ多くの事柄は、今後様々な局面でよい指針となってくれるでしょう。

日本学術振興会には、多大なる経済的援助を頂きました。

最後に、両親、並びに長谷川ふみかに感謝します。