

小規模並列プロセッサによる人工エージェント連続動作像の実時間生成

Real-time Synthesis of Continuous Motion Images of Artificial Agent Employing Small-scale Parallel Processors

土肥 浩*・石塚 満*

Hiroshi DOHI and Mitsuru ISHIZUKA

1. はじめに

映像や音声などのメディアをコンピュータによって高度処理し応用するメディア・コンピューティングの研究が進められている。われわれは、キーボードやマウスに換わる高次ヒューマンインタフェースとして、ビジュアル・ソフトウェア・エージェント (VSA: Visual Software Agent) を提案し、その具体化を進めている¹⁾。VSA は、コンピュータとユーザの自然なコミュニケーションの接面を形成する人工エージェントである。1枚の写真をもとにして、画像処理技術により現実感のあるエージェントがモニタ画面に登場する。その動きは、ユーザからのさまざまな入力により決定される。これまでに人の顔を持つ VSA、尻尾を振りながら泳ぐ金魚の VSA のプロトタイプが作られている²⁾。

本稿では、並列汎用マイクロプロセッサを用いて実現した、入力に対してほぼ自然な速度で応答するビジュアル・ソフトウェア・エージェントについて述べる。画像処理部は約11 cm×19 cm のボード 1 枚に実装されている。

2. 自然感の高い姿を有するビジュアル・ソフトウェア・エージェント

VSA は、変形可能な 3 次元立体モデルにテクスチャマッピング処理³⁾を施すことにより実現されている。テクスチャマッピング処理とは、複雑な「模様」を物体表面に転写する技術である。たとえば人の顔を持つ VSA を合成する場合、1枚の顔写真から顔の部分だけを数百の微小三角形に分割し、それを頭部の 3 次元立体モデルに張り付けていく。頭部の立体モデルは、やはり微小三角形により構成される 3 次元ワイヤフレームモデルである。この時、隙間ができないように三角形を伸縮させる。それぞれの模様は歪むが、三角形が小さければ、全体として連続性のある模様が再現できる。いま、立体モデル

が口を大きく開いているとすると、口を大きく開いた顔が合成できる。また微小三角形の張り付け方を変えることによって、その表情を変化させることもできる^{4),5)}。この立体モデルの変形から顔写真のテクスチャマッピングまでの処理を実時間で連続的に行うことができれば、1枚の写真に映っている人物が瞬きし、喋るわけである。顔写真は正面から撮影したものを使用しており、後ろを向くことはできないが、ある程度の範囲であれば任意の方向に顔を向けることも可能である。

金魚の VSA も、原理的には人の顔を持つ VSA と同じく、魚の 3 次元ワイヤフレームモデルに金魚の写真を張り付けることにより実現される。ただし金魚の場合には、任意の方向から見た画像を合成できなければならない。金魚の写真は、真横から撮影したものを 1 枚だけ使用した。見えない側は対称であると仮定している。金魚の立体モデルが紡錘形をしているため、頭から、あるいは尻尾からみた写真を用意しなくても立体感のある画像を合成することができる。これにより金魚の VSA は、尻尾を振りながら自由に泳ぎ回ることができる。

VSA は、さまざまな入力により立体モデルの移動、変形を制御することで、入力に対して応答するエージェントを実現しているのである。

テクスチャマッピングは非常に現実感のある画像合成が可能である反面、レンダリング等の処理に比べて計算機の負荷が大きい。このため複雑で数の多いポリゴンに対してテクスチャマッピングを行うには、専用ハードウェアを備えたグラフィック・ワークステーションや高速スーパーミニコンピュータなどが用いられてきた。これらのシステムは非常に高価であり、また装置もかなり大規模である。

われわれはこれまで、テクスチャマッピングによる実時間画像合成のために、独自に開発した VIT (Visual Interface to Transputers) 32台と汎用トランスピュータ 16台を接続した並列ビジュアル・コンピューティングシステム TN-VIT (Transputer Network with VIT) を使

*東京大学生産技術研究所 第3部

研究速報
用してきた⁶⁾。VITはトランスピュータを要素プロセッサとし、隣接するプロセッサ間に画像データ転送用の32 bit 単方向同期リンクを備えている。TN-VITは1フレームの画像を一台のVITに割り当てる時分割並列により30フレーム/秒のテクスチャマッピング画像生成を実現しているが、入力に対するレスポンス時間を小さくできない点がある。

また森島ら⁷⁾は顔全体の画像に対して、処理部分を鼻から下の口領域に限定することで、トランスピュータ20個を用いたネットワークでパイプライン処理することにより、毎秒およそ16フレームの表情合成を実現した。このシステムではマルチプロセッサによる非同期式演算パイプライン方式を採用しているが、負荷を均等にするために、各ステージのプロセッサの個数を静的に調節している。

3. 小規模並列ハードウェアによるVSAシステム

本稿は汎用の並列マイクロプロセッサを用いて並列処理を行うことにより、組み込み可能なサイズの小規模並列ハードウェアにより実現したVSAシステムについて記す。顔全体を処理対象とし、入力に対して自然な速度で応答することができる。要素プロセッサには、ハードウェア製作の容易さ、拡張性に優れている点で、トランスピュータ T800 を使用した。

トランスピュータ T800 は32 bit マイクロプロセッサであり、10 MIPS、1.5 Mflops (いずれも20 MHz クロック時) の演算能力を持つ。従来のマイクロプロセッサとは異なり、最初からマルチプロセッサ構成で使用することを前提に設計されている。プロセッサはそれぞれ4組のシリアルリンクを持ち、Point-to-Pointで他のプロセッサと接続する。シリアルリンクは双方向通信が可能で、転送速度は20 Mbits/sec、転送容量は単方向の場合1,740 Kbytes/secである。通常、共有メモリや共有バスは持たない。規格化されたボードにトランスピュータとメモリを実装したトランスピュータ・モジュール (TRAM) があり、マザーボードに挿すだけでプロセッサの拡張が行なえる。

並列処理においては、各要素プロセッサの負荷をできるだけ均等にし、実行効率を高く維持することが重要である。特にトランスピュータではプロセッサ間を20 Mbps のシリアルリンクで接続しているため、各プロセッサの相互依存関係を少なくし、プロセッサ間の同期通信のオーバーヘッドを小さくしなければならない。このためにそれぞれのプロセッサ (テクスチャマップ・エンジン) が、頂点移動から座標変換、テクスチャマッピングまでの一連の処理を行う。

テクスチャマップ・エンジンの各モジュールは、起動時に3次元ワイヤフレームモデルの頂点座標、面情報、テクスチャデータをローカルメモリにダウンロードし、WAIT状態になる。ワークステーションに接続されたルート・トランスピュータは入力 (たとえばXウィンドウ上のマウスの位置データなど) を受け取り、これをもとにして平行移動や回転移動の6自由度の変換パラメータ、瞬きや口の形の指示などのパラメータ、およびその他の制御情報を一つの packets にして、WAIT状態のテクスチャマップ・エンジンの一つに送る。合成される画像には順序関係が存在するので、テクスチャマップ・エンジンは順に循環的に使用される。ルート・トランスピュータは3次元ワイヤフレームモデルの頂点座標を直接操作することはない。合成された画像データはシリアルリンクを通してグラフィクスボードに送られる。シリアルリンクの転送容量は1フレーム全体の画像を送るのには十分ではないので、書き換えの必要がある部分データのみが送られる。

4. システム構成

本システムの構成を図1に示す。4個のトランスピュータ・モジュールからなるテクスチャマップ・エンジンを中心に、トランスピュータ6個で構成されている。SUNワークステーションとはVMEバスを介して接続されている。

テクスチャマップ・エンジンは、4個のトランスピュータ・モジュールで構成される。各モジュールは、それぞれトランスピュータ1個と2MBのローカルメモリで構成されている。1モジュールの大きさは、約5.5 cm×9.5 cm (Size 2 TRAM) である。モジュール4個でもダブルハイト標準VMEボード1/2枚 (約11 cm×19 cm) と非常に小さい。

グラフィクスボードは、テクスチャマップ・エンジンで合成された画像データをRGBのビデオ信号に変換する。グラフィクスボードと各トランスピュータ・モジュールとは、それぞれ1組のシリアルリンク (合計4組) で接続されている。グラフィクスボードは、トランスピュータ (T800)、メモリ2MB、フレームバッファ

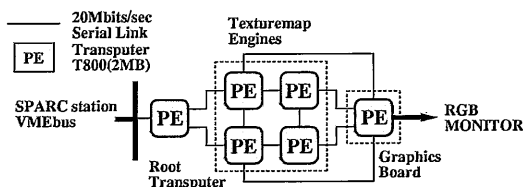


図1 システム構成

研究速報
 (デュアルポートメモリ) 1 MB, D/A コンバータ等で構成されており, これらを同じくダブルハイトの標準 VME ボード上に実装した. グラフィクスボードのトランスピュータはフレームバッファに直接アクセスできる. フレームバッファはデュアルポートメモリを使用しているため, トランスピュータによるアクセスとビデオタイミングコントローラのアクセスはほとんど競合しない.

トランスピュータのソフトウェア開発環境として PARALLELWARE 社の EXPRESS Ver. 3.2 を使用した. プログラミング言語は C 言語である. プログラムは Sun ワークステーション上でクロスコンパイルされ, VME バスを介して各トランスピュータにダウンロードされる. EXPRESS には, 各プロセッサの状態を解析できるツールが用意されている. プロセッサ間の通信履歴を経過時間とともに記録し解析できるツールを用いて, ソフトウェアの最適化を行った.

5. 実験結果

人の顔を持つ VSA では, 一つの要素プロセッサが 1 フレームの画像を担当する. この場合, ルート・トランスピュータからの 1 パケットが 1 フレームの画像生成に用いられる. また金魚の VSA では, 一つの要素プロセッサが一匹の金魚を担当し, 複数の要素プロセッサで合成された画像をまとめて 1 つのフレームが出来上がる. たとえば 4 匹の金魚が泳ぐ場合, 4 パケットで 1 フレームの画像生成が行われる. 合成された金魚は, 尻尾を振りながら画面の中を自由に泳ぎ回る. 遠くを泳いでいるときは小さく, 近くを泳いでいる時は大きく, 遠近感のある処理をしている.

尻尾を振りながら泳ぐ金魚の VSA の実行例を図 2 に示す. 合成されている背景が本システムである. 右側の

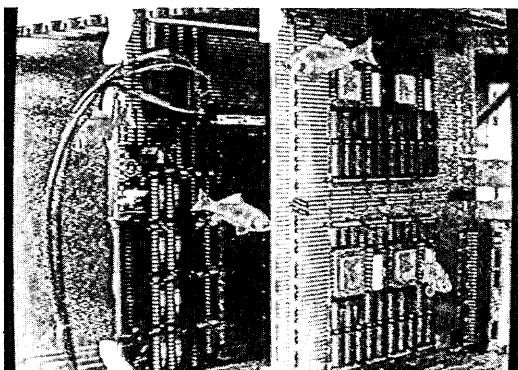


図 2 尻尾を振りながら泳ぐ金魚の VSA

ボードには 4 個のテクスチャマップ・エンジンが実装されており, 左側のボードがグラフィクスボードである.

表示速度は, 主に描画面積によって大きく変化する. たとえば人の顔を持つ VSA では, 1 枚のカラー画像をもとにして約 100×170 (pixel) の顔画像を合成する場合, プロセッサモジュール 4 個で, 平均およそ 12 frame/sec の性能を達成した. これは適当に瞬きをしながら, 上下左右 45 度以内で連続的に任意に首を振った場合である. またソフトウェアで, 静止画背景との合成をしている. 4 匹の金魚の VSA もほぼ同等の速度である. これよりテクスチャマップ・エンジンの入力に対するレスポンス時間は 400 msec 以下である. ヒューマン・インタフェースシステムとして, 入力に対してほぼ自然な速度で連続的に応答することが可能である.

6. ま と め

本稿では, 並列処理を用いることにより, 汎用の小規模なハードウェアでも入力に対して自然な速度で応答できるビジュアル・ソフトウェア・エージェントが実現できることを示した. 無人案内システムや知的符号化通信など非常に幅広い応用分野が期待できる. 今後, 動画像の実時間認識・生成技術, 知識ベース技術を集積した高次ヒューマン・インタフェースとして具体化していく予定である.

(1992年9月1日受理)

参 考 文 献

- 1) 石塚満, 土肥浩, 長谷川修, 藤木真和: 新しいヒューマンインタフェースに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). 生産研究, 本特集号, 1992
- 2) C-W. Lee, O. Hasegawa, W. Wongwarawipat, H. Dohi, and M. Ishizuka: Realistic Image Synthesis of a Deformable Living Thing based on Motion Understanding. *J. Visual Comm. Image Rep.*, 1991. 12
- 3) P. S. Heckbert: Survey of Texture Mapping. *IEEE CC&A*, pp. 56-67, 1986. 11
- 4) 崔昌石, 原島博, 武部幹: 顔の 3 次元モデルに基づく表情の記述と合成. 信学論, Vol. J73-A, No. 7, pp. 1270-1280, 1990. 7
- 5) F. I. Parke: Control Parameterization for Facial Animation. *Springer-Verlag*, pp. 3-14, 1991
- 6) O. Hasegawa, C-W. Lee, W. Wongwarawipat, and M. Ishizuka: Realtime Synthesis of Human-like Agent in Response to User's Moving Image. *11th ICPR*, 1992. 8
- 7) 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, Vol. J73-D-II, No. 10, pp. 1647-1654, 1990. 10