

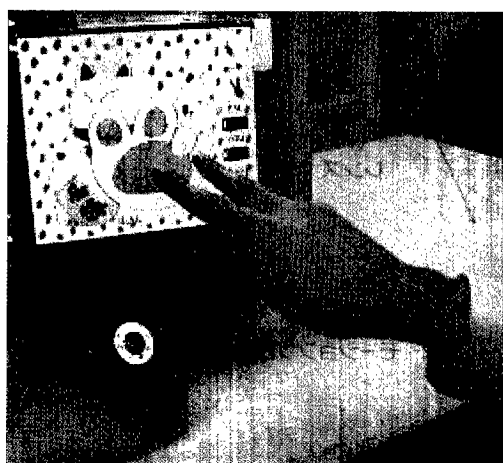
3

手振りで気持ちを伝えるインタフェース

～モーションプロセッサ™～

沼崎 俊一 (株) 東芝 研究開発センター

土井 美和子 (株) 東芝 研究開発センター



ポストGUI時代においては、ユーザが装置に縛られない自由なインタフェース環境が求められている。人間のジェスチャをコンピュータに入力する技術は、この1つの実現形態として期待されている。

CCDカメラで撮像した人物映像から動作環境を得る従来手法では、対象物体を切り出すために、大きなCPU負荷を必要とする画像解析を行うため、リアルタイムに処理できるシステムの構築は困難であった。筆者らは、新しい原理を用いて、背景から切り出された物体の画像を高速に取得できることを特徴とする動作入力装置、モーションプロセッサを開発した。本稿では、ジェスチャ入力の重要性を述べ、モーションプロセッサの原理・構成および応用範囲について述べる。

実	世	界	に
近	づ	く	イ
ン	タ	フ	エ
ー	ス	技	術

ディスプレイの呪縛からユーザを解放する
ポストGUI

コンピュータとのインタフェースは1960年代のバッチ処理、1970年代のコマンド入力CUI (Character User Interface)、そして現在我々が慣れ親しんでいるGUI (Graphical User Interface)と進化してきた。CUIではユーザがコマンドを覚えていないと、操作ができなかった。が、GUIでは、目の前の表示されているアイコンやメニューを選択するだけで操作でき、ユーザの負担を大いに軽減し、WSやPC普及の原動力となった。

しかし、残念ながら、GUIのモデルはデスクトップメタファ (机上の比

諭)であり、デスク→ディスプレイにユーザを縛りつけてきた。これに対し、ディスプレイの呪縛からユーザを解放し、いつでも、どこでも、誰とでも随意にコミュニケーションできることを目指すものがポストGUIである(図-1)。ユーザが相手にするのは、四角四面のディスプレイではなく、エージェント、ペットロボット、友達(ネットワークの向こうにいるかもしれないが)である。コンピュータがユーザから陽に見えることはない。

ユーザは机の前に座らずに、居間で寝ころびながら、あるいはベランダのベンチでリラックスしたままでもよい。ユーザの声は音声認識され、ユーザの手振り、身振りは画像解析され、ユーザがどの程度リラックスしているかは脈波センサなどによりセンシングされる。気圧などの外部状況、天気予報サイトからの情報で、外出には傘が必要なことを通知してくれる気のきいたインターフェイスがポストGUIである。ポストGUIの目指す自然で直感的なインターフェイス実現には、音声認識、画像解析、センサ、エージェントなどの要素技術を融合し、ユーザの意図に沿って、リアルタイムに応答できる対話技術の開発が必要である。

本稿では、まず、ユーザの意図に

沿うためのポストGUI実現のリアルタイムでの対話技術の1要素であるジェスチャ入力技術の動向について紹介する。次に、リアルタイムでのジェスチャ入力を可能とするために開発した物体を背景から切り出す新規入力デバイス(モーションプロセッサ)と、そのソフトウェア開発キット、これらを用いたアプリケーション例についてそれぞれ解説する。

ユーザが主役の手振り/身振り入力技術

人間同士では、テキスト、音声などの言語的(バーバル)チャンネルにより、意思伝達していると思われがちである。しかし、純粋にバーバルチャンネルのみで構成されるチャットや、電子メール、電子掲示板では、発話者の意図が明確に伝わらず、あらぬ誤解を生み、大論争、はては中傷合戦に発展する場合もある。

つまり、非言語的(ノンバーバル)チャンネルが伝える情報がないと、コミュニケーションが正しく行えないということである。ノンバーバルチャンネルとしては音声におけるピッチや抑揚などの周辺言語⁷⁾、表情、身振り、手振りのような身体動作、対話者同士の着席位置、距離などの場の情報などがある。Birdwhistell⁴⁾によ

れば、このようなノンバーバルチャンネルがコミュニケーションに占める割合は65~70%、Mehrabian⁵⁾によれば、93%であり、ノンバーバルチャンネルこそが、コミュニケーションの重要チャンネルといえる。このノンバーバルチャンネルの中でも身振り/手振りは、対話の主導権の移動を調整する調整子^{6), 9)}として、重要である。つまり、身振り/手振りによって対話の主導権を自分に持ってきたり、あるいは相手に譲ったりしながら、コミュニケーションを円滑に進めているのである。

MITのBobickら¹⁰⁾は時系列に動作を重ねたシルエットによるモーションテンプレートを用いて、ユーザが自ら選んだ音楽とインストラクタによりエアロビクス運動を行えるVirtualPATを研究している。ユーザの運動に応じて、「その調子」「もっと動いて」といった檄がバーチャルインストラクタから飛ぶ。また、ソニーの暦本らは、赤外光をマジックミラーの背面から照射し、ディスプレイに密着した手の動きだけを取り出し、コンピュータの表示内容と合成するHoloWall¹¹⁾を開発した。マイクロソフトでは、ルールベースで拍手などの手振り¹²⁾や、カメラ2台を使った全身像の追跡¹³⁾をする研究などが行われている。これらの研究¹⁴⁾では、

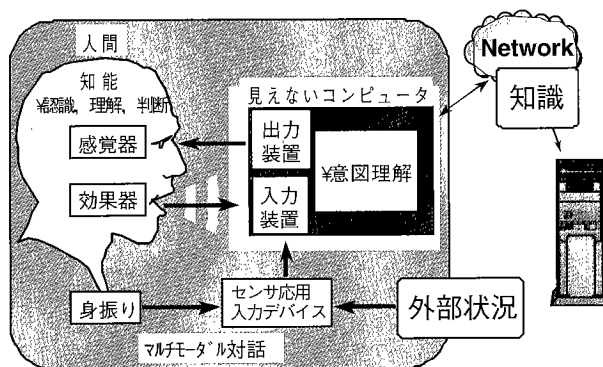


図-1 環境と対話する新しいヒューマンインタフェース (ポストGUI)

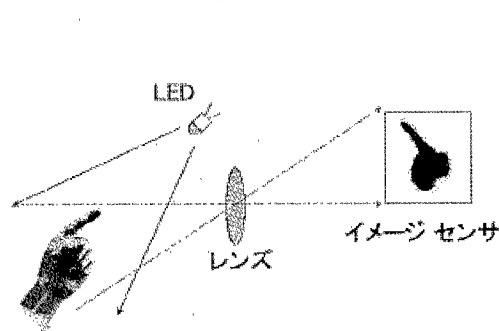


図-2 モーションプロセッサの原理

複数台のカメラを使ったり、画像処理用とアプリ用の複数台のコンピュータを使うなど、装置が大がかりになるのが問題になっている。

イメージセンサと並列演算素子を組み合わせた専用LSIで、物体認識を高速に行う手法¹⁵⁾も検討されているが、行える演算は算術演算など比較的単純なものに限られ、複雑な背景に対する誤認識などの問題は、画像解析を用いる手法と同様に存在する。

原理を用いた、動作入力装置モーションプロセッサの提案

ユーザに負担をかけずにジェスチャを入力するためには、非接触にユーザ情報を獲得するのが望ましい。しかし、CCDカメラなどを用いて取得した画像から手や身体などジェスチャとして認識すべき部分を背景などから切り出そうとすると、多大な計算量と処理時間を必要とする。このため、アプリケーションを含め、人間が許容できる最低限の応答時間⁸⁾ 0.1秒を満たすことができない。また、複雑な背景下では切り出しが適切に行われないことも多く、安定して人物を切り出すには、背景を単色にするブルーバックなどの手法が

必要であった。

著者らは、新しい原理を用いて、手などの対象物をリアルタイムで背景より切り出して取得することができる動作入力装置、モーションプロセッサを開発した¹⁾。モーションプロセッサは、従来の画像解析ベースの物体抽出と異なり、CPUに演算負荷をかけずにリアルタイムで物体画像を切り出すことができる。

図-2にモーションプロセッサの動作原理を示す。LEDから発光された近赤外光が物体で反射され、イメージセンサがその反射光を受光する。物体表面がLED光を拡散反射する場合、反射光の強さは、ほぼ物体までの距離に反比例する。したがって、遠い背景からの反射光はきわめて小さくなり、近くの物体の形状のみが反射光画像として撮像される。また、この物体の切り出しには画像処理演算を用いていないため、CPUに負荷をかけることなく高速にリアルタイムで取得できる。

このようにしてモーションプロセッサによって検出された手の画像の例を図-3に示す。背景は除去され、手形状のみが撮像されているのが分かる。

取得画像を3次的にポリゴン表示したものが図-4である。反射光量は、物体までの距離にほぼ反比例す

るので、物体のおよその3次元形状を表している。

モーションプロセッサを用いると、手の形状の切り出し、3次元空間内での手の動きの認識をリアルタイムに行える。これにCG技術などを組み合わせ、直感的な操作を実現した²⁾。

モーションプロセッサの仕様・構成

モーションプロセッサは、カメラとインタフェースカードから構成され、パソコンに接続して使用される。図-5はカメラの外観写真である。中央のレンズの周囲に、近赤外光を発するLEDを配している。またレンズの奥に受光部のイメージセンサチップ

実	世	界	に
近	づ	く	イ
ン	タ	フ	ェ
ー	ス	技	術



図-3 モーションプロセッサの取得画像の例

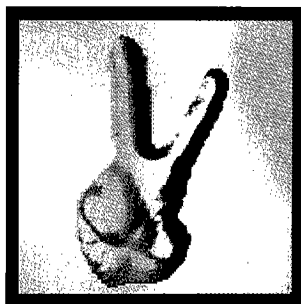


図-4 3Dポリゴン表示の取得画像



図-5 モーションプロセッサの外観

プがある。インタフェースカードは、イメージバッファ、PCIインタフェース機能などを搭載している。

表-1は、モーションプロセッサの仕様を表している。動きを見るための画像であるので、高い解像度は必要なく、64×64画素とした。各画素8bitのデータで反射光量が出力される(ただし、下位1bitは無効)。反射光を用いているため、撮像距離範囲は30～90cm程度である。ただし、原理的には強力な光源を使うことで距離を延ばすことは可能である。従来のように物体の切り出しにGPUの負荷を必要としないため、高速処理が可能であり、ビデオレートを上回る、毎秒50枚の画像を得ることができる。このことによって、動きの速いジェスチャも検出することができる。

パソコン側には、対象物体が切り出された画像がリアルタイムで送られてくる。これらの画像に対し、さまざまな演算を行い、手の位置や動き、ジェスチャの種類などを認識するアプリケーションソフトウェアを開発するための開発キットとして、モーションプロセッサSDK (Software Development Kit) を開発した³⁾。SDKは3次元空間における対象物の重心位置、重心速度、体積の計算、FFT演算といった低レベルの演算から、パターンマッチング、オプティカル・フロー抽出などの高次の処理を行うAPI (Application Programming Interface) 群から構成されている。

このSDKにより、ソフトウェア開

発者は、Microsoft® VisualBasic™あるいはVisualC++™を使って、アプリケーションを容易に開発できる。

モーションプロセッサを用いたアプリケーション

著者らは、モーションプロセッサSDKを用いて、いくつかのアプリケーションを開発した。そのいくつかを紹介する。

図-6は、じゃんけんゲームをしている図である。このアプリケーションでは、モーションプロセッサは手の前後方向の動きと形状を捉えている。前後方向の動きを捉えることで「じゃ～ん、け～ん、ぼん」のタイミングを検出している。「ぼん」のタイミングでの手の形状をパターン認識することで「グー、チョキ、パー」の形を認識して勝敗を判定する。

従来のジャンケンゲームであれば、機械のタイミングにユーザが合わせていた。しかし、モーションプロセッサを使うことで、ユーザのタイミングに機械(コンピュータ)が合わせることが可能となった。これにより、初めてコンピュータがユーザのタイミングを読むことが可能となり、気持ちを伝えることに一歩近づいた。

また、手の形や動きでさまざまな音を奏でるパーチャル楽器を作成した。これは、ドラムを手のひらで「打つ」動作をすることにより、さまざまな音を出させるアプリケーショ

ンである。モーションプロセッサにより、「打つ」速さを取得し、その速度に合わせて、音量を変えたり、音程を変えたりすることができる。パーチャル楽器はさらなる発展性を秘めており、「打つ」時の手の形を認識し、それによって音色(シンバルの音、太鼓の音、といった具合に)を変えたり、あるいは中間的な手の形でそれぞれの音色をブレンドしたりするとより面白い楽器ができると期待できる。

ほかにも、手振りでコミュニケーションがとれるパーチャルペット、手でこねる感覚で自由に画像を加工できるインタフェースなど、さまざまな応用アプリケーションを開発した。

また、ポストPC時代といわれる現代では、コンピュータはその姿を隠し、さまざまなところに存在する。情報家電や情報KIOSK、車載情報機器などの新しい装置、あるいはそれらに実装されるエージェントなどのインタフェースとして大きな可能性を秘めている。

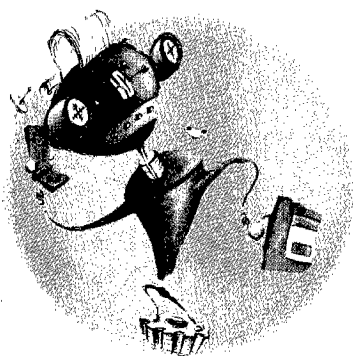
モーションプロセッサの効果

相手の反応が鈍いと、手振り、身振りで気持ちを伝えることも難しくなる。モーションプロセッサは適切な反応で気持ちを伝えることを可能とした。

図-7の左はCCDカメラを使ったジェスチャ認識システムの構成、同図

分解能	64 × 64 pixels
距離深度	128階調
撮像レンジ	30 ～ 90 cm
動作レート	50 frames / sec
カメラサイズ	幅75×高75×奥行78mm
対応VF	PCI
対応OS	Windows98/95

表-1 試作機の仕様



右はモーションプロセッサを用いたジェスチャ認識システムの構成である。CCDカメラを用いる場合は、その画像中にある背景や関係ない物体を取り除くための画像解析処理が必要となる。この処理がCPUに大きな演算負荷をかけるため、リアルタイムでジェスチャを認識するシステムを構成するのは非常に困難である。図-7の右側に示すようにモーションプロセッサは、この物体切り出しがモーションプロセッサ内部で行われているので、CPUに負荷をかけない。CPUは切り出した物体の認識処理(動きの認識、形状の認識)のみに注力できる。この結果、Pentium 300MHz程度のマシンパワーでも毎秒50枚の画像を処理することが可能であり、CCDカメラ画像を画像解析してジェスチャを認識する方法に比べ、非常に高速である。また、認識した結果は図-4のように立体的に表示できるので、遠隔地の人と、ネットワークを通じてジェスチャゲームやコミュニケーションを行う際に、お互いに相手の手の形を見ることができ、より臨場感を増すことができる。

後の展開

背景から物体の画像を、高速・安定的にかつ安価に切り出すモーションプロセッサにより、わずかな計算機負荷で、人間の身振りや手振りを用いるインタフェースの構築を可能とした。

モーションプロセッサ試作機は、1999年春より、大学や企業にてさまざまなアプリケーションプログラムの開発、評価を行っていただいている¹⁶⁾。さらに、現在は、SDKの高機能化、カメラの高解像度化に向け、開発を行っている。

今後は、小型化、モジュール化を図り、コンピュータが前面に出てこない情報家電やペトロボットなど、直感的にユーザの意図を伝えるインタフェースが必須な分野へ展開を図っていく。

参考文献

- 1) Numazaki, S. et al.: A Kinetic and 3D Image Input Device, Proceedings of CHI'98 (Summary), pp.237-238 (1998).
- 2) Umeki, N. et al.: A Motional Interface Approach Based on User's Tempo, Proceedings of CHI'99 (Extended Abstracts), pp.23-24 (1999).
- 3) 山内他: モーションプロセッサ用ソフトウェア開発キット, 情報処理学会第58回全国大会 No.4, pp.155-156 (1999).
- 4) Birdwhistell, R.L.: Kinetics and Context, University of Pennsylvania Press (1970).
- 5) Mehrabian, A.: Communication without Words, Psychology Today, Vol.2, pp.52-55 (1968).

- 6) Ekman, P.: Three Classes of Nonverbal Behavior, Aspects of Nonverbal Communication (ed.), Swets and Zeitlinger (1989).
- 7) Yoshimura, T. et al.: Pitch Pattern Clustering of User Utterances in Human-Machine Dialogue, Proceedings ICSLP 96, IEEE, Vol.2522, pp.837-840 (1996).
- 8) Card, S.K. et al.: The Psychology of Human Computer Interaction, Lawrence Erlbaum Associates (1983).
- 9) Wexelblat, A.: An Approach to Natural Gesture in Virtual Environments, ACM Trans. on Computer-Human Interaction, Vol.2, No.3, pp.179-200 (1995).
- 10) Davis, J.W. and Bobick, A.F.: VirtualPAT: A Virtual Personal Aerobics Trainer, Workshop on Perceptual User Interfaces (PUI'98) (1998).
- 11) Rekimoto, J. and Matsushita, N.: Perceptual Surfaces: Towards a Human and Object Sensitive Interactive Display, Workshop on Perceptual User Interfaces (PUI'97) (1997).
- 12) Cutler, R. and Turk, M.: View-based Interpretation of Real-time Optical Flow for Gesture Recognition, Proc. 1998 IEEE Conference on Automatic Face and Gesture Recognition (1998).
- 13) Jojic, N., Turk, M. and Huang, T.: Tracking Articulated Objects in Stereo Image Sequences, Proc. IT Workshop on Detection, Estimation, Classification, and Imaging, Santa Fe, NM (1999).
- 14) 1998 Workshop on Perceptual User Interface (PUI'98) (1998).
- 15) 石川正俊: スーパービジョンチップと今後の応用, Proceedings of STARC Symposium 99, pp.99-106 (1999).
- 16) モーションプロセッサに関するホームページおよび問合せ先:
http://mptech.infortech.co.jp/
E-mail:support@mptech.infortech.co.jp
(平成11年10月27日受付)

実	世	界	に
近	づ	く	イ
ン	タ	フ	エ
ー	ス	技	術

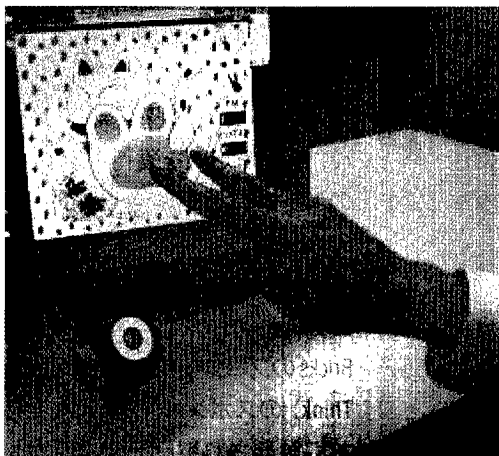


図-6 ジャンケンゲーム

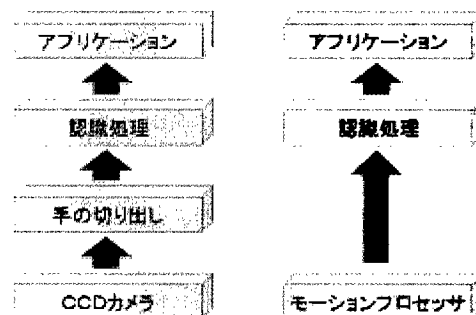


図-7 従来手法との処理方法の違い

