

Cell 画像処理アプリケーション

Cell Image Processing Applications

近藤 伸宏

■ KONDOH Nobuhiro

檜田 和浩

■ HIWADA Kazuhiro

谷口 恭弘

■ TANIGUCHI Yasuhiro

風間 久

■ KAZAMA Hisashi

Cell Broadband Engine (CBE) のパフォーマンスを引き出すためには、一つのチップに8個搭載されたSPE (Synergistic Processor Element) を有効に活用する必要がある。

東芝は、アプリケーションレベルでのCBEのパフォーマンスを実証するために、Cellリファレンスセットを利用して、主に画像処理やコンピュータグラフィックス処理を行うアプリケーションを開発した。この開発により、化粧シミュレーション、髪型シミュレーション、ジェスチャ入力、顔認識などのアプリケーションがCBE上で実装され、CBEのパフォーマンスの高さが実証された。

One of the key points to fully utilize the high performance of the Cell Broadband Engine (CBE) is to effectively use the eight Synergistic Processor Elements (SPEs) in a chip. To demonstrate the high performance of the CBE, Toshiba has developed sample applications that mainly deal with image data processing or computer graphics, based on the Cell reference set. We have built applications for makeup simulation, hair-styling simulation, gesture recognition, and identification by face recognition, thereby confirming the excellent performance of the CBE.

1 まえがき

画像処理アプリケーションは、大容量のデータを扱い、また、リアルタイム性が要求される、演算リッチなアプリケーションである。従来は、処理範囲を狭くしたり、処理するフレームレートを落としたりするなど、性能を犠牲にした研究が行われてきた。Cell Broadband Engine (CBE) の登場により、実用的な速度と精度で動作する画像処理アプリケーションを開発することが可能になった。

ここでは、Cellリファレンスセットを用いて作成された3種類のアプリケーションについて述べる。最初に“デジタルかがみF-TYPE”という、化粧シミュレーション及び髪型シミュレーションの技術を利用したシステムについて、次にジェスチャ認識技術を利用したユーザーインターフェースについて、最後に顔認識技術を利用した認証システムについて述べ、CBEのパフォーマンスを引き出すとこんなことができるのか、という新しい世界の一端を紹介したい。

2 デジタルかがみF-TYPE

デジタルかがみF-TYPE (図1) は、化粧や髪型の変更を擬似体験できるシステムである。このシステムでは、顔トラッキング技術、3次元(3D)化粧シミュレーションや髪型シミュレーションの技術が用いられている。これらの技術は画像処理やコンピュータグラフィックス処理をベースにしたもので



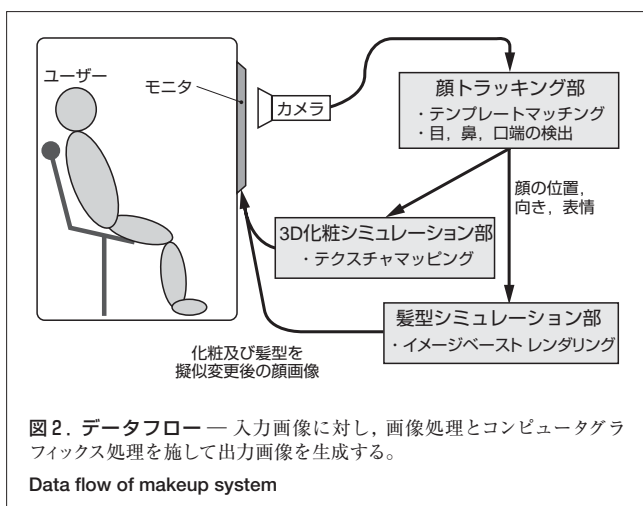
図1. デジタルかがみF-TYPE — 利用者は、化粧や髪型の変更を擬似的に体験できる。

Digital Mirror F-TYPE

ある。

2.1 全体構成

デジタルかがみF-TYPEにおけるデータの流れを図2に示す。まず、カメラで撮影された利用者の顔の映像がCellリファレンスセット上で動作する顔トラッキング部に入力され、顔の位置、向き、表情が認識される。その結果は3D化粧シミュレーション部と髪型シミュレーション部に入力され、あらかじめ用意した化粧や髪型の画像とカメラで撮影した顔の画像が自然に見えるように合成され、利用者の前に設置されたモニタに出力される。3D化粧及び髪型のシミュレーションは独立に行われており、化粧と髪型両方のシミュレーション



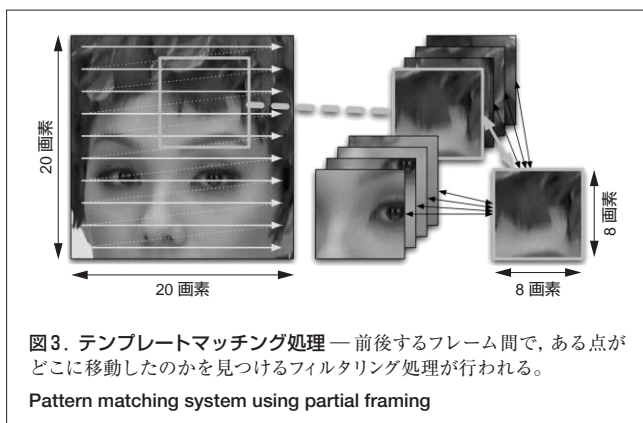
を同時に行うことができる。

2.2 顔トラッキング

顔トラッキング部では、最初に、目、鼻、口を自動認識し、顔の位置を把握する。次に、テンプレートマッチングという処理によって、顔全体から抽出した、およそ500個の点の動きを追跡する。この点の動きから、顔の向きや表情を認識することができる⁽¹⁾。

テンプレートマッチング処理は、動画像において、ある点が次のフレームでどこに移動したのかを認識するために利用される処理である。これは、図3に示すように、ある点の周囲にある8×8画素の画像が、周囲20×20画素の領域のどこに移動したかを判定するための処理で、8×8画素の画像どうしのフィルタリング処理が行われる。したがって、1回のテンプレートマッチングでは、169回のフィルタリング処理が行われることになる。更に、この処理は、追跡を行う500個の点すべてに対して行うため、1フレームの処理では84,500回のフィルタリング処理を行う必要がある。

シングルコアのプロセッサでは、すべての処理を一つのコアで順次実行しなければならない。ところが、この処理は500個の点について、どの順序で処理してもよく、並列に処理



することが可能である。このような処理では、CBEのSPEを用いると、最大8処理を同時に行うことができる。更に、この処理ではSIMD (Single Instruction Multiple Data) 命令を使用できるため、並列処理を行わない場合に比べると、32倍の処理性能を達成することができる。

2.3 3D化粧シミュレーション

3D化粧シミュレーション部では、テクスチャマッピングと呼ばれるコンピュータグラフィックス処理を行っている。このシステムでは、利用者の顔の表面形状をおよそ6,000個のポリゴンと呼ばれる三角形で表現しており、その表面形状に化粧の色をのせることで、あたかも化粧をしたかのような画像を作り出している。

テクスチャマッピングは、それぞれのポリゴンがどのような色で表示されるかを計算するための処理で、市販のコンピュータでは、グラフィックプロセッサと呼ばれる専用のLSIを利用して処理していることが多い。デジタルかがみF-TYPEでは、SPEを利用して、この機能をソフトウェアで実現した。この処理はポリゴンごとに並列に処理できるため、デジタルかがみF-TYPEでは8個のSPEを利用して並列に処理している。

2.4 髪型シミュレーション

髪型シミュレーション部では、イメージベースレンダリング処理を行っている。これもコンピュータグラフィックス処理であり、このアプリケーションでは、入力画像と過去に撮影済みの髪型画像という実写どうしの融合に使用している。

髪型シミュレーションでは、変更したい髪型の画像データベースを用いる。このデータベースには、いろいろな髪型のそれぞれに、いろいろな方向を向いた画像が登録されている。しかし、利用者が現在向いている方向とまったく同じ方向を向いている画像はないため、入力画像の顔の向きに近い複数の画像を画像データベースから選択し、現在の顔の向きに合った髪型の画像を生成している。このような画像を用いた大容量データを必要とする処理には、CBEに実装されているEIB (Element Interconnect Bus) のような高速内部バスや、XIO^(注1) のような高速メモリインタフェースが有効であり、データ転送が演算処理の妨げにならないように制御することができる。

2.5 処理結果

処理結果を図4に示す。デジタルかがみF-TYPEによって、化粧や髪型の擬似的な変更体験ができていていることが示されている。

このアプリケーションでは、2.2～2.4節で述べたような処理を、1秒間に30フレーム分処理しなければならない。しかも、ユーザーの顔や表情に追従するインタラクティブなアプリケーションのため、リアルタイムに処理する必要がある。東芝は、

(注1) XIOは、Rambus社の登録商標。



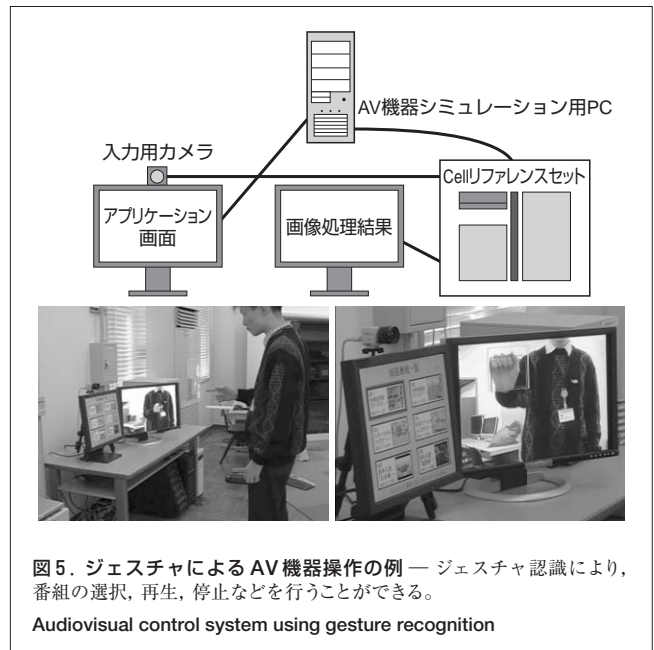
これらを一つのCBEプロセッサで処理できることを実証した。このアプリケーションの開発により、あらためてCBEのコンピュータアーキテクチャの有効性が実証された。すなわち、SPEに搭載されたローカルストレージを有効に活用することで、演算リッチなリアルタイムアプリケーションを動作可能なことが確認された。また、アプリケーションの中で並列に処理できる部分は8個のSPEに処理をさせ、制御はPPE (PowerPC Processor Element)で処理するというプログラミングモデルにより、CBEは、他のプロセッサの追従を許さない性能を発揮することが示された。

3 ジェスチャ認識を用いたユーザーインターフェース

ジェスチャ認識を用いたユーザーインターフェースの一例として、AV機器をジェスチャによって操作する試作システムを作成した(図5)。このシステムでは、カメラから入力された画像をCBEで処理することによって、画像中の手の形状や動きを検出し、これらの情報をAV機器シミュレーション用パソコン(PC)にネットワークを使って転送する。AV機器シミュレーション用PCでは、受け取った手の形状と動きをコマンドに変換して機器画面を操作する。

このようなアプリケーションでは、高速かつ安定的に手の形状や動きを検出することが重要となる。画像中から手の領域を検出する手法は数多く提案されているが、ここでは、CBEへの移植を考慮して、並列度の高いプログラムを作成可能なAdaBoost⁽²⁾と呼ばれる手法を用いた。この手法では、手を検出するのに効果の高い検出器の形状を多数の手画像を用いて学習し、学習によって得られた検出器を用いて、画面中から手を検出する。

実際の利用場面では、画像中の手の大きさはシーンによって大きく異なるため、学習した検出器のサイズを複数変更することによって対応した。一般に、形状やサイズは互い

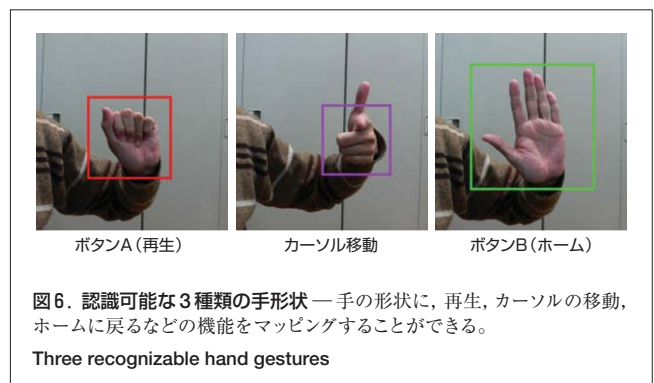


に独立した特徴と考えられるので、CBEで実行する場合には、異なったSPEで並列して処理することが可能となる。

今回試作したシステムでは、図6に示す3種類の手の形状を判別することが可能である。3種類の手形状を用いることによって、選択画面中のカーソルを移動させて自分が見たい映像を選択し、ボタンAを押すことによって映像の再生をする。また、ボタンBを押すことによって、いつでもホーム画面に戻ることが可能となっている。

ジェスチャ認識の手法としてCBEの特性を生かすことができるアルゴリズムを用いた結果、Cellリファレンスセットで試作したシステムでは、1秒間に20~30枚の画像処理を実現できた。このシステムで用いたAdaboost法のように、並列性が高く、CBEの特性を生かすことのできるアルゴリズムでは、周波数や演算器の個数比以上の高速化が可能となる。

今回の試作システムは、展示会のような多くの観客がいる中でも安定に動くように、全画面の中から手を検出しているが、一般的な用途を考えた場合には、手の検出位置や大き



さなどはある程度限定することが可能である。この場合には、数個のSPEだけで同様の処理を実現することが可能となり、顔認識など、他のアルゴリズムと協調したインテリジェントなシステムの構築が可能になると考えられる。

4 高速な顔認識エンジン

顔認識は研究・開発が盛んな分野であり、入退出管理装置への応用など実用化も進んでいる。顔認識は、セキュリティ以外にも幅広い応用が可能な技術である。例えば、AVアプリケーションと組み合わせることで、ドラマから特定の登場人物のカットを抽出したり、特定の歌手の映像を集めたりする応用などが考えられるし、視線方向や顔向きが識別できれば、機器のスイッチとしての応用も考えられる。

顔認識方式には、顔パーツの位置情報を用いる方法と、画像そのものをパターンとして用いる方法があるが、当社では後者を利用している⁽³⁾。この方式では、顔領域の検出とパターン識別の二つのステップを実行する。

顔領域の検出は、入力した画像内で顔の存在位置を検出するステップであり、空間並列性がある。パターン識別は、顔画像を特徴ベクトルというデータに変換し、辞書と呼ばれる個人ごとにクラス分けされた統計量と比較するステップであり、クラスごとの並列性がある。どちらも並列演算が可能であり、CBEのような並列演算プロセッサで高速化が可能である。

以上のことから、当社は、CBEの能力を生かした汎用可能な顔認識エンジンを開発している。試作システムとして、カメラ入力映像に対する同時多人数の顔認識システムを実装した。

処理結果出力を図7に示す。

SPEの並列性を生かすために、試作システムでは画像領域



図7. 試作システムの出力画面例 — 同時に複数の人の認識ができています。画面中の緑色の枠はシステムが検出した顔の位置と大きさを、黄色の小さい丸は目の位置を表している。

Example of test system display

を部分画像に分割し、それからをSPEに並列に処理させる構成をとった。従来型プロセッサでリアルタイム処理を実現するためには、探索空間を限定したり、人物を一人に限定してトラッキングしたりするなど、処理量を抑える必要があった。一方、試作システムでは探索範囲を画面全体に広げることができ、複数人を同時に検出・認識することができるようになった。

5 あとがき

画像処理研究の歴史は長いですが、膨大なデータ転送や演算性能を必要とするために、リアルタイムで処理することが難しかった。ここまで紹介してきたように、CBEを利用すると様々な画像処理をリアルタイムに行うことができるようになる。画像処理を利用すると、カメラを利用したヒューマンインタフェースを実現でき、従来なかったような機器を作ることにも可能となる。

当社は、画像処理アプリケーションの実装を通して、CBEの高度な演算性能を実証した。今後は、CBEと画像処理研究のコラボレーションから、新しい世界を切り開いていきたいと考えている。

文 献

- (1) Hiwada, K., et al. "Mimicking Video: Real-Time Morphable 3D Model Fitting". Proc. 10th ACM Symposium on Virtual Reality Software and Technology. Osaka, 2003-10. ACM. ACM Press, p.132 - 139.
- (2) Viola, P.; Jones, M. "Rapid object detection using a boosted cascade of simple features". Proc. of Computer Vision and Pattern Recognition, Kauai Marriott, Hawaii, 2001-12. IEEE Computer Society. p.511 - 518.
- (3) 小坂谷達夫, ほか. 制約相互部分空間法を用いた顔認識システムの開発と評価. 情報処理学会論文誌. 45, 3, 2004, p.951 - 959.



近藤 伸宏 KONDOH Nobuhiro

セミコンダクター社 ブロードバンドシステム LSI 事業統括部
ブロードバンドシステム LSI 応用技術部。並列処理、画像処理
システムの研究・開発に従事。
Broadband System LSI Div.



檜田 和浩 HIWADA Kazuhiro

セミコンダクター社 ブロードバンドシステム LSI 事業統括部
ブロードバンドシステム LSI 応用技術部。並列処理、画像処理
システムの研究・開発に従事。
Broadband System LSI Div.



谷口 恭弘 TANIGUCHI Yasuhiro, D.Eng.

研究開発センター マルチメディアラボラトリー 研究主務, 工
博。コンピュータビジョン関連の研究・開発に従事。電子情
報通信学会, ロボット学会会員。
Multimedia Lab.



風間 久 KAZAMA Hisashi

セミコンダクター社 ブロードバンドシステム LSI 事業統括部
ブロードバンドシステム LSI 開発センター 参事。画像処理,
画像認識, コンピュータビジョンの研究・開発と事業推進活動
に従事。電子情報通信学会会員。
Broadband System LSI Div.