

Cell Broadband Engine™ を用いた フェイス トラッキング

Real-Time 3D Face Tracking with Cell Broadband Engine™

檜田 和浩

■ HIWADA Kazuhiro

Cell Broadband Engine™ (注1) (以下、Cell/B.E.と略記) は、画像処理などのマルチメディアデータのリアルタイム処理に向けた高性能プロセッサである。このCell/B.E.の高い性能を引き出すには、複数搭載されたSPE (Synergistic Processor Element) の活用が不可欠である。

仮想鏡システム“デジタルかがみF-TYPE”は、Cell/B.E.の高い性能を生かした画像処理アプリケーションである。ここで用いられている、実写とコンピュータグラフィックス (CG) の自然な合成を可能にするフェイス トラッキング技術は、従来リアルタイム化の難しい処理であった。東芝は、高速化や並列化に向けた新しいアルゴリズムを提案し、Cell/B.E.に向けた最適化を行った。この結果、1フレーム10 msというリアルタイム動作に十分な性能を達成し、Cell/B.E.のリアルタイム画像処理におけるパフォーマンスの高さが実証できた。

The Cell Broadband Engine™ (Cell/B.E.) is a high-performance real-time processor that effectively handles multimedia data. However, to elicit the high performance of the Cell/B.E., it is necessary to use synergistic processor elements (SPEs) in a meaningful way.

Toshiba's "F-TYPE Digital Mirror" is a computer-vision application that exploits the high performance of the Cell/B.E.. This system uses a face-tracking algorithm, which was previously difficult to process in real time. Toshiba successfully developed an advanced face-tracking algorithm suitable for multicore processing and implemented it on the Cell/B.E.. As a result, a practical system has been realized that is able to process face tracking at a speed of 10 ms per frame.

1 まえがき

Cell Broadband Engine™ (以下、Cell/B.E.と略記) は、マルチメディアデータの高速でリアルタイムな処理に適したプロセッサである。マルチメディア処理のなかでも画像処理アプリケーションは、大容量の画像データをリアルタイムに扱う必要があることから、Cell/B.E.の応用先として期待されている。

東芝は、フェイス トラッキングという画像処理手法を用いた仮想鏡システム“デジタルかがみF-TYPE”をCellリファレンスセット上で開発した。このシステムは、利用者の顔画像を解析し、リアルタイムに化粧や髪型などを変更して表示できる。このようなシステムは、従来リアルタイムに動作させることが難しく、精度やフレームレートなどを犠牲にする必要があったが、当社はリアルタイム動作に向けた新しいアルゴリズムを提案し、Cell/B.E.を用いた効率良い実装を行うことで、精度などを犠牲にすることなくリアルタイムにシステムを動作させることに成功した。

ここでは、Cell/B.E.へのソフトウェア最適化の事例をとおり、Cell/B.E.の可能性の一端を示すとともに、今後のCell/B.E.向けソフトウェア開発の一助としたい。

(注1) Cell Broadband Engineは、(株)ソニー・コンピュータエンタテインメントの商標。

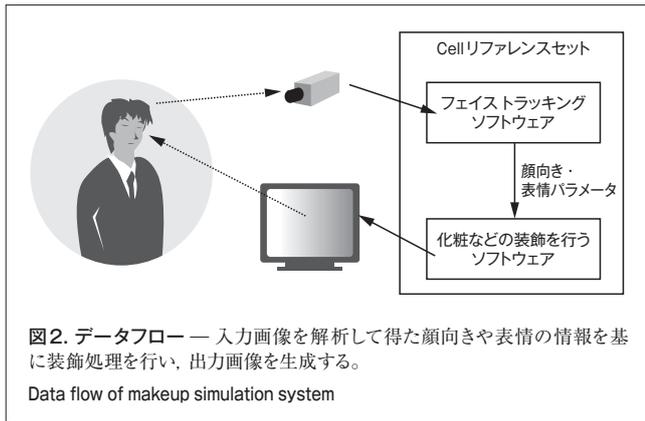
2 仮想鏡システム“デジタルかがみF-TYPE”

このシステムは、利用者が仮想的に化粧や髪型の変更を試すことができる仮想鏡システムである (図1)。このシステムにおけるデータの流れを図2に示す。まず、カメラから取り込まれた利用者の映像はCellリファレンスセットに送られ、Cell/B.E.によりフェイス トラッキング処理が行われる。その結果得られた姿勢や表情などの情報を用いて、化粧や髪型の変更と



図1. 仮想鏡システム“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。

“F-TYPE Digital Mirror” makeup simulation system



いった映像装飾処理を行い、被験者の前のディスプレイに表示する。この際、画像を左右反転し鏡像を表示することで、利用者は鏡を見るような感覚で、化粧や髪型の変更を試すことができる。

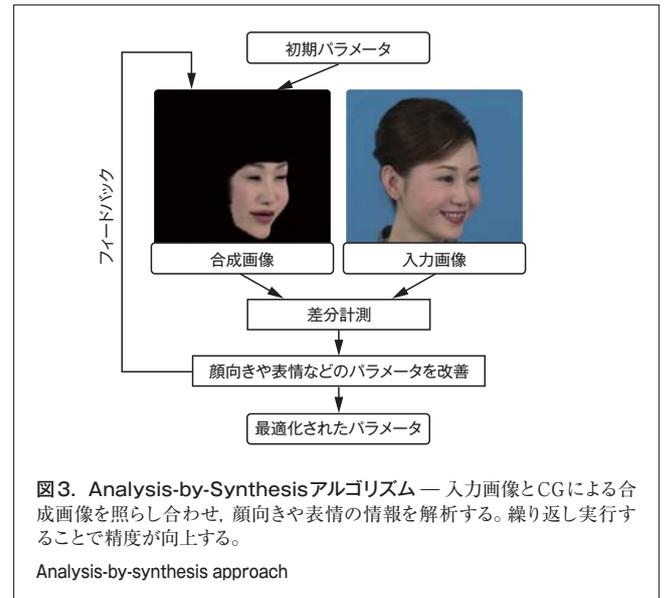
この、画像入力、フェイス トラッキング、及び化粧などの装飾処理を1秒間に30回繰り返し行うことで、リアルタイム動作する仮想鏡システムが実現できる。なかでもフェイス トラッキング処理には、画素単位の高い精度が要求されるため、リアルタイムに処理するのが難しかったが、当社は、高速化や並列化に向けたアルゴリズムを提案し、Cell/B.E.への効率良い実装を行うことで、リアルタイム動作を実現した。

3 フェイス トラッキング アルゴリズム

フェイス トラッキング技術は、コンピュータによる視覚情報処理分野において広く研究され、様々な手法が提案されている。なかでも、“Analysis-by-Synthesis (AbS)”手法は、実写とCGの合成といった高い精度を要求される場面向いたアルゴリズムである。当社は、この手法を拡張し、リアルタイム化に向けたアルゴリズムを提案した⁽¹⁾。

3.1 Analysis-by-Synthesis手法

AbS手法は、入力画像を模した画像を合成しつつ、その合成画像と入力画像を照らし合わせることで、詳細な解析を行う手法である(図3)。AbS手法では、まず顔モデルを用いて利用者の画像を合成する。次に、この合成画像と入力画像を比較し、それらの差分を計測する。この差分を最小化するように顔の姿勢や表情などのパラメータを最適化することで、より入力画像に近い利用者の画像を合成する。更に入力画像に近い合成画像を用いることで、より詳細な最適化が可能になるため、合成、比較、及び最適化の手順を繰り返すことで、高い精度の解析が可能になる。AbS手法は、もとより入力画像を模した画像を生成する手法であるため、結果として得られる画像が入力画像と自然に合成しやすいという特徴があり、仮想鏡システムの構築に向いている。しかし、合成、比較、及び最適化の



手順を繰り返すという重い処理を行うため、リアルタイムに動作させるにはいくつかの工夫が必要であった。

3.2 リアルタイムトラッキング手法

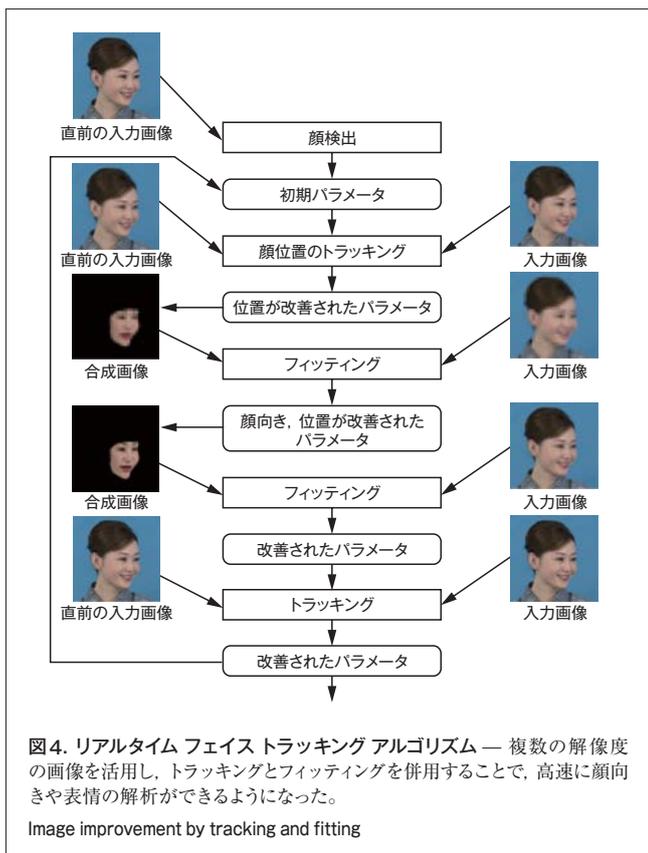
AbS手法を高速化するため、次の三つのアイデアを新たに導入した。

- (1) Coordinate-Oriented Error Minimization
- (2) Coarse-to-Fine Fitting
- (3) Fitting 手法と Tracking 手法の併用

Coordinate-Oriented Error Minimizationは、合成画像と入力画像の差分を計測しパラメータを最適化する手法の一つである。この手法では、差分の計測に、合成画像上の特徴点の位置と入力画像上の対応点の位置との相対位置ベクトルを利用する。例えば“左目の位置が10画素ずれている”といった画像のずれ情報が得られる。この差分計測を数百から数千点について行い、この結果をpseudo inverseという一般的な行列演算手法によって解析することにより最適化を行う。行列演算による最適化は比較的少ない演算量で実現できるため、差分演算の高速化が処理全体を高速化するうえで重要になる。

Coarse-to-Fine Fittingは差分演算を高速に行うための手法である。AbSは繰り返し最適化を行う手法であるが、この初期段階では少ない数の特徴点を用い、終盤では多くの特徴点を用いるのがCoarse-to-Fine Fittingの特徴である。初期段階では画像間のずれが大きいため広い範囲を探索する必要があるが、終盤では画像間のずれが少なくなり、探索範囲を狭くできるため多くの特徴点を用いても演算量を抑えることができる。結果として少ない演算量で精度を落とさずに差分演算を行える(図4)。

Fitting手法とTracking手法の併用により、AbSにとられない手法の拡張を行った。“Fitting”はAbS手法を意味する言葉であり、“Tracking”はまた別の解析手法である。Tracking



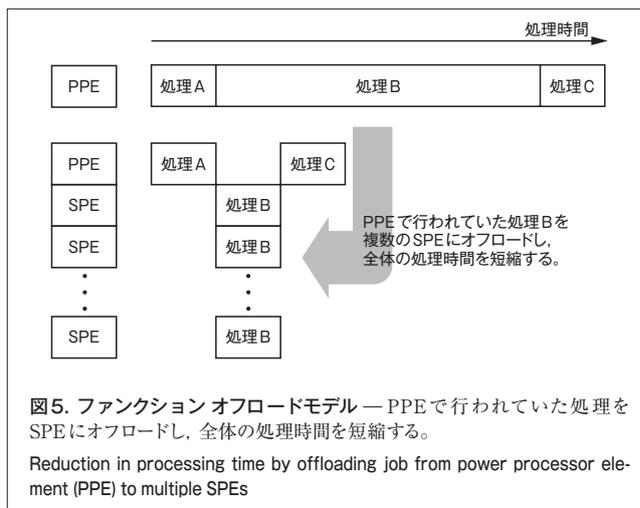
は、前のフレームの画像と現在のフレームの画像を比較することで、顔など対象の動きを推定する手法であり、前のフレームの結果を生かして効率的な推定ができるという特徴がある(図4)。当社は主にAbS手法に基づいたFitting処理を用い、補助的にTracking処理を行うことで、演算量の削減と高い精度を両立させた。

このような最適化の結果、提案したアルゴリズムは市販のハイエンドパソコン(PC)においてリアルタイムに近い、1秒間に15フレームの性能を実現した。

4 CBEへの実装

市販のPCでリアルタイムに近い動作が可能になったが、自然な動画画像を生成するには最低でも1秒間に30フレーム程度、1フレーム当たり33msで処理できることが望ましい。更に、フェイス トラッキングをアプリケーションの一部として使用するためには、10ms以下程度の短い時間で処理できることが必要になる。当社は、以下に述べるCell/B.E.への最適化を行うことでこの性能向上を実現した。

Cell/B.E.を用いた高速化のための開発モデルは多種あるが、当社はその中でも低い開発コストで高速化の効果が得られる“ファンクション オフロード”⁽²⁾という開発モデルを用いた(図5)。開発にあたり、まずCell/B.E.の持つ汎用プロセッサ



であるPPE (Power Processor Element) にすべての処理を移植する。次にプロファイリングを行い、多くの処理時間が掛かっている処理を探し出す。こうして見つかった時間の掛かる処理を複数のSPE (Synergistic Processor Element) に担わせる(オフロードする)ことで、全体の処理を高速化できる。また、SPEがオフロードされた処理を実行している間にPPEでは他の処理を実行できることもあり、単なるオフロードにとどまらず、PPEとSPEという異種プロセッサの並列動作も実現できる。

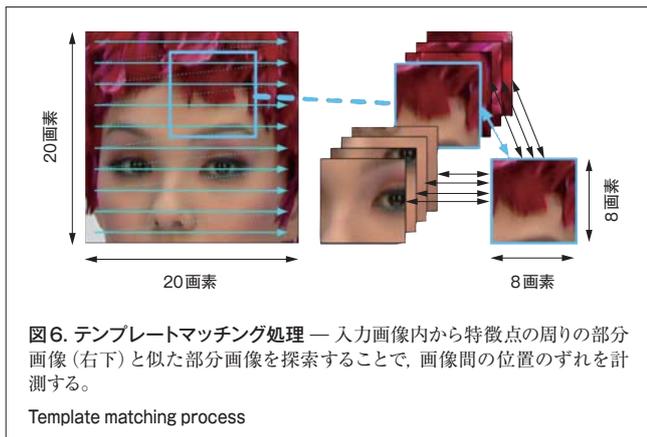
4.1 画像生成処理の並列化

画像生成には、ポリゴン レンダリングとテクスチャ マッピングという、一般的なCG描画処理を用いた。この処理では通常、ポリゴン(多角形)を画素ごとに分解するラスタライズ処理と、そのポリゴンにテクスチャ(表面画像)をはり付けるテクスチャ マッピング処理を同時に行う。しかし、これらの処理では、メモリ アクセス パターンに差があるため、当社は、ラスタライズ処理とテクスチャ マッピング処理を2段階に分けて実装し、それぞれを複数のSPEにオフロードする戦略を採った。このようにあらかじめ処理を分割することで、SPEへのオフロードと並列化を効率良く行うことができた。

4.2 画像間の差分計測処理の並列化

当社のアルゴリズムにおいて、もっとも重い処理がこの画像間の差分計測処理である。画像間の差分計測にはテンプレート マッチング処理を用いた。図6に示すように、合成画像上の特徴点の周りの部分画像に着目し、その部分画像にもっともよく似た入力画像上の部分画像を見つけるという処理を行う。当社の実装では部分画像は8画素×8画素とし、部分画像が似ていることの指標にはZSSD (Zero-mean Sum of Squared Difference) 関数を用いた。

SPEはSIMD (Single Instruction Multi Data) 演算器を持ち、16バイト単位のデータを一度に処理できる。また、一般的なSIMD演算器が得意とする浮動小数点演算だけでなく、char型(1バイト単位の文字型データ)やshort型(2バイト単



位の整数型データ) といった画像処理に向けたデータを一度に処理できるという特徴を持つため、テンプレート マッチング処理を通常の4倍から16倍という高い効率で実装することができた。また、多数の特徴点についてテンプレートマッチングを行う必要があるため、特徴点を複数のSPEに割り振ることで効率良い並列化を行った。

4.3 パラメータ最適化処理の高速化

パラメータ最適化処理についても、SPEへのオフロードを行った。パラメータ最適化処理は、主に行列の乗算とpseudo inverseという2種類の行列演算から成る。行列の乗算は、行列を分割し複数のSPEで同時に実行した。Pseudo inverseは、簡単のため一つのSPE上にオフロードした。

このシステムでは、行列演算処理の比率が低かったため急な最適化は行わなかった。しかし、元来、行列演算はCell/B.E.にもっとも向いた処理の一つであり、ライブラリが整備されれば容易にその効果を享受できるようになるだろう。

4.4 SPEの並列動作

以上のようなオフロードの結果、大部分の処理がSPEで並列に処理されるようになった。時系列に沿ったSPEの並列動作のようすを図7に示す。この結果、表1に示すように、フェ

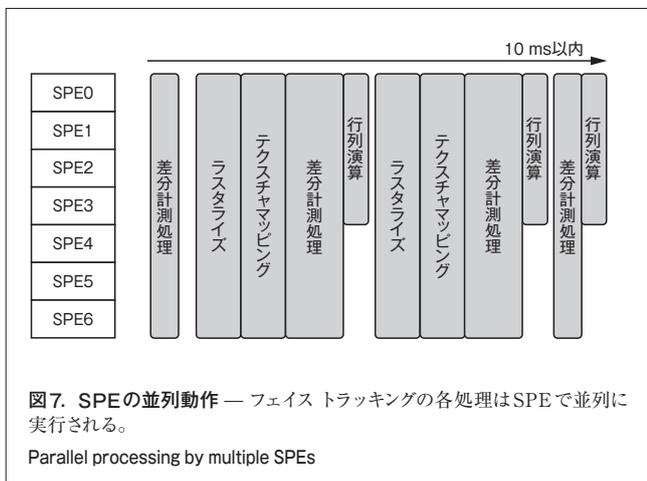


表1. Cell/B.E.向け最適化による処理時間の短縮

Results of reduction in processing time by offloading

項目	市販PCでの処理時間 (2.5 GHz)	CBEでの処理時間 (3.2 GHz)
画像合成処理	9.5 ms	5.5 ms
差分計測処理	21.0 ms	2.6 ms
行列演算	4.7 ms	1.3 ms

イス トラッキングをアプリケーションの一部として利用可能な、1フレーム当たり10 msの性能を実現した。しかし、SPEは常に動作しているわけではなく、休んでいる時間も多し。Cell/B.E.の性能を100%使い切るためには、よりむだのないSPEのスケジューリング手法や、他のアプリケーションやライブラリとのSPEのタイムシェアリングなどの手法を確立し、開発のコストを下げる努力を続ける必要がある。

5 あとがき

AbSは、従来リアルタイム化の難しい重い処理であったが、新しいアイデアを導入することで高速、かつ並列化の容易なアルゴリズムに拡張した。また、Cell/B.E.に向けて実装を最適化することで、1フレーム当たり10 msという速度目標を達成できた。

フェイス トラッキングを家電やPC上でリーズナブルなコストで実現できれば、新しいユーザーインターフェースとしてジェスチャ認識や目線認識などを用いた手法と同様に利用したり、それらと組み合わせて応用できる可能性がある。今後も、更に高速化や高精度化を進め、新しいユーザーインターフェースとしての可能性を開きたい。また、このフェイス トラッキング技術のCGやアニメーションといった映像制作現場への導入を予定している。人の感性に訴えかけるようなより高い精度を実現し、このような分野への貢献を進め、Cell/B.E.の活用分野を広げたい。

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. ACM, Osaka, 2003-10, ACM, New York USA, ACM Press, 2003, p.132-139.
- (2) J.A. Kahle, et al. Introduction to the cell multiprocessor. IBM Journal of Research and Development. 49, 4/5, 2005, p.589-604.



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.