

# Cell Broadband Engine™ を用いた ハンドジェスチャ ユーザーインターフェース

Hand Gesture User Interface Implemented on Cell Broadband Engine™

池 司      ビヨン シュテンガー      岸川 晋久      近藤 伸宏  
 ■ IKE Tsukasa      ■ Björn STENGER      ■ KISHIKAWA Nobuhisa      ■ KONDOH Nobuhiro

ハンドジェスチャ認識技術は、ビデオカメラで撮影したユーザーの画像を用いて、ユーザーの手の形状及び動きを認識する技術である。東芝は、ハンドジェスチャ認識技術を用いることにより、リモコンを使わずハンドジェスチャで機器を操作するシステムを試作した。環境変化に対して頑健なオブジェクト検出手法を用いて手を検出することにより、様々な環境でハンドジェスチャを認識できるようにした。また、Cell Broadband Engine™(注1)を用いて検出処理を並列に行うことで、ハンドジェスチャをリアルタイムで認識できるようにし、ユーザーインターフェースとして快適なレスポンスを実現した。

Toshiba has developed a hand gesture user interface (HGUI) system that enables users to control equipment without the use of manipulation devices. Hand gesture recognition is a technique for recognizing the user's hand postures and movements using images captured by a video camera. The system recognizes hand gestures under various environments with a robust detection method. We have also introduced parallel operations for feature computation and implemented them on the Cell Broadband Engine™. The system runs at sufficiently high speed to recognize the user's hand gestures in real time.

## 1 まえがき

近年、テレビやビデオレコーダ、エアコン、照明器具など、家庭内の様々な機器をリモコンで操作できるようになってきている。これにより、わざわざ移動することなく手元で機器を操作できるようになり便利になった反面、部屋にはそれぞれの機器に対応したリモコンが氾濫(はんらん)するようになり、かえってユーザーが操作にとまどったり、部屋が乱雑になったりする問題が生じている。こうした問題を解決するための手段の一つとして、ユーザーの手の形及び動き(ハンドジェスチャ)を機器に認識させるハンドジェスチャ認識技術を用いることにより、リモコンを用いずユーザーのハンドジェスチャによって機器を操作する試みが盛んに行われている。

東芝は、ビデオカメラによって撮影したユーザーの画像を用いてハンドジェスチャの認識を行い、認識結果を用いて機器を制御するハンドジェスチャ ユーザーインターフェース (Hand Gesture User Interface: HGUI) システムを試作した。

一般家庭での利用を想定し、照明条件やユーザーの背後に映っている物体が異なる環境においても、高い精度でハンドジェスチャを認識できる技術を開発し、試作システムに搭載している。また、高い演算処理能力を持つ Cell Broadband Engine™ (以下、Cell/B.E.と略記) 上でハンドジェスチャ認識処理を実行することでシステムのレスポンスを向上させ、快適に操作できるようにしている(図1)。

(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。



図1. HGUIシステム — ユーザーは、AV機器シミュレータの画面が表示されたモニタの前に立ってハンドジェスチャを行うことにより、映像の選択や再生などの操作を行うことができる。

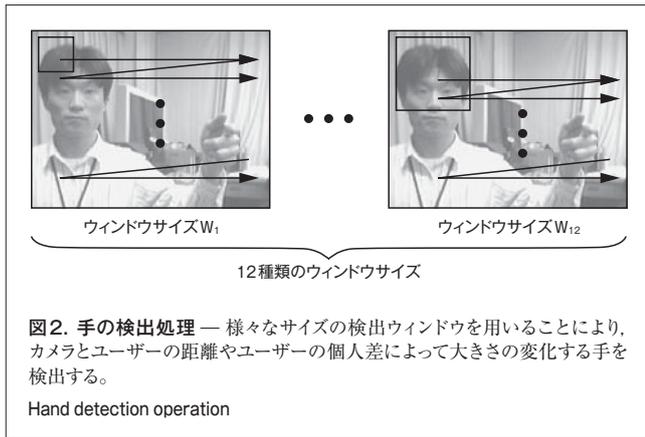
HGUI system

ここでは、ハンドジェスチャ認識の手法、Cell/B.E.を活用した認識処理の高速化、及び試作したHGUIシステムについて述べる。

## 2 ハンドジェスチャの認識

### 2.1 概要

ハンドジェスチャの認識は、ビデオカメラを用いてユーザーの画像を毎秒30枚程度の速度で撮影し、各画像に含まれる手を



検出して、その形と位置と大きさを求めることにより行われる。手の検出は、**図2**に示すようにカメラ画像に対して検出ウィンドウを走査させ、検出ウィンドウ内の画像に対してそれが手であるかどうかを識別することにより行う。ここで、画像上での手の大きさはユーザーの個人差やカメラからの距離によって変わるため、あらゆる大きさの検出ウィンドウを用意してそれぞれ画像を走査させることにより、大きさにかかわらず画像に含まれるすべての手を検出する。

次に、検出されたすべての手のなかから、検出対象となる手の形にもっとも似ていると識別処理によって評価されたものを選択し、その形と位置と大きさを認識結果として出力する。

## 2.2 手の識別処理

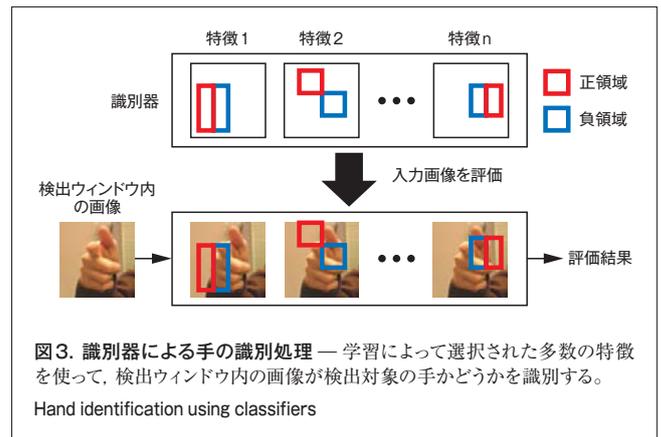
手の識別処理では、検出ウィンドウ内の画像について、その輝度値パターンを用いてそれが検出したい手であるかどうかを識別する。

検出ウィンドウ内の画像には、手とともにその背景にある物体も含まれるため、検出ウィンドウ内の輝度値パターンは同じ手の形であっても様々である。特に、一般家庭での利用を考えた場合、室内の明るさやユーザーの背後にある物体などの環境は部屋によって異なるため、ハンドジェスチャ認識技術では、こうした環境変化に対してロバスト（頑健）な識別手法が不可欠である。

そこで、試作したシステムでは、当社が考案した環境変化に対して頑健な顔識別手法<sup>(1)</sup>を手の識別処理に応用し、環境変化に強いハンドジェスチャの認識を可能にした。

この手法では、手形状ごとに対応する識別器を用いて検出ウィンドウ内の画像を評価することにより、手の識別処理を行う。ここで用いる識別器は、多くのサンプル画像を用いた学習によって生成したものを事前に用意する。

識別器は、**図3**に示すような多数の特徴によって構成されている。それぞれの特徴はHaar-like特徴と呼ばれ、これらは正と負の2種類の矩形（くけい）領域によって構成される。識別処理では、これら2種類の矩形領域内の輝度値平均を算出したうえで、これらの差が特徴ごとに定義されたしきい値を超



えているか（もしくは超えていないか）で類似度を算出する。これら類似度の総和が、識別しきい値を超えていれば手と識別する。

ここで、識別器を構成する特徴の位置、種類、及びしきい値は、サンプル画像として用意した数千～数万枚の正解画像及び不正解画像を用いた学習処理によって自動的に選択される。特徴の選択基準としては、正解画像と不正解画像を効率よく識別でき、かつ多くの正解画像に含まれる特徴（すなわち、環境変化の影響を受けにくい部位）を選択するアルゴリズムである。したがって、様々な環境で撮影したサンプル画像を用いて学習を行うことにより、例えば、手の輪郭部分や指と指の間の溝の部分など、環境変化の影響を受けにくい部位が自動的に選択されるため、環境変化に対して頑健な識別器が構成される。

更にこの手法では、二つ以上の特徴の相関も踏まえて手の識別処理を行う。例えば、特徴1と特徴2が同時に検出された場合に、類似度を高くするといった処理を行う。こうした相関情報の活用により、識別精度はいっそう向上する。

## 2.3 検出処理の高速化

ハンドジェスチャによって機器を快適に操作できるようにするには、ユーザーのハンドジェスチャを即座に認識し、機器が応答することが求められる。そこで、認識処理にかかる時間の大半を占める手の検出処理にCell/B.E.を活用することで、認識処理を高速化した。

Cell/B.E.は、3.2 GHzで動作し、汎用的な処理を行う1個のPower Processor Element (PPE)、及び高い算術演算性能を持つ7個のSynergistic Processor Element (SPE)によって構成されるマルチコアプロセッサである<sup>(2)</sup>。各SPEは、SIMD (Single Instruction Multiple Data)<sup>(注2)</sup>命令による4並列の32ビット算術演算機能を持つことから、7個のSPEを活用することにより28個の算術演算を同時に実行できる。そこで、手の検出処理を並列化することで、Cell/B.E.上でハンドジェス

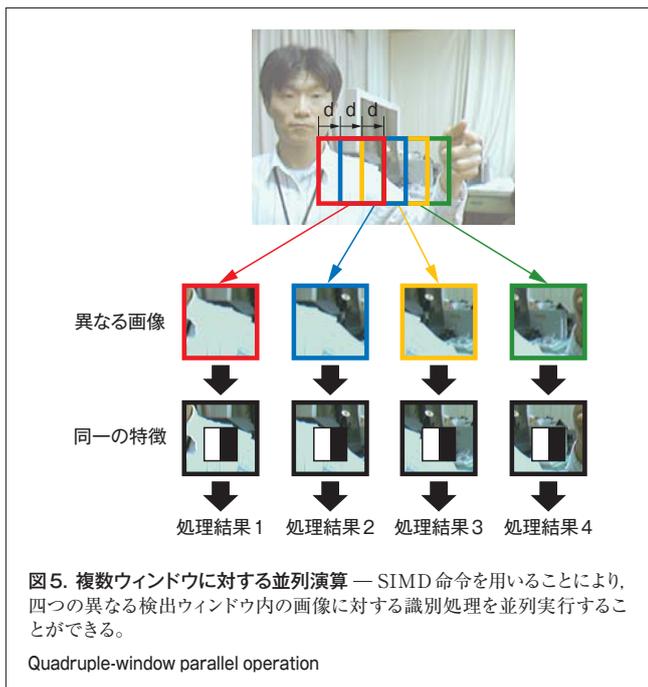
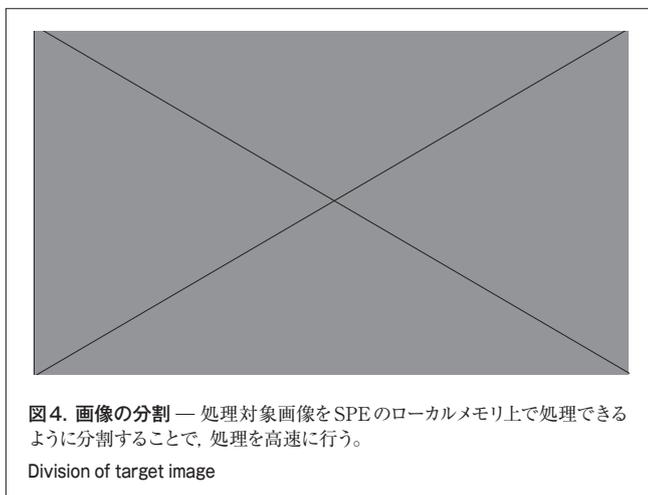
(注2) 演算装置設計手法の一つで、演算装置において、1回の命令で複数データに対する処理を同時に行うもの。

チャ認識処理を高速化した。

まず、手の検出処理(図2)において、異なるサイズの検出ウィンドウに対する処理は並列に実行できる。そこで、各ウィンドウサイズの処理を異なるSPEに割り当てることにより、処理を並列に実行した。

また、図4に示すように、処理対象画像を複数の部分画像に分割し、各SPEに搭載された256 Kバイトのローカルメモリ上で手の検出処理を高速化した。なお、部分画像の境界で手が検出されることを考慮し、部分画像間に一定のオーバーラップ領域を持たせている。

一方、部分画像内における手の検出処理では、各SPEでSIMD命令を用いることにより、複数の検出ウィンドウ内の画像に対する識別処理を並列に行い、検出処理を更に高速化した。すなわち、図5に示すように、それぞれ間隔 $d$ ずつ位置が



異なる四つの検出ウィンドウ内の画像に対して、同一の特徴を用いた識別処理をSIMD命令で並列処理することで、これら四つの検出ウィンドウ内の画像を同時に識別できる。

### 3 HGUI試作システム

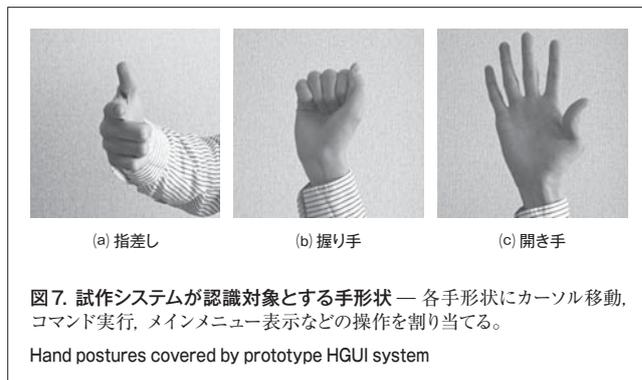
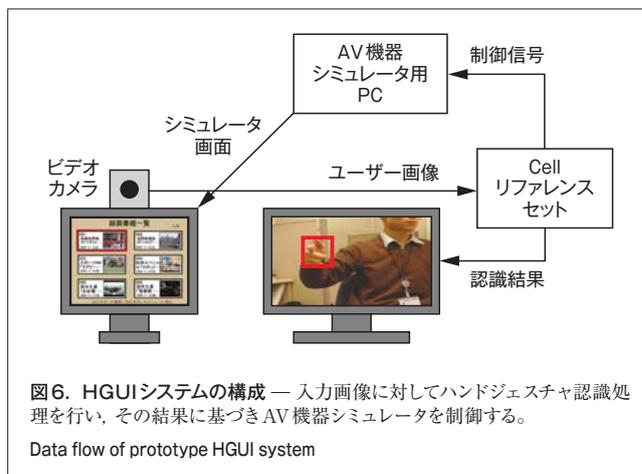
#### 3.1 システム構成

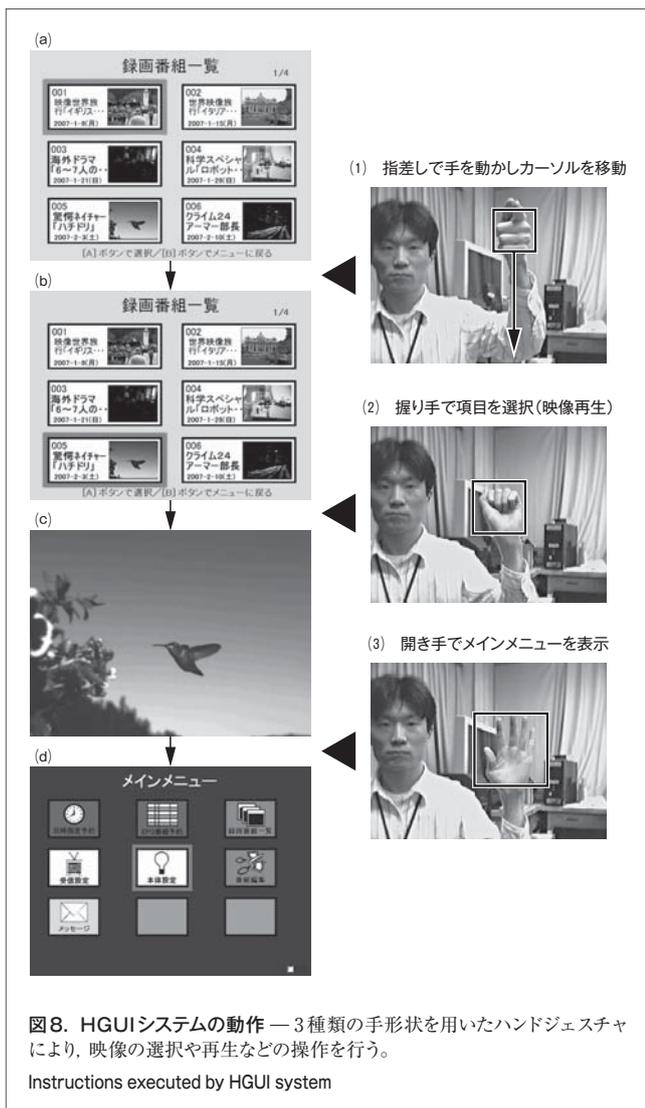
試作したHGUIシステム(図1)は、ユーザーのハンドジェスチャを認識し、認識結果に基づいてAV機器シミュレータを制御する。ユーザーはシミュレータ画面が表示されたモニタの前でハンドジェスチャを行うことにより、シミュレータの操作を行う。

試作システムの構成を図6に示す。シミュレータ画面の表示されたモニタの上部中央に、ユーザー撮影用のビデオカメラが設置されている。撮影されたユーザー画像は、Cell/B.E.を搭載したCellリファレンスセットに入力される。Cellリファレンスセットは、これらの画像を用いてハンドジェスチャ認識処理を行い、認識結果に基づいてAV機器シミュレータ用パソコン(PC)に制御信号を送信する。AV機器シミュレータ用PCは、受信した制御信号に基づきカーソルの移動や映像再生などの動作を実行する。

#### 3.2 HGUIシステムの動作

試作システムでは、図7に示す3種類の手形状(指差し、握





り手、開き手)を認識する。これらは、それぞれカーソル移動、コマンド実行、メインメニュー表示の各操作に割り当てられている。

図8(a)は、シミュレータが出力する録画番組一覧の画面である。左上の番組(001)の周囲を囲む太枠は、その項目が選択されていることを示すカーソルである。この状態で、例えば手を指差しの形(図8(1))にして移動させると、手の動きに応じてカーソルが移動する(図8(b)の番組(005))。カーソルを再生したい映像の項目へ移動させ、手を握り手の形(図8(2))にすると、その映像が再生される(図8(c))。また、手を開き手の形(図8(3))にすることにより、メインメニューが表示される(図8(d))。

### 3.3 処理速度

試作システムでは、図7に示した3種類の手形状について、それぞれ12種類のサイズの検出ウィンドウを用いて手の検出を行う。手の検出処理は手形状ごとに行うため、1枚の画像に対して36回検出ウィンドウを走査して手の検出を行う。こうし

た一連の検出処理をCellリファレンスセットで行わせた結果、処理時間は平均約34msとなった。ハンドジェスチャ認識処理をビデオカメラの撮影間隔と同程度の速度で完了することにより、ユーザーインターフェースとして快適な操作性を実現した。

## 4 あとがき

環境変化に対して頑健な手の識別手法を用いたハンドジェスチャ認識技術、及びCell/B.E.の高い演算能力を活用し、一般的な環境下で快適なレスポンスで動作するHGUIシステムを実現した。今後は、ユーザーの負担が少ないハンドジェスチャについて検討するとともに、引き続き認識技術の改良による認識精度のいっそうの向上に取り組んでいく。

## 文献

- (1) Mita, T., et al. "Joint Haar-like Features for Face Detection". Proc. International Conference on Computer Vision. Beijing, China, 2005-10, IEEE. IEEE, 2005, p.1619-1626.
- (2) 黒澤泰彦, ほか. 次世代プロセッサCell Broadband Engine. 東芝レビュー, 61, 6, 2006, p.9-15.



池 司 IKE Tsukasa, Ph.D.

研究開発センター マルチメディアラボラトリー, 博士(情報科学)。画像認識に関する研究開発に従事。IEEE, 電子情報通信学会会員。  
Multimedia Lab.



ピヨン シュテンガー Björn STENGER, Ph.D.

東芝欧州研究所 ケンブリッジ研究所 コンピュータビジョングループ, 博士(工学)。画像認識に関する研究開発に従事。IEEE, BMVA会員。  
Toshiba Research Europe Ltd., Cambridge Research Lab.



岸川 晋久 KISHIKAWA Nobuhisa

セミコンダクター社 システムLSI事業部 ブロードバンドシステムLSI開発センター。画像認識に関する研究開発に従事。情報処理学会会員。  
System LSI Div.



近藤 伸宏 KONDOH Nobuhiro

セミコンダクター社 システムLSI事業部 ブロードバンドシステムLSI応用技術部主務。画像処理アプリケーション開発業務に従事。  
System LSI Div.