

インタラクションシステムのための 高精度な3次元ハンドジェスチャ認識手法

玉 城 絵 美^{†1} 味 八 木 崇^{†2} 暦 本 純 一^{†2,†3}

カメラ動画像を用いたハンドジェスチャ認識は、インタラクションシステムの重要な技術の1つである。従来のハンドジェスチャ認識手法として、大量の手の特徴情報を持ったデータベースから類似データ検索し、認識する手法が多数提案されている。しかし、複雑な3次元ハンドジェスチャ認識には限界があった。特に、ヒトのハンドジェスチャの多く含まれる前腕回旋動作時に認識精度が低下するという問題点があった。そのため、従来手法は実用的なジェスチャ認識には適応できない。本研究では、手指の輪郭特徴情報に爪の位置情報を追加することにより、高精度なハンドジェスチャ認識システムを構築した。評価実験を行った結果、CPUがPentiumIV、2.8GHzのPCで、前腕回旋動作に対応しつつ各手指関節角度の誤差の標準偏差7.23度、処理速度100fpsが実現できた。

A Robust and Accurate 3D Hand Posture Estimation Method for Interactive Systems

EMI TAMAKI,^{†1} TAKASHI MIYAKI^{†2}
and JUN REKIMOTO^{†2,†3}

In this paper, a new 3D hand posture estimation system using a single camera is proposed. Existing hand gesture recognition systems estimate hand's 3D models based on image features such as contour or skin texture. However, it was difficult to estimate the wrist rotation because the contour and the texture data do not have enough information to distinguish hand's sides. To solve this problem, we propose a new 3D hand posture estimation system that uses data of nail positions. Nail positions are an important factor to recognize hand's sides. Using nail positions, it becomes possible to detect whether the camera is facing palm or dorsum. In addition, nail areas can be robustly extracted from a skin area by a simple image processing technique. Our proposed system uses a database consists of data-sets of the hand's contour, the nail positions, and finger joint angles. To estimate the hand posture, the system first extracts the hand's contour and the nail positions from the captured image, and searches for

a similar data-set from the database. The system then outputs the finger joint angles of the searched data-set. Our experimental results show high accuracy in the hand posture estimation with the wrist rotation, such as approximately 7.23 degree of standard deviation of error in the finger joint angles estimation.

1. はじめに

ハンドジェスチャ認識は、インタラクションシステムの基礎技術として重要な要素の1つである。ヒトが日常のコミュニケーションの1つとしてハンドジェスチャを使っていることから、ハンドジェスチャ認識がインタラクションシステムとして欠かせないことは明白である。図1にカメラを使ったハンドジェスチャ認識システムのコンセプトを示す。ユーザは、手に何も持たず、装着せずに情報入力を行うことができる。

ハンドジェスチャ認識には、大きく分けて3つの手法がある。1つ目は、直接関節角度を計測する手法である。データグローブ^{1),2)}は、ハンドジェスチャ認識にあたり、広く利用されている。ハンドジェスチャ動作の計測や実験を行う場合には、最適な手法である。しかしながら、データグローブを持ち歩き、装着する必要がある、常々使い続けるには不便である。

2つ目は、カメラとマーカを使ってハンドジェスチャ認識をする手法^{3)–5)}である。この手法では、まず手に複数のマーカを装着し、カメラでマーカ追跡し、モデルマッチングを行うことによって、ハンドジェスチャ認識を行う。すでに、高精度な認識手法と結果が出ているが、マーカ設置の点を考慮すると、日常生活で使用するには、まだ厳しいところがある。

3つ目は、カメラ1台のみを使ってハンドジェスチャ認識する手法である。この手法は、携帯電話やウェアラブル装置上の超小型カメラへの応用の点から、期待されている。そのため、多数の研究成果が発表されている。カメラ1台のみを使用する認識では、手の外郭線を抽出し、外郭線の特徴量化もしくはエッジ情報を用いて指先の位置認識を行い、ハンドジェスチャ認識する手法^{6)–9)}や、外郭線のデータベースもしくはハンドモデル照会することによって認識を行う手法^{10)–12)}などが多く提案されている。これらの研究成果により、カメラ1台のみで、高速で高精度にハンドジェスチャ認識を行うことができる。しかし、手の外

^{†1} 東京大学大学院学際情報学府

Graduate School of Interdisciplinary Information Studies, The University of Tokyo

^{†2} 東京大学大学院情報学環

Interfaculty Initiative in Information Studies, The University of Tokyo

^{†3} 株式会社ソニーコンピュータサイエンス研究所

Sony Computer Science Laboratories, Inc.

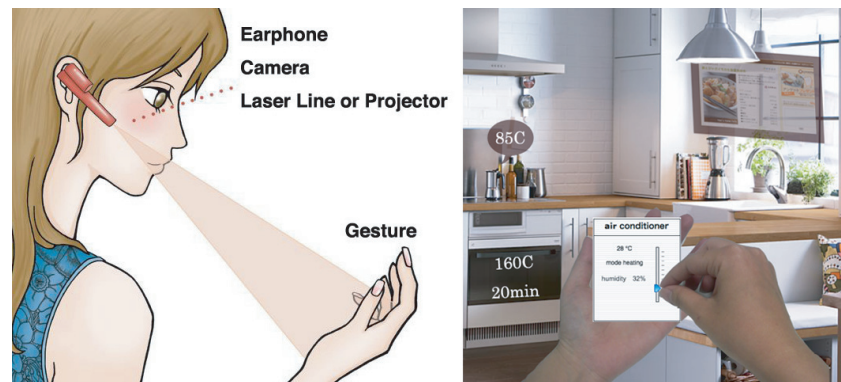


図 1 ハンドジェスチャ認識を用いたインタラクションシステムの例

Fig. 1 Examples of interaction devices using hand gesture recognitions.

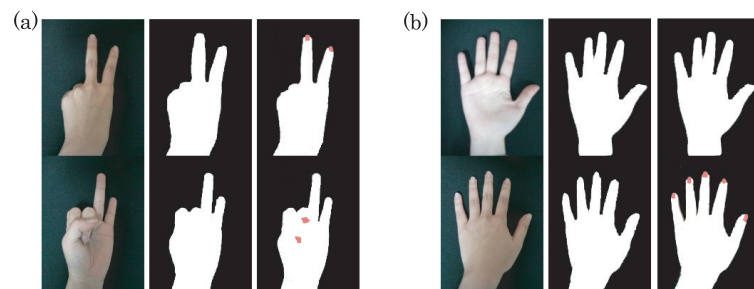


図 2 認識間違いが起きやすいハンドジェスチャの例．左側：カメラで撮影した画像．中央：外郭線．右側：外郭線と爪

Fig. 2 Examples of misunderstood hand posture. Left: Captured Images. Center: The contour. Right: The contour and the nails.

郭線のみでは、手の表裏情報を含まないため、手の回転すなわち前腕が回旋したときに認識間違いが発生する．図 2 は、実際に認識間違いが起きやすいハンドジェスチャの例である．

この問題に対し、Shimada ら¹²⁾ は、多量な手の外郭線データを用意し、データベースと各外郭線データ間の近傍マップを作成することによって、外郭線データのみで前腕回旋時のハンドジェスチャ認識する手法を確立した．しかしながら、同手法は、個人ごとの正確なデータデータベース作成の後に、正確に外郭線が抽出できていることが前提にある．そのた

め、現在の超小型カメラにそのまま転用することは難しい．

手の外郭だけでなく内側の輪郭線をも検出し、データベース照会を行う手法¹⁴⁾ も提案された．この手法で、図 2-(a) に示すような認識間違いを回避することができる．ただし、図 2-(b) に示すように、内側の輪郭線がない場合の認識間違い問題は残ったままである．

内側の輪郭線がない場合の認識間違い問題に対して、手の皺を利用した手法^{15),16)} が提案されている．この手法は、図 2-(a), (b) に示すような認識間違いを回避することができる．しかし、手の皺の検出自体が照明環境変化に左右されやすく、安定したハンドジェスチャ認識は期待できない．

同様に、手に映し出される影を利用した手法¹⁷⁾ も提案されているが、近年普及が著しい多灯分散照明下での認識の点を考慮すると、インタラクションシステムに用いることはできない．

インタラクションシステムのためのハンドジェスチャ認識の重要点は、取得したカメラ画像から安定して表裏情報を検出することである．本論文では、爪の位置情報を用いた新しいハンドジェスチャ認識手法を提案する．図 2-(a), (b) の 3 列目に示すように、外郭線に爪の位置情報を加えることで、手の表裏情報を検出することができる．また、簡単な画像処理¹⁸⁾ により爪領域を抽出することができる．

我々が提案する手法では、手の外郭線特徴量と爪の位置と手指の関節角度の 3 つの情報を持ったデータセットを、データベースとして多数用意する．ハンドジェスチャ認識の際には、カメラ画像から外郭線と爪の位置を検出し、この 2 つの情報をもとに、最も類似するデータセットをデータベース内から検索する．類似するデータセットが持つ手指の関節角度から 3 次元の手の形を再構成することによってハンドジェスチャを認識することができる．最後に提案システムの有効性を示すため、データグループを用いた評価実験を行う．

2. システム構成

2.1 概要

システム全体の処理はデータベース作成と実際のハンドジェスチャ認識に分かれる．図 3 にシステム構成の概要図を示す．データベースは手指の輪郭特徴量、爪位置、各関節角度の 3 つをデータセットにして作成される．そして、ハンドジェスチャ認識を行うときには、手指の輪郭特徴量と爪の位置情報を基にデータベース中で類似した手指形状を見つけ出し、それが持つ関節角度データを推定結果とする．

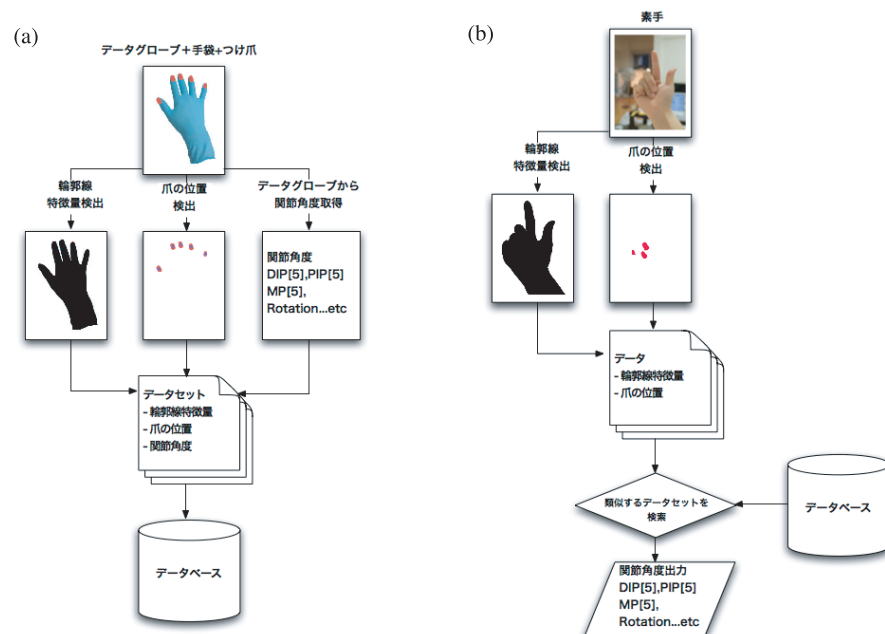


図 3 システム概要図。(a) データベース生成、(b) ハンドジェスチャ認識

Fig. 3 Our system outlines. (a) An construction of a database, (b) a hand gesture estimation.

2.2 データベース生成

2.2.1 処理の流れ

はじめに手指の画像を取得する。腕の位置を固定した状態で、カメラの前で自由に手指を動かして、データ取得を行う。手指画像の取得時は、照明環境を特定せず、カメラから 1 m 程度離れた位置の手指を撮影する。手指を撮影するカメラは、PointGray 社製のカラー高速カメラを使用した。解像度は 320×240 画素、画像取得速度は 30 fps で撮像した。手には関節角度取得のために Virtual Technologies 社製のデータグローブ CyberGlove を装着する。このとき、複雑なテクスチャを含んだデータグローブが手指全体を覆ってしまうため、白い手袋と赤い付け爪を装着した。また、付け爪は、装着後に大きさを切りそろえ、余計な装飾はとりはずしている。

次に、白と赤い領域を肌領域として輪郭線の抽出を行い、赤い領域を爪として爪の位置を

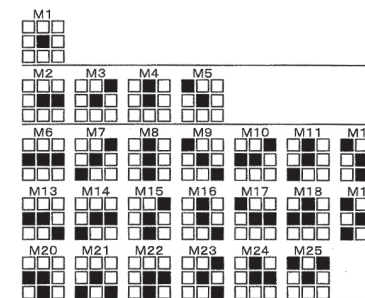


図 4 高次局所自己相関の局所パターン

Fig. 4 Higher-order local auto-correlational patterns.

検出する。そして、手指の輪郭線情報を数値化するために、特徴量化を行う。このとき、手指に装着したデータグローブより手指の関節角度を取得し、同時に前腕回旋角度も取得する。手指の輪郭線特徴量データ、爪の位置データ、関節角度データが揃うと、それを 1 つのデータセットとして保存する。最後に、データセットが十分な数に達すると、画像取得を終了し、データベースを構成する。

2.2.2 輪郭線の特徴量化

本研究では、Otsu ら¹⁸⁾ が提案している高次局所自己相関関数を、輪郭の特徴量の抽出に用いた。高次局所自己相関関数特徴は画像の認識や計測のために有用な特徴であり、次式で定義される高次相関関数について、参照点とその近傍についての計算をしたものである。

$$X^N(a_1, \dots, a_N) = \int f(r + a_0) \dots f(r + a_N) dr \quad (1)$$

ここで、 X^N は N 次元での r 点近傍の相関関数である。一般に自然画像を処理対象とする場合、対象点まわりの画素が重要なため、係数 N を 2 次までとした。また、平行移動による等価なものを除いて、図 4 で示す 25 の特徴量で表現できる。

ただし、No.1 から No.5 までは、No.6 以降の特徴量に比べスケールが小さいので、スケール差を合わせる必要がある。No.2 から No.5 までは参照点の画素をさらに掛け合わせ、No.1 は参照点の画素の 2 乗を掛けることで他の特徴量に一致させた。手指の画像は、全体の 320×240 画素からさらに手指部分だけ切り出し、輪郭抽出を行った後に 18×21 画素に縮小し、縦 6、横 7 ブロック分割を行い、そのすべてに特徴量の次元低減を行った。

2.2.3 爪の位置検出

本研究では、手指の領域内から爪の領域を抽出し、ラベリング処理を行った結果、得られた爪の領域の重心点を爪の位置とした。ただし、どの指の爪であるかは推定しない。なぜならば、手指の輪郭特徴量と爪の位置情報には、爪がどの指に対応しているかの情報が含まれているからである。今回は、爪の領域の抽出精度を確かなものとするため、データベース作成時に一般的な赤い付け爪を使用した。

2.2.4 データセット作成

2.2.2 項と 2.2.3 項の手指の輪郭特徴量、爪の位置と、データグローブより取得した各関節角度からデータセットを作成した。図 3-(a) にデータセットの作成概要図を示す。

爪の位置は、画像を 18×21 画素に縮小処理を行った後の x と y 座標で表し、爪がない場合は $(-1, -1)$ 座標とした。手指関節角度は前述のデータグローブを使用し、18 自由度の取得を行った。前腕回旋角度は、30 度ごとの 6 段階で計測を行った。

2.2.5 データベース構成

前腕回旋角度を基にして、3 層以上に階層化されたデータベースを作成した。第 1 階層では、前腕回旋角度を基にしている。第 2 階層以降では、データ収集時に、指の関節のうちで最も角度の分散が少ないものを基にする。データ量に応じて、この第 2 階層目と同じような階層を作成する。さらに、最下位であるデータセットは、指の関節角度順に並べられている。データベース作成前には、重複したデータセットが存在しないように、手指の輪郭特徴量と爪の位置が極端に近い値を持っているデータセットを削除した。

実験時のデータベース内のデータ数は合計 1 万個弱程度で、第 2 層の 1 つのクラスが持つデータベースは平均して 1,500 個であった。このときの第 2 層以下のデータベース階層は 5 階層である。なぜならば、前腕回旋が固定された状態でのハンドジェスチャ認識のためのデータベース階層は 5 階層が最も計算量が少ないと報告がなされているからである¹¹⁾。

2.3 ハンドジェスチャ認識

2.3.1 処理の流れ

本節では、ハンドジェスチャ認識の処理の流れを説明する。はじめに、手指の画像を取得する。腕の位置を固定した状態で、カメラの前で自由に手指を動かして、データ取得を行う。このときのカメラとカメラからの手指の距離、取得画像の解像度は、データベース作成時と同じである。ただし、手には何も装着していない。

次に、肌の領域抽出と爪の領域抽出するために、取得画像の表色系を個人別の表色系¹⁷⁾に変換する。その後、個人別表色系を基に肌と爪の領域を抽出し、輪郭線と爪の位置を検出

する。そして、手指の輪郭線情報を数値化するために、特徴量化を行う。

最後に、取得画像より検出した手指の輪郭線特徴量データと爪の位置データを用いて、データベース内の検索を行う。データベース中で、最も類似した手指形状のデータセットを見つけ出し、それが持つ関節角度データを推定結果とする。ただし、一連の処理はカメラ画像取得処理とは別スレッドで動作している。

2.3.2 肌と爪の領域抽出

手と爪の領域を抽出する際には、個人別表色系 YSaSb を用いた¹⁷⁾。個人別表色系は、照明変化に対応可能であると報告されており、テーブルや肌色モデル参照に比べ抽出処理速度が速い。本研究では、個人別表色系の肌色軸 S_a と任意の環境時の肌色の最大値と最小値を閾値として使用し、手の領域を抽出した。その後 S_a 軸に直行する S_b 軸を用いて、肌と同様に爪領域の抽出を行った。抽出の様子を図 5 に示す。図中の黄色が肌色、赤色が爪領域を示しており、爪領域を抽出可能であることが分かる。

個人別表色系 YSaSb の作成方法について説明する。YSaSb は、肌と爪の抽出処理を行う前に、個人ごとに事前に表色系を作成しておく必要がある。作成までの流れを以下に示す。

処理 1：取得画像の平滑化処理。

2：輝度と色差空間の分離。

3：肌色データ取得。

4：個人の肌色分布図を作成。

5：個人の表色系を作成。

処理 1 では、カメラからの取得画面上の細かなノイズを除去するために、平滑化を行った。

処理 2 では、肌色分布図を取得する前に、照明の強さの影響を省くため YIQ 表色系を利

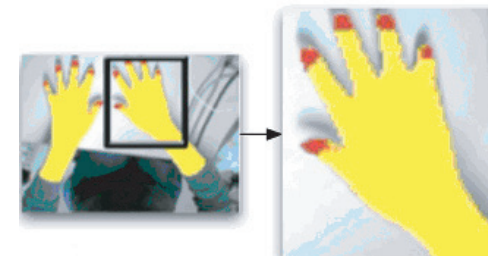


図 5 爪領域の抽出

Fig. 5 An example of nail area extractions.

用して, RGB 入力のカメラ画像を輝度 (Luminance) と色差や色度 (Chrominance) に分離した. 次式に RGB から YIQ 表色系への変換式を示す.

$$\begin{aligned} R' &= \begin{cases} 4.5 \times R & R < 0.018 \\ 1.099 \times (R^{0.45}) - 0.099 & R \geq 0.018 \end{cases} \\ G' &= \begin{cases} 4.5 \times G & G < 0.018 \\ 1.099 \times (G^{0.45}) - 0.099 & G \geq 0.018 \end{cases} \\ B' &= \begin{cases} 4.5 \times B & B < 0.018 \\ 1.099 \times (B^{0.45}) - 0.099 & B \geq 0.018 \end{cases} \end{aligned} \quad (2)$$

$$\begin{aligned} Y &= 0.299 \times R' + 0.587 \times G' + 0.114 \times B' \\ I &= 0.596 \times R' - 0.274 \times G' - 0.322 \times B' \\ Q &= 0.211 \times R' - 0.522 \times G' + 0.311 \times B' \end{aligned} \quad (3)$$

ただし, 式中の R と G と B は RGB 表色系のそれぞれの値であり, Y と I と Q は YIQ 表色系のそれぞれの値である. YIQ 表色系は北アメリカのテレビ放送に使用され, 肌色表現に優れている表色系である. 肌色が分布すると思われる領域を, 式 (2) の γ 補正により伸張して粒度が高く表現できるため分離に使用した. しかし, 厳密に言えば, 後に個人の表色系を作成するので, 輝度と色差空間の分離が行えるのならば, YUV (YCrCb) 表色系による色差空間での表現でもよい.

処理 3 では, 100 フレーム分の肌色データを取得する際には, 取得画像の中心部に掌より小さい枠を作り, 1 秒弱ほど枠内に手領域をあててデータ取得を行った.

処理 4 では, 個人の表色系の肌色軸の傾きを肌色モデルの重心と原点を結ぶ線から求めた. なお, 重心に対する重みは色の出現回数 m とした. 重心点 $\alpha(Q', I')$ を求める式を以下に示す.

$$\begin{aligned} Q' &= \frac{\sum m q_n q_n}{M} \\ I' &= \frac{\sum m i_n i_n}{M} \end{aligned} \quad (4)$$

ただし, $m q_n$ は, Q 軸上の位置 n の色の出現回数, $m i_n$ は I 軸上の位置 n の色の出現回数, q_n は Q 軸上の位置, i_n は I 軸上の位置, M は取得画像の画素数を示している. 式 (4) より求めた重心点 α を単位ベクトル化し, $\overrightarrow{\alpha(Q', I')}$ とするとき, 式 (3) の I を I' 傾けることにより, 肌色軸 S_a の軸成分 S'_a を求める. 次式に S'_a を示す. 同様に肌色直交軸 S_b の軸

成分 S'_b も単位ベクトル, $\overrightarrow{\alpha(Q', I')}$ に直交する単位ベクトルから計算することができる.

$$\begin{aligned} S'_a &= (0.596 \times I' + 0.211 \times Q') + (-0.2741 \times I' - 0.522 \times Q') \\ &\quad + (-0.322 \times I' + 0.311 \times Q') \end{aligned} \quad (5)$$

本論文では, 個人別表色系 YSaSb を用いたときの爪領域の抽出の有用性を検証するために, 精度実験を行った. 肌領域の抽出についての精度実験は参考論文 17) に示されているが, 爪領域の抽出に関する詳細な評価実験結果はない. 爪領域の抽出に関して, YSaSb の有用性を確認するため, 様々な環境下で抽出精度を実験する必要がある. 背景が黒, 青, 緑, 赤と白の 5 種類と, 照明が蛍光灯, 青, 赤, 室内太陽光と室外太陽光の 5 種類, 合計 10 種類の環境下での爪領域の抽出精度実験を行った. 各環境下で, 5 個の爪を含んだ 640×480 pixels の画像を 3 枚, 合計 150 個の爪が含まれる画像 30 枚用意した. 640×480 pixels の解像度で, カメラから 1 m 手をはなした状態でハンドジェスチャ認識を行った場合, 爪 1 つにつき平均して 1,008 pixels の領域がある. 200 pixels 程度の領域が抽出できれば, ノイズ除去処理で削除される可能性は低い. このことから, 1 つの爪領域に対し 20% 以上の領域抽出ができれば, 爪の検出ができる. 爪の領域の真値として, 筆者が実際に目視でマーキングした画像を用いた. YSaSb 表色系を用いて爪領域の抽出を行った画像と, マーキングした画像とを比較し, 爪領域抽出の再現率を測定した. そのうち, 爪の検出に必要な領域を抽出できた爪の割合を, 表 1 と表 2 に示す.

結果より, 青の照明環境下では, 爪として十分に検出できる確率が 60% となり, 位置検出に影響することが分かった. 肌色軸 S_b が, 青緑色に沿った軸になりやすい傾向のため, 特に青の照明に弱いと考えられる. しかしながら, 9 種類の環境下では, 爪の検出が可能であることを確認した.

表 1 個人別表色系 YSaSb を用いた背景別の爪検出率 (%)

Table 1 Results of the nail area extractions using YSaSb in 5 backgrounds (%).

Black	Blue	Green	Red	White
100.00	100.00	100.00	93.33	100.00

表 2 個人別表色系 YSaSb を用いた照明別の爪検出率 (%)

Table 2 Results of the nail area extractions using YSaSb in 5 lighting enviroment (%).

Fluorescent light	Blue	Red	Sunlight (Inside)	Sunlight (Outside)
100.00	60.00	100.00	93.33	100.00

2.3.3 データ探索

計算量短縮のため、現在のデータセットからみて近傍のデータのみを検索範囲として、手指の輪郭特徴量と爪の位置から、最も類似度の高いデータセットを検索する。

次に、類似度を算出する際には、単純なユークリッド距離計算をする。類似度 E_r を次式に示す。

$$E_r = \sum_{i=1}^P (x_i - x'_i)^2 + \sum_{k=1}^N (y_k - y'_k)^2 \quad (6)$$

ただし、 P は高次局所パターン数、 x_i はデータベース内のデータセットの手指輪郭の i 番目の特徴量、 x'_i は取得画像の手指輪郭の位置 i 番目の特徴量、 N は爪の数、 y_k はデータベース内のデータセットの k 番目の爪の位置、 y'_k は取得画像の k 番目の爪の位置を示している。 E_r を最小にするデータセットを検索し、そのデータセットが持つ手指の関節角度を出力することによって、3 次元ハンドジェスチャ認識を行う。結果の例として図 6 に、形状推定時の取得画像と、推定結果を用いて動作させた CG モデルを示す。これらの写真より、本システムが高精度で推定を行っていることが感覚的に分かる。

3. 評価

3.1 実験方法

推定された関節角度を定量的に評価するため、手にデータグローブ装着し実験を行った。このときの動作システムは、2.3 節の形状推定とまったく同じシステムである。提案システムより出力される関節角度の推定値と、データグローブより出力される測定値を真値として評価実験を行った。ただし、複雑なテクスチャを含んだデータグローブが手指全体を覆ってしまうため、白い手袋と赤い付け爪を装着した。手袋と付け爪の色については、人工色のため、どの色においても単色であることに変わらず、手袋と付け爪色によって抽出精度が変化することはないので、任意の色とした。

被験者は先述の高速カメラの前方 1 m 程度の位置に手をかざし、腕の位置を固定した状態で前腕と手指を自由に動かした。CPU Pentium IV、クロック周波数 2.8 GHz の PC で動画像を 30 fps で撮像し、撮像とは別スレッドで手指の形状推定を行った。また、データベース内のデータ数は 8,934 個であり、そのときの手指の形状推定の処理速度の平均は 100 fps であった。

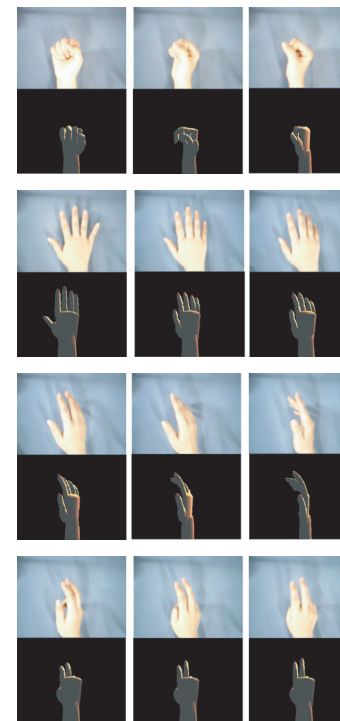


図 6 形状推定の結果。背景青：撮影された画像，背景黒：推定結果
Fig. 6 Results of 3D hand posture estimations. Blue background: Captured images, Black background: Results.

3.2 爪の位置情報あり

前腕回旋動作時の手指関節角度の推定精度実験を行った。初めに、中指 PIP 関節角度推定の実験結果を図 7 に示す。ただし、図 7-(a) の縦軸は関節角度、横軸はフレーム数を示しており、図中の太い実線はデータグローブによる手指関節角度の測定値、細い実線は本システムによる手指関節角度の推定値、破線はそのときの前腕回旋角度の推定値を示している。図 7-(b) の縦軸は実験時の中指 PIP 関節角度の推定誤差、横軸はフレーム数を示している。実験時には、推定値は 10 フレーム前からの移動平均を求めることにより、平滑化を行っている。前腕回旋動作は 30 度ごとの 6 段階で計測されており、右手の掌がカメラに向いてい

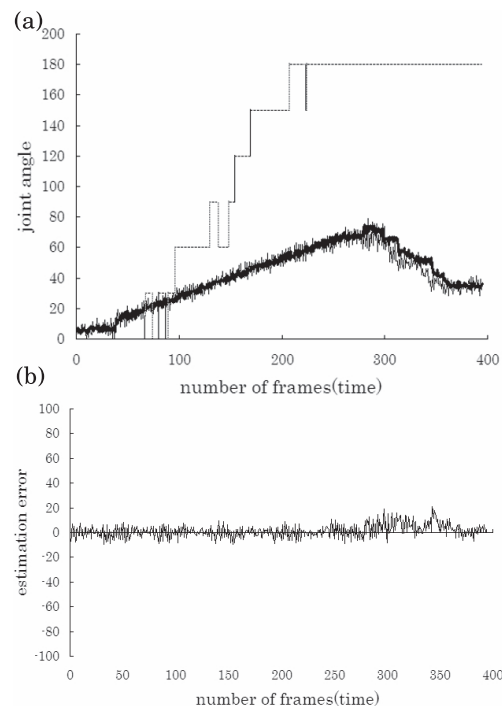


図 7 前腕回旋時の中指 PIP 関節角度 . (a) 推定精度 , (b) 推定誤差

Fig. 7 Middle finger joint angle with the wrist rotation. (a) An estimation result, (b) an estimation result error.

るときの前腕回旋角度を 0 度, 手の甲がカメラに向いているときの前腕回旋角度を 180 度とした . 図 7-(a) より, 前腕回旋時でも中指 PIP 関節角度が十分に推定できていることが確認できる . しかし, 図 7-(b) より, 前腕回旋角度が 0 度で固定のときに比べ, 動作時は推定誤差の標準偏差が 8.40 度増加し, 精度が低下する . また, 前腕回旋角度が 180 度で中指 PIP 関節の角度の値が高いとき, つまり, 爪が見えないときは, 実際よりも値が低く推定され, 推定誤差が大きくなっている . しかしながら, 前腕回旋時の中指 PIP 関節角度推定の誤差平均は 8.22 度, 標準偏差 6.09 度で, ハンドジェスチャ認識として十分に良い精度が得られた .

3.3 爪の位置情報なし

本システムと爪の位置情報を持たないシステムとの比較のために, 輪郭線特徴量と手指関節角度のみをデータセットとした形状推定システムを構築し, 前腕回旋時の推定精度実験を行った . データベース構築手法は, 本システムと同じである . 形状推定時の類似度算出には次式を用いた .

$$E_r = \sum_{i=1}^{25} (x_i - x'_i)^2 \quad (7)$$

ただし, E_r は類似度, x_i はデータベース内のデータセットの手指輪郭の i 番目の特徴量, x'_i は取得画像の手指輪郭の位置 i 番目の特徴量, 25 は図 4 で示す局所パターン数を示している . E_r を最小にするデータセットを検索し, そのデータセットが持つ手指の関節角度を出力することによって, 3 次元ハンドジェスチャ認識を行う .

カメラの前で自由に手を動かしたときの人差し指 PIP 関節角度推定の実験結果を図 8 に示す .

ただし, 図 8-(a) の縦軸は関節角度, 横軸はフレーム数を示しており, 図中の太い実線はデータグループによる手指関節角度の測定値, 細い実線は本システムによる手指関節角度の推定値, 破線はそのときの前腕回旋角度の推定値を示している . 図 7-(b) の縦軸は実験時の中指 PIP 関節角度の推定誤差, 横軸はフレーム数を示している . 実験時には, 推定値は 10 フレーム前からの移動平均を求めることにより, 平滑化を行っている . 前腕回旋動作は 30 度ごとの 6 段階で計測されており, 右手の掌がカメラに向いているときの前腕回旋角度を 0 度, 手の甲がカメラに向いているときの前腕回旋角度を 180 度とした . このときの推定の誤差の標準偏差 22.0 であった .

爪の位置情報を含まないシステムは, ジェスチャ認識に必要な推定精度に達していない . 前腕回旋角度が 80 度から 180 度の間は, 0 度から 80 度のときの推定に比べ, 推定誤差平均が 39.7 度増加していた . つまり, 前腕回旋角度が 80 度以降のとき, 推定精度が落ちる . 3.2 節の爪の位置情報を含んだシステムは, 前腕回旋角度 80 度以降も高い推定精度を保っており, 前腕回旋動作時のハンドジェスチャ認識には, 爪の位置情報が有用であると確認できる .

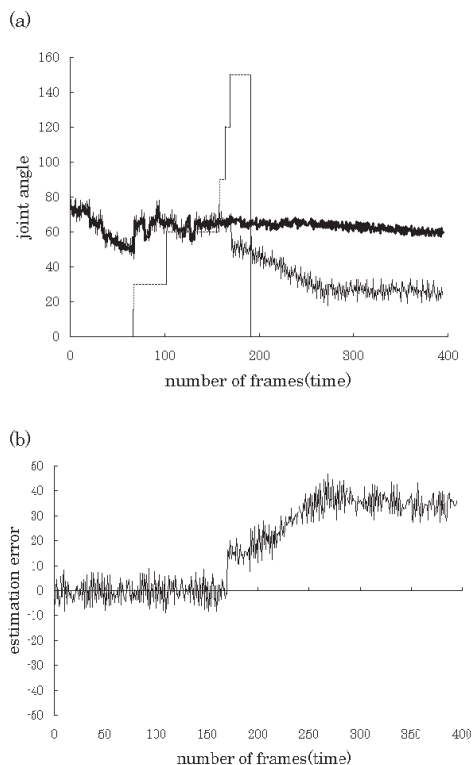


図 8 爪の位置情報を持たないシステムでの人差し PIP 関節角度の推定精度. (a) 推定精度, (b) 推定誤差
Fig. 8 Thumb joint angle with the wrist rotation. (a) An estimation result, (b) an estimation result error.

4. 議 論

4.1 データベースの自動生成

2.2 節でのデータベース生成手法も有用であるが、データベースを作成するためにカメラの前で人が手を動かす手間がかかってしまう。この問題を解決するために、本節ではデータベースを自動生成する手法を提案する。

まず、ユーザ手の写真を複数枚撮影し、手の形に合わせたハンド CG モデルを作成する

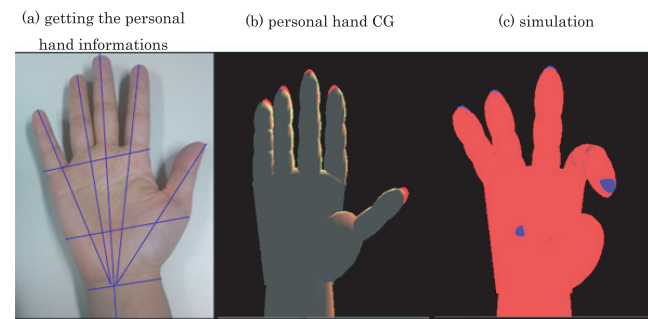


図 9 手の CG モデル作成の例
Fig. 9 An example of making hand CG model.

(図 9). 次に、そのハンド CG モデルを動作させている様子から、各ハンドジェスチャの輪郭線と爪の位置と関節角度情報を取得し、データセットを多数作成する (図 10). その多数のデータセットから 2.2 節と同様にデータベース生成を行う。

4.2 素手での評価実験

4.1 節で作成した各関節 15 度刻みのデータベースを用いて、実際に素手でのハンドジェスチャ認識ができるかどうか評価実験を行った。しかし、データグローブを装着しない状態で、素手の関節角度の真値を検出することが困難であるため、以下の評価実験手法を用いた。

前腕回旋と指関節角度に関し自由に動作する素手に対し、2 視点から別々にハンドジェスチャ認識を行う。この 2 つの視点からの認識結果が一致する場合、ハンドジェスチャ認識可能であるとした。今回は、被験者の手を中心として 107 度あけ、2 視点のカメラを設置した。カメラは MacBook に内蔵されている iSight2 台を用いた。iSight には、環境光の変化に合わせて輝度を自動調整する機能が標準で設定されているが、これを切った状態で実験を行った。取得した画像の解像度は 640×480 pixels である。人差し指 MP 関節 (中手指節間関節, Metacarpophalangeal joint) の実験結果を図 11 に示す。図 11-(a) より、2 視点からのハンドジェスチャ認識が同様に追従していることが確認できる。また、爪の位置情報を用いない図 11-(b) の実験では、18 フレーム目から 2 視点からの認識結果が異なる。フレーム数 0 から 17 までの間、把持のジェスチャ (グー) とすべての指を広げたジェスチャ (パー) の状態を繰り返していたが、18 から人差し指と他の指とを連動させないジェスチャ (チョキ) を行ったところ、片方のハンドジェスチャ認識システムが手を表裏逆に認識してしまった結果である。素手でのハンドジェスチャ認識結果で、再度爪の位置情報の有効性を

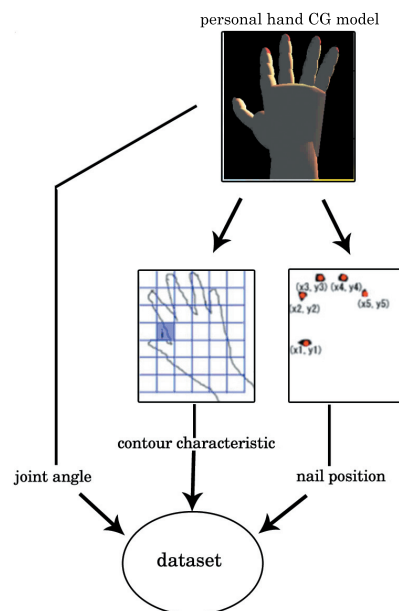


図 10 手の CG モデルを用いたデータセット作成の概要
Fig. 10 The construction of data-set by using hand CG model.

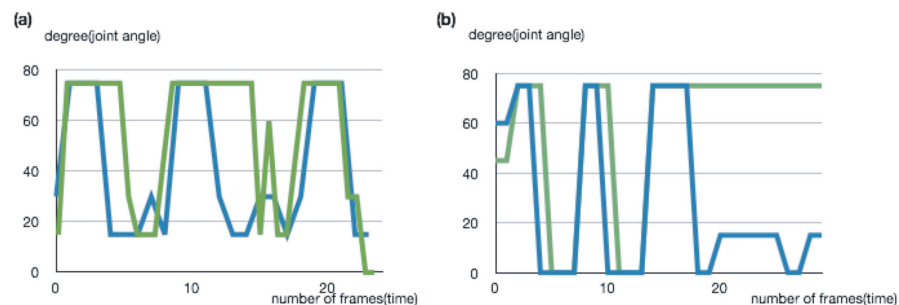


図 11 2 視点からのハンドジェスチャ認識結果：人差し指 MP 関節角度。(a) 爪の位置情報あり，(b) 爪の位置情報なし
Fig. 11 Results of the hand gesture estimation system from the 2 viewpoints: the thumb MP joint angles, (a) Using the data of contour and nail positions, (b) using the data of only contour.

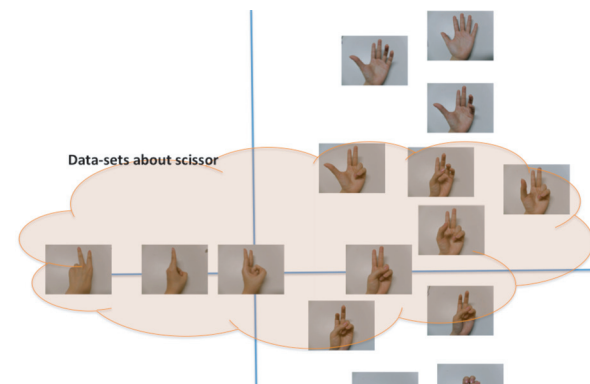


図 12 データベース上でのハンドサイン定義の例
Fig. 12 Definitions of hand signs on database.

確認できた。

4.3 ハンドサインの認識

本論文で提案したジェスチャ認識システムから出力された手指の関節角度からハンドサインとして情報入力を行う手法について説明する。たとえば、チョキのサインを定義する場合は、ユーザが定義したチョキの手指の関節角度と任意の許容角度前後を設定し、さらにすべての前腕回旋角度を許容することによって関節角度からハンドジェスチャサインを定義することができる(図 12)。

4.4 応 用

本節ではインタラクシオンシステムへの応用例を示す。図 13 は、Tamaki らによって提案されたインタラクシオン装置 Brainy Hand¹⁹⁾である。同装置上の超小型カメラを用いてハンドジェスチャを認識し、情報を入力している。また小型のプロジェクタで掌に映像を投影することができる。同装置の超小型カメラ MO-S1588-8F は、 656×492 -pixels の動画像を取得することができるが、照明光の強さにより、精密な外郭線抽出が難しい。しかしながら、外郭線と爪の位置情報を用いた本論文のハンドジェスチャ認識手法を用いることによって、安定したハンドジェスチャ認識が可能となっている。また、単純な情報入力装置としてではなく、ロボットの操作としても応用することもできる¹⁶⁾。



図 13 インタクション機器への応用例

Fig. 13 A practical application on interaction device.

5. ま と め

ハンドジェスチャ認識は、インタクションシステムを実現するための重要な基礎技術である。しかし、従来手法の前腕回旋動作を含んだハンドジェスチャ認識には、手の表裏の誤認識により、推定精度が著しく低下するという問題があった。この問題を解決するためには、手指の表裏情報が必要と考えられた。本研究では、手指の輪郭の特徴量に爪の位置情報を追加することにより、指先の位置と表裏の 2 つの情報も併用するハンドジェスチャ認識システムを提案した。

手指の輪郭と爪の位置を検出するために、個人別表色系 YSaSb¹⁷⁾ を用いた。10 種類の背景と照明環境下での、爪の位置検出の有効性について検証した。結果、青の照明環境時には、爪の位置検出が難しく、ハンドジェスチャ認識システムの精度が低下する。しかしながら、9 種類の環境下で、爪の位置検出が可能であることが確認できた。

実際にハンドジェスチャ認識システムを構築し、前腕回旋動作を含んだハンドジェスチャの映像から手指の関節角度を推定する実験を行った。結果、データ数 8,934 個のとき、CPU が PentiumIV, 2.8 GHz の通常仕様の PC で、前腕回旋動作に対応しつつ手指関節角度の誤差の標準偏差 7.23 度、処理速度 100 fps が実現できた。また、爪の位置情報を抜いたシステムとの比較を行い、前腕回旋動作時の爪の位置情報の有効性を示した。

加えて、手の CG モデルからデータベースを自動生成する手法、ハンドジェスチャサイ

ンとして認識する手法やインタクション装置の応用例を示した。

参 考 文 献

- 1) Data Glove Ultra Series, 5DT Data Glove 5 Ultra/5DT Data Glove 14 Ultra.
- 2) The CyberGlove, Immersion Corporation, 18-joint angles sensor.
- 3) Amaya, K., Calvert, T. and MacKenzie, C.L: Posture Estimation of a Human Hand from a Limited Number of Markers, IEICE technical report, Pattern recognition and understanding, Vol.12, No.95, pp.63–69 (1995).
- 4) Wayne, P. and Bruce, T.: Using ARToolKit for 3D Hand Position Tracking in Mobile Outdoor Environments Mobile Outdoor Environments (2002). <http://www.tinmith.net/papers/piekarski-art-2002-poster.pdf>
- 5) Takegawa, Y., Terada, T. and Nishio, S.: Design and Implementation of a Real-Time Fingering Detection System for Pianists, *WISS2005*, pp.95–96 (2005).
- 6) 蔵田武志, 興梠正克, 加藤丈和, 大隈隆史, 遠藤 健: ウェアラブルビジュアルインターフェースのための機能分散型ハンドトラッキング手法とその応用, 日本 VR 学会論文誌, Vol.10, No.4, pp.555–564 (2005).
- 7) 佐々木博史, 黒田知宏, 眞鍋佳嗣, 千原國宏: 『てのひらめにゅう』: ウェアラブルコンピュータ用入力インタフェース, 日本バーチャルリアリティ学会論文誌, Vol.7, No.3, pp.393–401 (2002).
- 8) Kurata, T., Kato, T., Kouroggi, M., Keechul, J. and Endo, K.: A functionally-distributed hand tracking method for wearable visual interfaces and its applications, *Proc. MVA*, pp.84–89 (2002).
- 9) Sasaki, H., Kuroda, T., Manabe, Y. and Chihara, K.: HIT-Ware: A menu system superimposing on a human hand for wearable computers, *Proc. ICAT'99*, pp.146–153 (1999).
- 10) Lee, T. and Hollered, T.: Hybrid Feature Tracking and User Interaction for Markerless Augmented Reality, *Proc. IEEE VR 2008 (10th Int'l. Conference on Virtual Reality)*, pp.1–8 (2008).
- 11) Lee, T. and Hollered, T.: Initializing Markerless Tracking Using a Simple Hand Gesture, *ISMAR 2007, 6th IEEE and ACM International Symposium*, pp.259–260 (2007).
- 12) Shimada, N., Kimura, K. and Shirai, Y.: Real-time 3D Hand Posture Estimation based on 2 D Appearance Retrieval Using Monocular Camera, *Proc. Int. WS. on RATFG-RTS (satellite WS of ICCV2001)*, pp.23–30 (2001).
- 13) Shimada, N., Shirai, Y., Kuno, Y. and Miura, J.: 3D Pose Estimation and Model Refinement of An Articulated Object from A Monocular Image Sequence, *Conf. on Face and Gesture Recognition*, pp.268–273 (1998).
- 14) Rafael Diaz, M., Travieso, C.M., Alonso, J.B. and Ferrer, M.A.: Biometric sys-

- tem based in the feature of hand palm, *Security Technology, 38th Annual 2004 International Carnahan Conference*, pp.136–139 (2004).
- 15) Lee, H.-J. and Chung, J.-H.: Hand Gesture Recognition using Orientation Histogram, *IEEE TENCON 99, Proc. IEEE Region 10 Conference*, No.2, pp.1355–1358 (1999).
 - 16) de La Gorce, M., Paragios, N. and Fleet, D.J.: Model-Based Hand tracking with texture, Shading and Self-occlusions, *CVPR08*, pp.1–8 (2008).
 - 17) Tamaki, E. and Hoshino, K.: Personalized color system for robust extraction of skin-color, *The Virtual Reality Society of Japan*, Vol.12, No.4, pp.471–478 (2007).
 - 18) Otsu, N. and Kurita, T.: A new scheme for practical, flexible and intelligent vision systems, *Proc. IAPR Workshop on Computer Vision*, pp.431–435 (1988).
 - 19) Tamaki, E., Miyaki, T. and Rekimoto, J.: Brainy hand: An ear-worn hand gesture interaction device, *CHI2009*, Work in Progress (2009).
 - 20) Hoshino, K., Tomida, M. and Tamaki, E.: Copycat Arm: An aping arm mimicking human motions without delay, *SIGGRAPH New Tech. Demos* (2008).

(平成 21 年 4 月 20 日受付)

(平成 21 年 11 月 6 日採録)



玉城 絵美 (正会員)

2006 年琉球大学工学部情報工学科卒業。2008 年筑波大学大学院システム情報工学研究科修士課程修了。現在、東京大学大学院学際情報学府博士課程在学中。主に、ハンドジェスチャに関するインタラクシオンシステムの研究に従事している。ACM、日本バーチャルリアリティ学会各会員。



味八木 崇 (正会員)

2008 年東京大学大学院新領域創成科学研究科博士後期課程修了。同年東京大学情報学環特任助教。実世界センシング、画像処理に関する研究に従事。ACM、IEEE、電子情報通信学会各会員。博士 (科学)。



暦本 純一 (正会員)

1986 年東京工業大学大学院情報科学科修士課程修了。1994 年より株式会社ソニーコンピュータサイエンス研究所に勤務。2007 年より東京大学大学院情報学環教授。理学博士。ヒューマンコンピュータインタラクシオン、特に実世界指向インタフェース、拡張現実感等に興味を持つ。1990 年情報処理学会 30 周年記念論文賞、1998 年 MMCA マルチメディアグランプリ技術賞、1999 年情報処理学会山下記念研究賞、2003 年日本文化デザイン賞、2005 年 iF Communication Design Award、2007 年 ACM SIGCHI Academy、2008 年日経 BP 技術賞等を受賞。