

距離画像から得られる形状空間に基づく 手の形状識別と姿勢推定

菊池 優介^{1,a)} 日野 英逸^{2,b)} 福井 和広^{2,c)}

概要: 本稿では 1 枚の距離画像から得られる形状空間に基づいて、手形状とその姿勢を推定する方法を提案する。提案法ではまず比較する 2 つの手形状の距離画像を撮影し、各距離画像から形状空間を生成する。次に得られた 2 つの形状空間の距離 (類似度) を両者のなす正準角で計り、与えられた手形状同士の 3 次元形状類似度とする。形状空間は視点と対象物の位置関係に依存せずに生成されるので、提案法はカメラや手の位置に不変な識別法となる。顔面微小特徴点群に基づく個人認証の従来研究では、動画像列から抽出したホクロやシミなど特徴点群に因子分解法を適用して、この形状空間を生成していた。しかし、手の場合にはその 3 次元形状を十分に反映した密な特徴点群を抽出・追跡することは極めて困難である。提案法では、これを解決するために、距離画像から生成した手形状の 3 次元メッシュを生成する。手の凸領域に含まれるメッシュを重点的にランダムサンプリングし、そのメッシュ上に配置した点群から形状空間を生成する。10 種類の手形状を用いた形状識別と姿勢推定の実験により提案法の有効性を示す。

キーワード: 手形状認識, 姿勢推定, 距離画像, 形状空間, 正準角

1. はじめに

顔や手などの 3 次元物体の形状識別や姿勢推定は、様々な画像認識システムを構築するうえで不可欠な要素技術である。これまでに様々な手形状の識別や姿勢推定の方法が提案されている [1–3]。

本稿では、これらの中で、手の形状クラスと姿勢の推定が同じ枠組みで実現できるという点で、形状空間に基づいた 3 次元物体認識法 [4] に着目する。形状空間は高次元ベクトル空間における 3 次元線形部分空間であり、1 つの 3 次元物体に対して一意に定まる。2 つの形状空間の類似度は両者の成す正準角により計ることができ、識別においては、入力距離画像から生成した入力形状空間と予め用意した辞書形状空間との類似度を計り、最も高い類似度クラスに入力距離画像を分類する。形状空間は座標系に対して不変なので、形状空間に基づく物体認識はカメラと物体の位置関係に依存しない。

動画像列において追跡した特徴点群の運動軌跡に対して、因子分解法を適用することで 3 次元復元が可能なこと [5] は良く知られているが、形状空間はその計算過程に

おいて生成される副産物である [4]。つまり、ある形状空間はその生成材料である特徴点群の 3 次元構造配置に対応している。従って、顔面微小特徴点に基づく個人認証のように、特徴点群の 3 次元構造そのものを識別対象と見なす場合には有効なフレームワークであるが、手形状を対象とする場合には、その 3 次元形状情報を十分に反映した密な特徴点が存在しないという大きな問題がある。仮にそのような特徴点を利用できるとしても、多数の特徴点群を安定に抽出・追跡することは難しい課題である。

この問題に対して、特徴点抽出・追跡や因子分解法を介さずに、1 枚の距離画像から形状空間を生成する方法を提案する。距離画像は距離センサから対象までの距離値をピクセル値として持つ画像であり、因子分解法を適用して得られる形状空間に含まれる情報を同様に含んでいる。提案法の概要を図 1 に示す。まず入力された距離画像から手形状の 3 次元メッシュを生成し、このメッシュ上に多数の点群をランダム配置する。この時、凸領域を重点的に選択することで、配置された点群が手形状の 3 次元情報を可能な限り反映したものにする。ここで N 個の点の 3 次元座標値を行ベクトルとして並べた $N \times 3$ の形状行列を考えると、この行列の 3 本の列ベクトルが N 次元ベクトル空間中の 3 次元形状空間の基底ベクトルとなる。

提案フレームワークが有効に働くためには、比較する 2

¹ 筑波大学大学院 システム情報工学研究科

² 筑波大学大学院 システム情報系

a) kikuchi@cvlab.cs.tsukuba.ac.jp

b) hinohide@cs.tsukuba.ac.jp

c) kfukui@cs.tsukuba.ac.jp

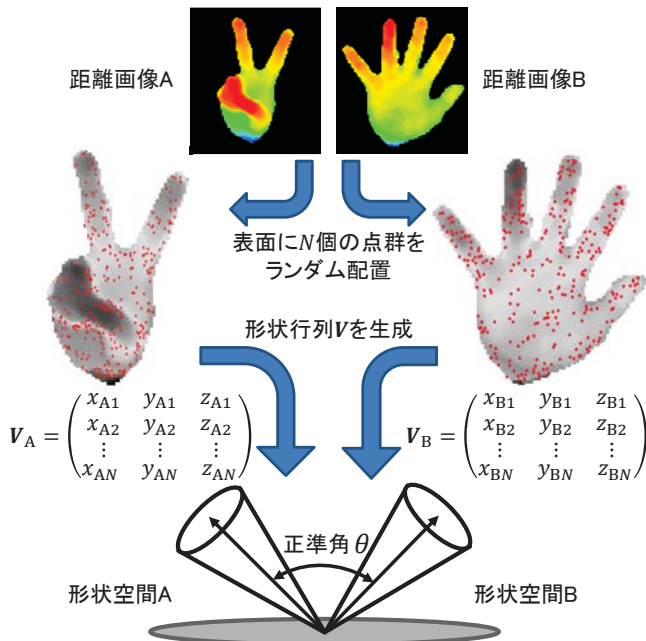


図1 提案法の概要: 距離画像から生成される形状空間に基づく3次元物体認識

つの手形状同士で各点の対応が取れている必要がある。対応が取れていない場合には、2つの形状空間を比較しても意味をなさない。この問題に対しては、上記の形状行列の3つの列ベクトルから計算される自己相関行列を用いた対応付け方法 [6] を適用する。この方法は自己相関行列が点群を表す座標系に依存せず一意に決まるという特性に基づき、一方の自己相関行列を他方に可能な限り近づけるように成分を並び替えていく。

本稿では、手形状のクラス識別に加え、その姿勢も併せて推定する方法を提案する。具体的には、先の処理で対応付いた2つの点群の関係から、一方を他方に変換する回転行列を求めることで、一方を基準とした他方の姿勢を推定する。

以下、本稿では第2章で因子分解法を用いた従来の形状空間の生成法とそれに基づく3次元物体形状識別の流れを概説する。第3章では距離画像から形状空間を生成する方法、およびそれに基づく識別・姿勢推定のフレームワークを提案する。第4章では10種類の手形状を用いた評価実験と結果に対する考察を述べる。第5章はまとめである。

2. 動画像列を用いた従来法

動画像列に対して因子分解法を適用して形状空間を生成する従来法について概説し、その問題点について述べる。

形状空間に基づいた顔画像認識の研究 [4] では、動画像列において追跡したホクロヤシミなどの微小特徴点群に対して因子分解法を適用して形状空間を生成していた。

この方法では、カメラに対して顔を動かしながら連続撮影した F 枚の時系列画像から、円形分離度フィルタ [7] を

適用してホクロヤシミのような P 個の顔面微小特徴点を抽出し、さらに KLT 法 [8] によって追跡する。次に、追跡した各特徴点の運動軌跡を表す計測行列 W ($2F \times P$) を作成する。因子分解法では、この W を運動情報のみを持つ行列 M ($2F \times 3$) と、形状情報のみを持つ行列 S ($3 \times P$) に分解する。この時 S は形状行列と呼ばれ、 S の列ベクトルが張る3次元部分空間が形状空間となる [9]。ここで、 W の分解は一意でないため、 S も一意に定まらないが、形状空間は特徴点群の相対関係が変化しなければ一意に定まる。2つの顔特徴点群の比較は、2つの形状空間同士の成す正準角に基づいて行う

しかし、この方法は3次元物体認識を行うとしながらも、識別対象は顔面微小特徴点群であり、3次元物体の形状そのものを対象にしたアルゴリズムとは言えなかった。手形状を対象とする場合、大きな問題としてそのような特徴点が存在しない事が挙げられる。仮にそのような特徴点が多分に取得できたとしても、多数の特徴点の抽出・追跡には大きな困難が伴い、実用的とは言えない。

3. 1枚の距離画像を用いた形状空間生成法

先に述べた従来法の課題を解決するために、1枚の距離画像から形状空間を生成する方法について述べる。

3.1 提案法の流れ

形状空間は対象物上の点群の3次元座標を反映したものである。従って、対象物から多数の点群の3次元座標を取得できれば、形状空間は因子分解法を介さずにそれから直接計算できる。提案法では、距離画像から得た手表面メッシュ上に点群をランダム配置することで形状空間の生成を行う。提案法は以下の3ステップの処理から構成される。

第1ステップ: 形状空間の生成

第2ステップ: 点群の対応付け

第3ステップ: 手形状の識別と姿勢の推定

以上を踏まえ、各ステップの処理をさらに要素分解した提案法の全体の流れを図2に示す。これに従って、全体の流れを説明する。

3.1.1 学習フェーズ

まず、学習フェーズにおける第1ステップの各処理について述べる。識別する手形状のクラス数を C 、各クラスごとの学習データ数を I とした時、学習データの距離画像を $I_{c,i}$ ($c = 1, \dots, C; i = 1, \dots, I$) とする。なお、簡単のため各クラスで学習データの数は同一であるとする。 $I_{c,i}$ に対して、はじめに手領域の切り出しを行う。本研究で用いるデータは、距離センサに対して手を突き出すように撮影しているため、抽出すべき手領域の距離値分布と背景領域の距離値分布に隔たりが生じる。この分布の隔たりから、手領域とその他領域を区別することで切り出しを行う。次の処理として、切り出された $I_{c,i}$ から

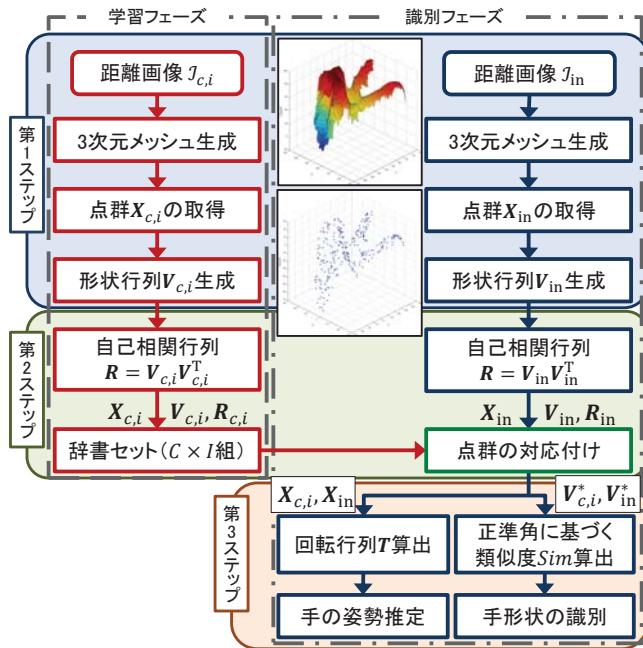


図 2 提案法の全体の流れ

手の表面形状の 3 次元メッシュを生成する。全メッシュの表面積に対する各メッシュ面積の割合をそのメッシュが選択される確率として、ランダムにメッシュを選択する。選択されたメッシュ内に点を配置し、その 3 次元座標 $\mathbf{x}_{c,i,n} = (x_{c,i,n}, y_{c,i,n}, z_{c,i,n})^T$ を取得する。この手続きにより N 点を取得し、点群 $\mathbf{X}_{c,i} = \{\mathbf{x}_{c,i,n} | n = 1, \dots, N\}$ を得る。点群 $\mathbf{X}_{c,i}$ の 3 次元座標値を行ベクトルとして並べた行列を形状行列 $\mathbf{V}_{c,i} = (\mathbf{x}_{c,i,1}, \dots, \mathbf{x}_{c,i,N})^T$ とする。以上より、画像 $I_{c,i}$ に対応する手形状の形状空間を、 $\mathbf{V}_{c,i}$ の各列ベクトルが張る 3 次元部分空間として得る。

第 2 ステップでは、点群の対応付けを行うために形状行列の自己相関行列 $\mathbf{R}_{c,i} = \mathbf{V}_{c,i} \mathbf{V}_{c,i}^T$ を求める。辞書セットとして $\mathbf{X}_{c,i}$ と $\mathbf{V}_{c,i}$, $\mathbf{R}_{c,i}$ のセットを $C \times I$ 組保持する。

3.1.2 識別フェーズ

識別フェーズは、自己相関行列を求める処理まで学習フェーズと同一である。識別フェーズにおいて、入力された距離画像を I_{in} , 取得した点群を $\mathbf{X}_{in} = \{\mathbf{x}_n | n = 1, \dots, N\}$, 形状行列を $\mathbf{V}_{in} = (\mathbf{x}_1, \dots, \mathbf{x}_N)^T$, \mathbf{V}_{in} の自己相関行列を $\mathbf{R}_{in} = \mathbf{V}_{in} \mathbf{V}_{in}^T$ とする。

識別フェーズの第 2 ステップでは、 \mathbf{X}_{in} と任意の辞書データにおける点群 $\mathbf{X}_{c,i}$ を対応付ける。この点群の対応付け問題に対して、松本らの提案した形状行列の自己相関行列を利用する方法 [6] を適用する。この方法ではまず、 \mathbf{R}_{in} と $\mathbf{R}_{c,i}$ における行ベクトルのそれぞれの要素をソートする。 \mathbf{R}_{in} と $\mathbf{R}_{c,i}$ のソートされた行ベクトル間の全組み合わせにおいて、 L_1 ノルムを計算する。この L_1 ノルムが最小となる組み合わせにおいて、それらの行ベクトルに対応する点同士を対応付ける。これを N 組の対応関係が求められるまで繰り返すことによって、 \mathbf{X}_{in} と $\mathbf{X}_{c,i}$ の対応付

けを行い、 $\mathbf{V}_{c,i}$ の行ベクトルの並び順を入れ替える。この対応付けを \mathbf{X}_{in} と辞書セット内の全データ $C \times I$ 個に対して行う。

第 3 ステップでは、手形状の識別と手の姿勢を推定する。まず手形状の識別について述べる。 \mathbf{V}_{in} と $\mathbf{V}_{c,i}$ に対して、Gram-Schmidt の正規直交化法を適用した形状行列をそれぞれ \mathbf{V}_{in}^* と $\mathbf{V}_{c,i}^*$ とする。この時、 \mathbf{V}_{in}^* と $\mathbf{V}_{c,i}^*$ の各列ベクトルは、各形状行列における形状空間を張る正規直交基底となる。求めた正規直交基底から、 \mathbf{V}_{in}^* と辞書セットの全ての $\mathbf{V}_{c,i}^*$ の間で正準角を計算し、類似度を定義する。正準角およびそれに基づく類似度の定義は 3.5 で述べる。定義した類似度が最も高くなる辞書形状空間の所属するクラスを \hat{c} とした時、このクラス \hat{c} を I_{in} の手形状として識別する。

手形状の姿勢推定において、対応付いた \mathbf{X}_{in} と手方向の基準となる $\mathbf{X}_{c,i}$ が可能な限り一致するような回転行列 \mathbf{T} を求める。提案法では \mathbf{T} を求めるために、Procrustes 解析 [10] を適用する。

3.2 点群の取得

距離画像 I の手形状から形状空間を生成するために、手表面から点群を取得する。この時どのような点を取得すればよいかが問題となる。提案法では、以下の 2 点に注意した点群の取得方法を考える。

- (i) 手形状の全体を表現できる点群であること。
- (ii) 特徴的なキーポイントの決定を要しない点群であること。

(i) は、手形状を識別する上で、手形状全体をまんべんなく表現できていなければ正確な識別は不可能であるという理由からである。(ii) は、手形状から安定したキーポイントを得ることは難しいため、本研究では考慮すべき事項とする。

以上を考慮し、提案法では次の 5 段階の処理によって点群を取得する。

- (1) 指先のような手形状に特徴的な凸領域を検出する。
- (2) 距離画像から得た手形状の 3 次元メッシュを生成する。
- (3) 各メッシュの面積を計算する。この時、非凸領域のメッシュ面積が小さくなるような係数を乗ずる。
- (4) 総面積に対する各メッシュの面積比をそのメッシュを選択する確率とし、この確率に応じてランダムに選択したメッシュ上に点を配置する。

3.2.1 凸領域の検出

細く尖った指の部分は、手形状を特徴付ける 1 つの要素である。従って、指部分のような凸領域を検出し、この領域に重みをおいて点群を配置すれば、手の 3 次元形状をよく表せると考えられる。本研究では凸領域の検出を、混合ガウス分布によるクラスタリング [11] を用いて行う。混合ガウス分布モデルは、複数のガウス分布の線形和からなる確率モデルである。モデルの推定には EM アルゴリズム



図 3 凸領域検出の例

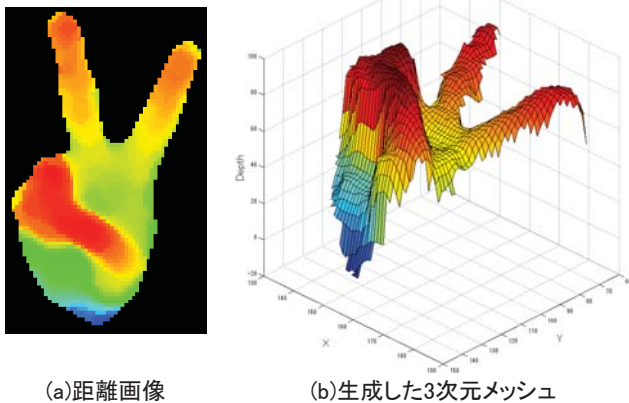


図 4 3次元メッシュの生成例

ム [12] を用いる。このモデルに従い、距離画像 I の手領域に含まれる画素をクラスタリングし、凸領域を検出する。クラスタリングを行うためには、カテゴリ数 K を決定しなければならない。本研究では、カテゴリ数の決定にベイズ情報量規準 (Bayesian Information Criterion: BIC [13]) を用いる。実際の距離画像に対してクラスタリングを行った結果を図 3(b) に示す。

次に、クラスタリングされた領域から凸領域を検出する。手形状の距離画像における凸領域は、非手領域と隣接している画素が多いことに着目し、クラスタリングされた領域内において非手領域と隣接する画素の割合が一定以上のクラスタを凸領域として検出する。実際に凸領域を検出した例を図 3(c) に示す。

3.2.2 3次元メッシュの生成

距離画像はセンサから対象までの距離値を格納した画素の集合であり、これらの画素から点群を直接サンプリングすることも考えられるが、本研究では、ノイズの影響を低減し、識別により有効な点群を生成するために、手形状を3次元メッシュで近似した上で、これに対して点群をランダム配置する。今回、3次元メッシュの生成には MATLAB の標準ライブラリである `griddata` 関数を距離画像の画素に適用した。実際に距離画像から生成された3次元メッシュの例を図 4 に示す。

3.2.3 メッシュ上への点群の配置

検出した凸領域と生成した3次元メッシュから、3次元空間上の点群 $\mathbf{X} = \{(x_n, y_n, z_n)^T | n = 1, \dots, N\}$ を取得する。この時、メッシュの面積 A_j に比例した確率で j 番目のメッシュを取り出すものとする。ただし、凸領域に重みをおいて点群の配置を行うために、凸領域のメッシュ面積に対して非凸領域のメッシュ面積が小さくなるような係数を乗じて面積を調整する。調整を行った j 番目のメッシュの面積 \tilde{A}_j を次式で定義する:

$$\tilde{A}_j \propto \begin{cases} A_j & (\text{凸領域に含まれる場合}) \\ 0.5A_j & (\text{otherwise}) \end{cases} \quad (1)$$

調整を行ったメッシュの総面積 \tilde{A} とした時、 j 番目のメッシュに対して、そのメッシュを選択する確率を次式で定義する:

$$p_A(j) = \frac{\tilde{A}_j}{\tilde{A}} \quad (2)$$

この確率 $p_A(j)$ に従ってランダムにメッシュを選択し、選択されたメッシュの表面上に点をランダムに配置する。点を規定の点数 N 点配置して得られたデータ行列を \mathbf{X} とする。

3.3 形状行列の生成

\mathbf{X} の各座標値を行ベクトルとし、その行ベクトルを列方向に並べた行列を形状行列 \mathbf{V} とする:

$$\mathbf{V} = \begin{pmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ \vdots & \vdots & \vdots \\ x_N & y_N & z_N \end{pmatrix} \in \mathbb{R}^{N \times 3} \quad (3)$$

形状空間は、 \mathbf{V} の列ベクトルで張られる3次元部分空間として定義される。

2つの形状空間の角度である正準角を計算することによって、2つの形状空間の類似度が定義できる。しかし点群の順番、つまり \mathbf{V} の行ベクトルの順番が異なれば、張られる形状空間も変化してしまう。従って、2つの形状空間同士の正準角を計算するためには、2つの点群の対応関係を求める必要がある。

3.4 点群の対応付け

生成された2つの形状空間から正準角を計算するためには、点群同士の対応付けが必要となる。ここで対応付けたい2つの点群は、各点の座標値しか情報を持たないため、この座標情報のみから対応関係を求めなければならない。さらに、それぞれの点群は手の形状表面からランダムに配置されたものであり、例え同じ手形状であっても点群の座標が完全に一致することはないため、ノイズに対してロバストな対応付けが求められる。また、同一の手形状であっ

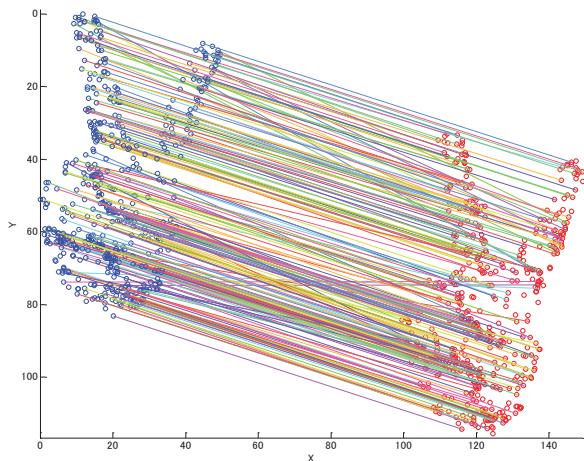


図 5 点群を対応付けた結果の例

でも回転による変動が加えられ、配置された点群にも回転が付加されている場合があるため、点群の回転に対応した対応付け手法が必要である。

このような問題に対して本研究では、松本らの提案した形状行列の自己相関行列を利用した点群の対応付け方法 [6] を適用する。自己相関行列は回転に不変であるため、これは点群の回転に影響を受けない対応付け方法である。また、点群の座標情報のみから対応付けが可能であり、ノイズにロバストな対応関係が求められる性質がある。詳細なアルゴリズムは、参考文献を参照されたい。

ほぼ同じ手形状の異なる距離画像から取得した2つの点群の対応付け結果の例を図5に示す。いくつかの対応において誤りが見られるが、全体的には正しく対応関係が求められていることが分かる。

3.5 正準角に基づく手形状の識別

互いに対応付けられた点群に基づく入力画像及び学習画像に対応する形状行列に対して、Gram-Schmidtの正規直交化法を適用したものを V_{in}^* , $V_{e,i}^*$ とする。 V_{in}^* と $V_{e,i}^*$ の各列ベクトルは、各々の形状空間における正規直交基底となる。この正規直交基底から、2つの形状空間同士において両者のなす正準角を計算し、類似度を定量的に評価することができる。

まず、部分空間同士の角度である正準角を定義する。 L 次元部分空間 \mathcal{V}_1 と M 次元部分空間 \mathcal{V}_2 ($L \leq M$) に対して、 L 個の正準角 ($0 \leq \theta_1 \leq \theta_2 \leq \dots \leq \theta_L \leq \frac{\pi}{2}$) を次式で定義する。

$$\cos^2 \theta_i = \max_{\substack{\mathbf{u}_i \perp \mathbf{u}_j, \mathbf{v}_i \perp \mathbf{v}_j \\ 1 \leq i, j \leq L, i \neq j}} \frac{\langle \mathbf{u}_i, \mathbf{v}_i \rangle^2}{\|\mathbf{u}_i\|^2 \|\mathbf{v}_i\|^2} \quad (4)$$

ここで $\langle \cdot, \cdot \rangle$ と $\|\cdot\|$ はそれぞれユークリッド内積とノルムを表し、 $\mathbf{u}_i \in \mathcal{V}_1$, $\mathbf{v}_i \in \mathcal{V}_2$, $\|\mathbf{u}_i\| \neq 0$, $\|\mathbf{v}_i\| \neq 0$ である。 \mathcal{V}_1 , \mathcal{V}_2 を張る正規直交基底をそれぞれ並べた行列を $\Phi = (\phi_1 \dots \phi_L)$, $\Psi = (\psi_1 \dots \psi_M)$ とすると、第 i 正準

角 θ_i の余弦 $\cos \theta_i$ は、 $\Phi^T \Psi$ あるいは $\Psi^T \Phi$ の i 番目に大きい特異値 σ_i として得られる [14]。

形状空間は3次元部分空間であるため、形状空間同士から3個の正準角 θ が算出できる。これらを用いて2つの形状空間 S_1 と S_2 との間で類似度 Sim を次式で定義する。

$$Sim(S_1, S_2) = \frac{1}{3} \sum_{i=1}^3 \cos^2 \theta_i^{(1,2)} \quad (5)$$

Sim は正規化量であり、2つの形状空間が完全に一致した場合 $Sim = 1.0$ となり、完全に直交した場合には $Sim = 0.0$ となる。

識別フェーズでは、入力データの形状空間に対して、辞書セットに保持してある全ての形状空間との正準角から上で述べた類似度 Sim を算出し、最も Sim の高かった辞書データの所属するクラスに手形状を識別する。

3.6 手の姿勢推定

取得した点群は、先の対応付け手法によって対応関係が求められている。従って、対応付いた点群同士が可能な限り一致するような回転行列を見つけることができれば、基準とする手形状に対する入力された手形状の回転角を求めることができる。

回転行列の推定のために、Procrustes解析 [10] を用いる。Procrustes解析は、対応関係が既知である点群に対して、対応付いた点同士の二乗距離の総和が最小となるような回転・反射・スケール・変位を求める手法である。本研究ではスケールと変位について正規化が行われているものとして、回転と反射のみについて説明する。

基準となるモデルの点群を \bar{X} , 回転角を求めたい点群を X とする。ここで、直交回転と反射を表す行列を T とし、変換された点群を TX と置くと、Procrustes解析は次のように定式化される。

$$\begin{aligned} \operatorname{argmin}_T \quad & \sum_{i=1}^3 \sum_{j=1}^N \|\bar{X} - TX\|_F^2, \\ \text{s.t.} \quad & T^T T = I \end{aligned} \quad (6)$$

ただし、 $\|\cdot\|_F$ はフロベニウスノルムである。行列 T から算出した回転角に基づき、対応付いている点群が基準モデルに対してどれだけ回転しているのかを求める。この手続きにより、入力された手の方向を求めることが可能となり、手の姿勢を定量的に推定できる。

4. 実験

提案法の有効性を検証するため、実際に撮影した距離画像を用いて3種類の実験を行った。まず、実験に用いるデータについて説明し、手形状の識別に関する2種類の実験について述べる。続いて、手の姿勢推定に関する実験について述べる。

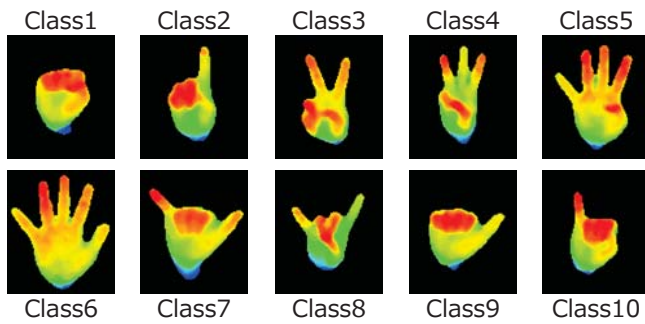


図 6 実験で識別すべき手形状

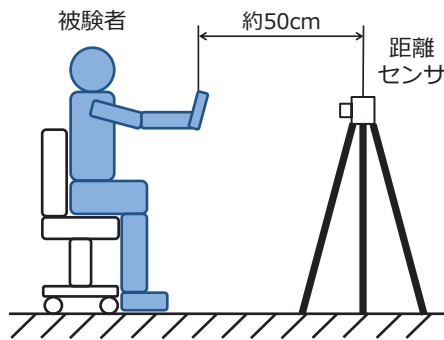


図 7 実験データの撮影環境

4.1 実験条件と評価データ

実験で識別すべき手形状は、図 6 に示す 10 クラスとする。実験に用いる手形状は、できる限り互いに類似した手形状となるように選択した。例えば、Class1~Class6 までは、それぞれ指を 1 本ずつ開いていった形状となっている。また Class7 と Class8 は共に両端の指を拡げており、極めて類似した形状となっている。さらに、これらの手形状と指 1 本の有無の違いのみを持つ手形状として Class9 と Class10 を用意した。

距離画像の撮影に際して、距離センサに SoftKinetic 社の DS325 を使用した。DS325 は、対象に照射した赤外線が反射して返ってくるまでの時間から距離を計測する Time of Flight (TOF) 方式の距離センサであり、 240×320 ピクセルの距離画像を取得可能である。データの撮影環境の模式図を図 7 に示す。被験者 1 名はキャスター付き椅子に座り、突き出した手と三脚に固定された距離センサまでの距離が約 50cm となるよう撮影してデータを収集した。収集したデータ 1 セットに付き各クラス 25 枚、つまり $25 \text{ 枚} \times 10 \text{ クラス} = 250 \text{ 枚}$ の距離画像を撮影する。撮影した後、一度椅子から立ち上がり、椅子を動かしてから座り直し、再度データを収集する。この動作には、撮影環境の差異によるノイズを擬似的に再現するという意図がある。これらの撮影を 4 回繰り返すことで計 4 セットのデータセット、距離画像にして $250 \text{ 枚} \times 4 \text{ セット} = \text{合計 } 1,000 \text{ 枚}$ を収集した。距離画像を撮影する際には、ノイズを加えるために意図的に手を動かしながら撮影を行ったため、同じ手形状

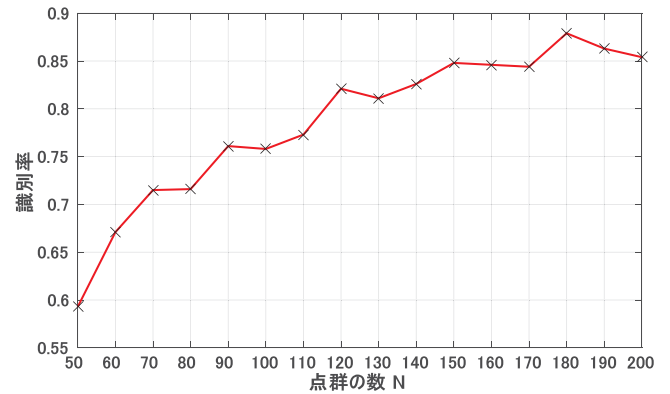


図 8 点数 N が変化した場合の識別率

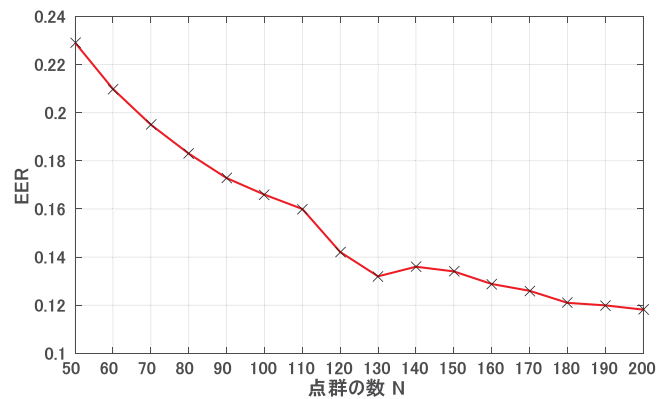


図 9 点数 N が変化した場合の EER

の距離画像であっても、見え方が大きく変化したデータとなっている。

4.2 手形状の識別

提案法における手形状の識別性能を評価するために、2 つの実験を行った。1 つ目の実験では、形状行列 V を構成する点群の点数 N を変化させた際に、手形状の識別性能がどのように変化するかを調べる。2 つ目の実験では、点群を配置する際における凸領域検出の有効性を検証するために、凸領域の検出を行わなかった場合との識別性能を比較する。評価実験では 4 セット収集したデータセットのうち、3 セットを学習データ、残り 1 セットをテストデータとした全組み合わせで識別実験を行う交差検証を行った。

4.2.1 点数の変化と識別性能の関係

配置する点群の点数 N を変化させた時、各点数における識別率とエラー等価率 (Equal Error Rate: EER) をそれぞれ図 8 と図 9 に示す。この実験結果から、全体的な識別率は N の増加とともに改善されていく傾向にあることが分かる。一方、必ずしも識別率は配置する点数に対して単調に改善するとは限らないことも分かる。これは提案法において、点群の取得方法にランダムな要素が存在するため、その影響で識別率が変動してしまうことによるものと考えられる。また、EER においても全体的な識別性能の改

Class	識別されたクラス										
	1	2	3	4	5	6	7	8	9	10	
実際のクラス	1	56	1	1			1	1		21	19
	2	1	77	1				1	2	13	5
	3			100							
	4			11	88	1					
	5				1	85	14				
	6					10	90				
	7	1						45		53	1
	8								100		
	9	1							15		84
	10	10	3	1				1	1	1	83

図 10 凸領域検出を行わない場合の混同行列 ($N = 200$ 点): 各セルは分類されたデータの百分率を表す. class1 と class7 において, 類似した手形状へ多く誤識別されている.

Class	識別されたクラス										
	1	2	3	4	5	6	7	8	9	10	
実際のクラス	1	94					1		3	2	
	2		81	1		5		1	1	8	3
	3			96	3					1	
	4		1	14	83	1				1	
	5			3		94	3				
	6			1		11	88				
	7			1		1	1	57	15	18	7
	8							1	99		
	9	3	4			2	2	12	1	76	
	10	3	7	1				3			86

図 11 凸領域検出を行う場合の混同行列 ($N = 200$ 点): 各セルはデータの百分率を表す. 凸領域検出を行わない場合に比べ, 識別率が向上している.

善は見られるが, $N = 140$ の時のように改善が見られない場合が見受けられる. これも上記の理由と同じく, 点群の取得方法におけるランダム性によると考えられる.

以上より, 形状空間を構成する点群の点数 N を増やすことによって, 手形状の識別性能を改善させることができる. しかし, 点数が増えると計算コストが増大してしまい, それと共に処理に要する時間も増えてしまう. さらに, 点数が増えるに従って識別性能の改善度合いが低下していることが見て取れる. 最適な N は問題毎に予備実験により決定する必要がある.

4.2.2 凸領域検出の有効性

凸領域の検出が有効であるかどうかを調べるため, 検出を行わない場合と行う場合を比較する. 検出を行わない場合とは, 手全体から無制約に点群をランダム配置することに相当する. ここでは $N = 200$ に統一し, 凸領域の検出を行わない場合と行う場合の混同行列をそれぞれ図 10

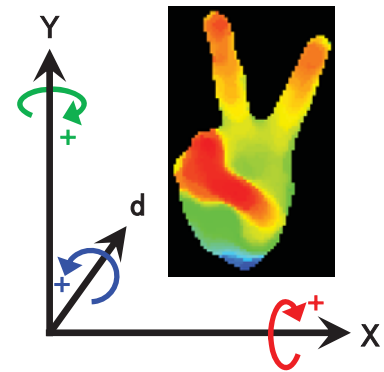
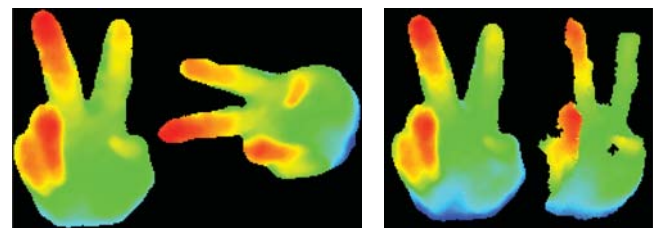


図 12 軸周りの回転角における正の方向



(a)基準モデル(左)と姿勢を推定したいモデル(右)
(b)算出した回転角により回転させたモデル(右)

図 13 基準モデルと姿勢推定対象の距離画像ならびに推定に基づき回転させた結果

と図 11 に示す. 各セルは, 分類されたデータの百分率を表す.

まず, 凸領域の検出を行わない場合について考察する. 7つのクラスにおいて識別率が80%を超えるが, いくつかのクラスにおいて識別率が低くなり, 特定のクラスへ誤った識別をしてしまっている. 中でも Class1 と Class7 はその傾向が顕著に現れている. Class1 は, Class9 と Class10 への誤った識別が多く存在する. これらの形状は, Class9 対しては親指の開閉による違い, Class10 に対しては小指の開閉の違い, といった微小な形状差異しかないため, 識別が不安定になったためだと考えられる. また, Class7 は Class9 への誤識別が非常に多くになっており, これも小指を開閉しているか否かの差異である. 従って, 手表面全体から無制約に点群のランダム配置を行うだけでは高い識別性能は得られないと言える.

一方, 凸領域を検出した場合には, 8つのクラスにおいて識別率が80%を超えている. Class1 においては, 凸領域の検出を行わなかった場合に比べ識別率が56%から94%へ大きく改善していることが分かる. また, Class7 についても識別率が45%から57%へ改善されている. 従って, 凸領域の検出を行うことで, 識別性能の向上が確認できた.

4.3 手の姿勢推定

本実験では提案法により手の姿勢推定がどの程度可能かを検証する. 各軸周りの回転の正方向を図 12 のように定める. 距離画像から真の手方向を定めるのは難しい. そこ

で図 13 (a) に示すように、基準となる手形状に対して d 軸周りにおおよそ $+90^\circ$ 回転させた手形状の距離画像を撮影した。これらの距離画像セットに対して回転行列を求め、各軸周りの回転角度を求めた。結果は以下の通りである。

- x 軸周りの回転角： $+19.1^\circ$
- y 軸周りの回転角： -6.1°
- d 軸周りの回転角： $+77.1^\circ$

上記から、さらに定性的な判断を行うため、算出された回転角の結果に基づき距離画像を 3 次元的に回転させた (図 13 (b))。2 次元投影された情報を強制的に 3 次元空間内の情報として回転させているため、回転させた距離画像には欠けなどが生じているが、全ての指が概ね一致していることは分かる。以上より、提案法は回転による変動を伴った手形状に対しても、手の姿勢推定が可能であると言える。

5. まとめ

本稿では、距離画像から形状空間を生成する方法を提案し、これに基づいて手形状クラスの識別とその姿勢推定を行った。提案法の基本アイデアは、動画像列からの特徴点群の抽出・追跡といった不安定な処理は行わず、1 枚の距離画像から生成した手形状の 3 次元メッシュ上に点群を凸領域に重点的にランダム配置することで、3 次元形状をより正確に反映した潤沢な点群を得ることにある。これにより、その表面に密な特徴点群を有しない手に対しても形状空間に基づく 3 次元物体のフレームワークを適用可能とした。さらに識別に加えて、対応付いた点群の情報を用いて回転行列を求めることで、手形状の姿勢も同じ枠組みで推定できることを示した。

提案法の有効性を検証するために、3 種類の実験を行った。第 1 の実験では、形状空間を構成する特徴点の数に依存して識別性能が如何に変化するかを評価した。第 2 の実験では、凸領域の重点配置アルゴリズムの有効性を確かめた。第 3 の実験では、同一手形状を回転して撮影した 2 枚の距離画像に対して、回転角度が推定できることを確認した。

謝辞

本研究は JSPS 科研費 25282173 の助成を受けたものです。

参考文献

- [1] Ali Erol, George Bebis, Mircea Nicolescu, Richard D. Boyle, and Xander Twombly. Vision-based hand pose estimation: a review. *Computer Vision and Image Understanding*, Vol. 108, Issues 1–2, pp. 52–73, 2007.
- [2] Siddharth S. Rautaray, and Anupam Agrawal. Vision based hand gesture recognition for human computer interaction: a survey. *Artificial Intelligence Review*, Vol. 43, Issues 1, pp. 1–54, 2012.
- [3] Suarez Jesus. Hand gesture recognition with depth images: a review. *IEEE RO-MAN*, pp. 411–417, 2012.

- [4] 五十嵐洋介, 福井和広. 形状空間の幾何学的な関係に基づく三次元物体認識. 電子情報通信学会論文誌 D, Vol. J94-D, No. 7, pp. 1125–1134, 2011.
- [5] Carlo Tomasi, and Takeo Kanade. Shape and motion from image streams under orthography: a factorization method. *International Journal of Computer Vision*, Vol. 9, No. 2, pp. 137–154, 1992.
- [6] 松本清吾, 福井和広. 3D モデル探索のための形状アライメント. 電子情報通信学会技術研究報告, パターン認識・メディア理解 (PRMU), Vol. 112, No. 441, pp. 32–42, 2013.
- [7] 福井和広, 山口修. 形状抽出とパターン照合の組み合わせによる顔特徴点抽出. 電子情報通信学会論文誌 (D.II), Vol. J80.D.II, No. 8, pp. 2170–2177, 1997.
- [8] Carlo Tomasi, and Takeo Kanade. Detection and tracking of point features. Technical Report CMU-CS-91-132, Carnegie Mellon University, 1991.
- [9] Carlo Tomasi, and Takeo Kanade. A sequential factorization method for recovering shape and motion from image streams. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 8, pp. 856–867, 1997.
- [10] John R. Hurley, and Raymond B. Cattell. The procrustes program: producing direct rotation to test a hypothesized factor structure. *Behavioral Science*, Vol. 7, pp. 258–262, 1962.
- [11] C.M. ビショップ (著), 元太浩, 栗田多喜夫, 樋口知之, 松本裕治, 村田昇 (監訳). パターン認識と機械学習 上 ベイズ理論による統計的予測. シュプリンガー・ジャパン, 2012.
- [12] Arthur P. Dempster, Nan Laird, Donald Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 39, No. 1, pp. 1–38, 1977.
- [13] Gideon Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, Vol. 6, No. 2, pp. 461–464, 1978.
- [14] F. シャトラン (著), 伊理 正夫, 伊理 由美 (訳). 行列の固有値—最新の解法と応用. シュプリンガー・ジャパン, 2003.