

LMNN を用いた超二次関数表現に基づく基本形状認識

八馬 遼^{1,a)} 小篠 裕子^{1,b)} 斎藤 英雄^{1,c)}

概要：本稿では、距離センサにより得られた基本形状物体の一部から物体形状パラメータを推定し、そのパラメータを用いて物体認識を行う手法について述べる。本手法では、物体表現手法として超二次関数を用いる。距離画像から推定された超二次関数パラメータを特徴ベクトルとし、その特徴ベクトルを用いて距離計量学習の手法の1つである Large Margin Nearest Neighbor (LMNN) を利用して物体認識を行う。既存手法である k -Nearest Neighbors や Support Vector Machines と比較し我々の LMNN を用いた手法が最も精度の良い認識率となった。

Primitive Shape Recognition via Superquadric Representation using Large Margin Nearest Neighbor Classifier

RYO HACHIUMA^{1,a)} YUKO OZASA^{1,b)} HIDEO SAITO^{1,c)}

1. はじめに

近年、距離センサ技術の発達に伴いシーンや物体の3次元情報を Kinect [1] などを用いて容易に取得できるようになった。3次元点群を用いた物体認識はロボット操作やシーン理解にとってとても重要なタスクの一つとなっている。一方でどのように人間が物体を認識するかという研究も広く行われてきた [2], [3]。Biederman ら [2] は人間は物体を複数の基本形状物体の組み合わせとその配置から認識しているとした。我々は、そのような人間の知覚システムに基づく基本形状レベルの3次元一般物体認識の実現を目標に研究を行っている。

基本形状に基づく物体認識を実現するには、より簡潔に、かつ一貫性を持たせて基本形状を表現することが重要となる。そのような物体の表現方法の1つとして超二次関数がある [4]。超二次関数は基本形状物体の表現方法の1つであり、比較的少数のパラメータで様々な形状を表現することができる。

本稿では、基本形状物体を超二次関数を用いて表現する

ことで基本形状の物体認識を行う。超二次関数パラメータを用いて基本形状物体認識を行った既存研究としては、識別器として k -Nearest Neighbor (k NN) を用いた手法 [5] や、Support Vector Machines (SVM) を用いた手法 [6] が挙げられる。一方、近年、距離計量学習が高い認識率を出す識別手法として注目を浴びている。その中の一つに Large Margin Nearest Neighbor (LMNN) がある [7]。LMNN はマハラノビス距離を学習することで、 k NN を行う際の精度を上げる手法である。また、SVM は本来2クラスを識別するための識別超平面を学習に求める手法であったため多クラスの識別においては複数の2クラス識別器を組み合わせる必要があった。しかし、LMNN では多クラス識別のために複数の識別器を用意する必要はない。本稿では、LMNN を用いて超二次関数表現された基本形状を識別する。

Kinect などに見られる安価な距離センサが開発されるまでは、実物体に対して推定された超二次関数パラメータを用いて物体認識を行うことは困難であったため Raja ら [5] は、あらかじめ各実物体に対して推定された超二次関数パラメータとユークリッド距離のマッチングをとることで物体を識別していた。また、Xing ら [6] は仮想的な超二次関数パラメータを用意し、そのパラメータに対して SVM を用いることで識別を行った。そして我々の実験では Kinect v.1 [1] により取得された基本形状の実物体に対してデー

¹ 情報処理学会
IPSI, Chiyoda, Tokyo 101-0062, Japan

a) ryo-hachiuma@keio.jp

b) yuko.ozasa@keio.jp

c) hs@keio.jp

タセットを構築しそのデータセットに対して k NN や我々の既存手法である SVM を用いた超二次関数パラメータ識別 [8] と比較し本研究の LMNN を用いた手法がどれほど有効かを検証した。

2. 基本形状認識

我々は、超二次関数を用いて基本形状を表現し基本形状に基づく物体認識手法を提案する。本手法は2つの段階から成っている。まず、物体の3次元点群に対して超二次関数パラメータを推定する。次に、推定された超二次関数パラメータを特徴ベクトル \mathbf{F} とし、LMNN を用いて特徴ベクトル \mathbf{F} を識別する。

2.1 超二次関数

超二次関数は二次曲面の拡張であり、比較的少数のパラメータで様々な基本形状を表現できる手段として、コンピュータグラフィックス分野 [4] やコンピュータビジョン分野 [9] で用いられ始めた。超二次曲面を陰関数表現したものを下記に示す。

$$f(x_s, y_s, z_s) = \left\{ \left(\frac{x_s}{a_1} \right)^{\frac{2}{\varepsilon_2}} + \left(\frac{y_s}{a_2} \right)^{\frac{2}{\varepsilon_2}} \right\}^{\frac{\varepsilon_2}{\varepsilon_1}} + \left(\frac{z_s}{a_3} \right)^{\frac{2}{\varepsilon_1}} = 1. \quad (1)$$

上式で、パラメータ a_1, a_2, a_3 はそれぞれ x, y, z 軸方向のスケールを表し、パラメータ $\varepsilon_1, \varepsilon_2$ は z 軸方向と x - y 平面の直角具合を表す形状パラメータであり、 x_s, y_s, z_s は超二次曲面上の点を表す3次元座標、添え字 s は超二次関数を中心とした座標系であることを示す。図1に形状パラメータである $\varepsilon_1, \varepsilon_2$ を変化させたときの様々な超二次曲面を示す。

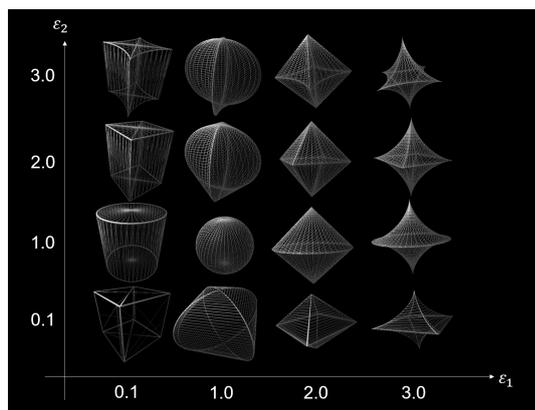


図1: $\varepsilon_1, \varepsilon_2$ の変化に伴う様々な超二次曲面

図1により、超二次曲面は $\varepsilon_1 \ll 1, \varepsilon_2 = 1$ のときに円柱のような形状となり、 $\varepsilon_1 \ll 1, \varepsilon_2 \ll 1$ のときに直方体のような形状となり、 $\varepsilon_1 = 1, \varepsilon_2 = 1$ のときに球のような形状をとることがわかる。

2.1.1 座標系変換

前述のように、超二次曲面は超二次曲面を中心とした座標系をとる。しかし、距離センサから得られた物体の3次元データが座標系の中心にくることはない。よって、世界座標系から超二次曲面を中心とした座標系への変換行列を求める必要がある。世界座標系のある点を (x_w, y_w, z_w) とし、超二次曲面を中心とした座標系のある点を (x_s, y_s, z_s) とすると、その2点の関係は式2のように書ける。式2において、変換行列 T は回転成分 $\mathbf{R}(\theta_x, \theta_y, \theta_z)$ と並進成分 $\mathbf{t}(t_x, t_y, t_z)$ から成っている。

$$\begin{pmatrix} x_s \\ y_s \\ z_s \\ 1 \end{pmatrix} = (\mathbf{R}|\mathbf{t}) \begin{pmatrix} x_w \\ y_w \\ z_w \\ 1 \end{pmatrix}. \quad (2)$$

2.1.2 超二次関数パラメータ推定

超二次関数は陰関数形式で表すことができるため、超二次関数 f とある点 $\mathbf{p}(x, y, z)$ に対して、 $f(x, y, z) > 1$ ならその点 p は超二次曲面の外側にあり、 $f(x, y, z) = 1$ なら点 p は超二次曲面上にあり $f(x, y, z) < 1$ なら点 p は超二次曲面の内側にある。また、前章における世界座標 (x_w, y_w, z_w) を用いて式1を書き直すと、超二次関数の5つのパラメータと位置姿勢パラメータの6つで合計11個のパラメータの式となる。物体の3次元点群が N 個与えられた時に、その3次元点群に対して超二次関数パラメータを推定する。超二次曲面は式1で表されるので、11個のパラメータを推定するには式3を最小化する必要がある。

$$\sum_{i=0}^N (f(x_{w_i}, y_{w_i}, z_{w_i}) - 1)^2. \quad (3)$$

しかし、距離センサで物体を撮影すると物体の部分的な3次元点群しか得ることができないため、様々な大きさの超二次曲面が式3を満たすことができる。よって、重みとしてスケールパラメータを導入することで大きさが最小となる超二次関数パラメータを推定することができる。そのときの最小化する式を下記に示す。

$$\sum_{i=0}^N (\sqrt{a_1 a_2 a_3} (f(x_{w_i}, y_{w_i}, z_{w_i}) - 1))^2. \quad (4)$$

関数 $\sqrt{a_1 a_2 a_3} (f(x_{w_i}, y_{w_i}, z_{w_i}) - 1)$ は11パラメータの非線形関数であるため、式3を非線形最小二乗問題として解くことができる。本論文では、Levenberg-Marquardt アルゴリズム [10] を用いてパラメータ推定を行った。さらに、 $\varepsilon_1, \varepsilon_2 < 0.1$ の時に式1は不安定になることが知られており、また図1が示すように $\varepsilon_1, \varepsilon_2 > 2.0$ の時は超二次曲面は凹状の形状となるが我々はそのような形状を推定の対象としていない。よって、式を最小化させる際に形状パラメータにおいては $0.0 < \varepsilon_1, \varepsilon_2 < 2.0$ 、スケールパラメータ

タにおいては $a_1, a_2, a_3 > 0.0$ の制約を導入する。しかし、Levenberg-Marquardt アルゴリズムは局所的最適化として解く手法なため解が局所解に陥ることもありパラメータの初期値を適切に定めることは重要となる。

2.1.3 初期値決定

推定する 11 個のパラメータは並進成分、回転成分、スケールパラメータ、形状パラメータの 4 種類に分けることができるため、各種類のパラメータで適切な初期値を決めることで局所解に陥らないようにする必要がある。物体のおおよその形状をあらかじめ求めるのはとても困難であるため、形状パラメータ $\varepsilon_1, \varepsilon_2$ の初期値は常に楕円体を表す 1.0 とする。次に、得られた物体の 3 次元点群の重心を算出することでそれを並進成分の初期値とする。また、回転成分の初期値を求めるためにまずは 3 次元点群に対して分散行列とその分散共分散行列の 3 つの固有ベクトルと固有値のペアを求める。そして、この固有ベクトルを回転成分の初期値に、そして固有値をスケールパラメータの初期値とする。

2.2 超二次関数認識

本論文では、超二次関数パラメータ $(\varepsilon_1, \varepsilon_2, a_1, a_2, a_3)$ を特徴ベクトル \mathbf{F} とし、特徴ベクトル \mathbf{F} に対して機械学習を用いることで基本形状物体を識別する。 L を識別するクラス数、 N を各クラスの学習用データ数、 l_i をそれぞれの特徴ベクトル \mathbf{F}_i に対応するクラスとすると、データセットは $\{(\mathbf{F}_i, l_i)\}_{i=1}^{L \times N}$ と書ける。共分散行列を M 、2 データ $\mathbf{F}_p, \mathbf{F}_q$ 間のマハラノビス距離を $D_M(\mathbf{F}_p, \mathbf{F}_q)$ とすると、マハラノビス距離は以下のように定義される。

$$D_M(\mathbf{F}_p, \mathbf{F}_q) = \sqrt{(\mathbf{F}_p - \mathbf{F}_q)^T M (\mathbf{F}_p - \mathbf{F}_q)} \quad (5)$$

LMNN [7] は距離計量学習と呼ばれる手法のうちの 1 つである。距離計量学習とは、学習データの統計的な特徴を学習し、識別に適したような距離空間を得る学習手法である。特に LMNN は対象データと同じクラスに属するデータとのマハラノビス距離を小さくし、違うクラスに属するデータとのマハラノビス距離をマージンを考慮しつつ大きくすることによって、より精度よい識別空間を作る手法である。しかし、対象データと同じクラスに属するすべてのデータとの距離を計算してしまうと計算コストが膨大になるため、LMNN ではターゲットネイバを導入する。ターゲットネイバとは、対象データと同じクラスに属し対象データの k 近傍の点のことである。さらに、LMNN ではマージンを導入し異なったクラスに属するデータはマージンを用いて対象データとの距離を大きくする。以上より、共分散行列 M は次式を最適化することにより学習される。

$$\text{Minimize } \sum_{ij} \eta_{ij} D_M(\mathbf{F}_i, \mathbf{F}_j) + c \sum_{ijh} (1 - \delta_{ih}) \xi_{ijh}$$

subject to:

$$D_M(\mathbf{F}_i, \mathbf{F}_h) - D_M(\mathbf{F}_i, \mathbf{F}_j) \geq 1 - \xi_{ijh}$$

$$\xi_{ijh} \geq 0$$

$$M \succeq 0.$$

(6)

ここで、 η_{ij} は \mathbf{F}_j が対象データ \mathbf{F}_j のターゲットネイバである場合は 1 をそうでなければ 0 をとるインジケータ関数である。また、 δ_{ih} は同様にクラス l_i とクラス l_h が同じクラスなら 1 をそうでなければ 0 をとる。そして c は重みのパラメータである。

3. 実験

本論文では、実物体に対する超二次関数パラメータ推定の妥当性と超二次関数パラメータ識別における LMNN の有効性を検証するために 2 つの実験を行った。まず、5 つの日用品の実物体に対して超二次関数パラメータ推定を行った。次に、基本形状実物体に対して推定されたパラメータでデータセットを構築し LMNN を用いて識別を行い k NN, SVM と比較することで LMNN の有効性を検証した。

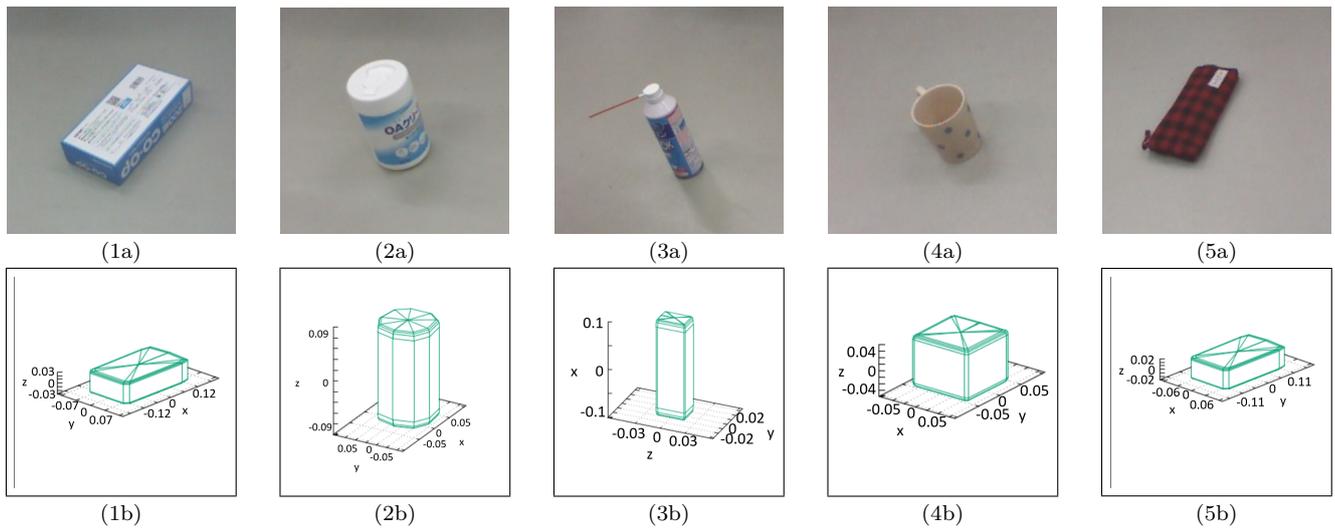
3.1 超二次関数パラメータ推定

まず、実物体に対する超二次関数パラメータ推定の結果を図 2 に示す。Kinect v.1 を用いて平面の上に置かれた物体を撮影し得られた 3 次元点群に対して RANSAC アルゴリズムを用いて平面を除去し、孤立点を除去するためにユークリッド距離を用いたクラスタリングを行った。

図 2 において (1a), (2a), \dots , (5a) は撮影した物体を、(1b), (2b), \dots , (5b) は推定された超二次曲面を示す。また、推定された各超二次関数パラメータを図 2 の下に示した。図 2 と図 1 を比べながら推定されたパラメータについての考察を行う。まずは、物体 (1a) の形状パラメータは $(\varepsilon_1, \varepsilon_2) = (0.10, 0.18)$ となっており、これは図 1 において直方体と示しているため妥当であることがわかる。また、物体 (1a) のスケールパラメータは $(a_1, a_2, a_3) = (0.12, 0.07, 0.03)$ となっているが物体 (1a) の実寸の縦横高さの比は ($width : height : length = 4.47 : 2.29 : 1.0$) であるためスケールパラメータも形状パラメータと同様に妥当であることがわかる。

3.2 基本形状認識

本実験では、5 種類の基本形状を用いた。用いた物体の例を図 3 に示す。図中の 1 の物体は 2 辺の長さが等しい直方体、2 の物体は立方体、3 の物体は円柱、4 の物体はピラミッド形状、5 の物体は立方体である。2 の物体と 5 の物



- (1) $(\varepsilon_1, \varepsilon_2) = (0.10, 0.18), (a_1, a_2, a_3) = (0.12, 0.07, 0.03)$
 (2) $(\varepsilon_1, \varepsilon_2) = (0.10, 0.80), (a_1, a_2, a_3) = (0.05, 0.05, 0.09)$
 (3) $(\varepsilon_1, \varepsilon_2) = (0.10, 0.10), (a_1, a_2, a_3) = (0.10, 0.02, 0.03)$
 (4) $(\varepsilon_1, \varepsilon_2) = (0.12, 0.10), (a_1, a_2, a_3) = (0.05, 0.05, 0.04)$
 (5) $(\varepsilon_1, \varepsilon_2) = (0.10, 0.10), (a_1, a_2, a_3) = (0.06, 0.11, 0.02)$

図 2: 超二次関数推定の結果 (a: RGB 画像 b: 超二次曲面)

体はともに立方体であるが、異なったスケールの立方体である。本実験ではこの 5 種類の物体を向き、置き方、位置をランダムに変えて各 240 データ撮影し取得した物体の 3 次元データを用いて 5 クラスの物体認識を行った。撮影には Microsoft Kinect v.1 を用いた。そして 240 データのうち各物体につき 80 データをテストデータ、160 データを学習用データとした。LMNN を有効性を検証するために、3 つの特徴ベクトルを用いて実験を行った。

$$\begin{cases} \mathbf{F}_1 = (\varepsilon_1, \varepsilon_2), \\ \mathbf{F}_2 = (a_1, a_2, a_3), \\ \mathbf{F}_3 = (\varepsilon_1, \varepsilon_2, a_1, a_2, a_3). \end{cases} \quad (7)$$

ネル, RBF カーネル) を用いた。表 1 により、すべての特徴ベクトル $\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3$ において RBF カーネルを用いたときが最も高い認識率となり、また特徴ベクトル \mathbf{F}_3 が他の特徴ベクトルと比べてすべてのカーネルにおいて最も良い認識率となった。このことにより、形状パラメータだけでなくスケールパラメータも超二次関数に基づく物体認識において有効であることがわかる。また、特徴ベクトル \mathbf{F}_1 と \mathbf{F}_2 を比べてみると、 \mathbf{F}_2 のほうがより良い認識率となっている。このことから、スケールパラメータよりも形状パラメータのほうが認識にはより有効であることがわかる。図 4 (a) は RBF カーネルで最適化により求めたパラメータが $C = 100.0, \gamma = 0.1$ の SVM の混同行列を示している。

表 1: SVM を用いた物体認識精度 (%)。

	SVM		
	Linear	Poly	RBF
F_1	45.3	51.2	62.3
F_2	42.5	38.8	52.0
F_3	60.8	64.2	73.5

表 2: NN と LMNN(k = 3,5) を用いた物体認識精度 (%)。

	kNN		LMNN	
	k=3	k=5	k=3	k=5
F_1	62.2	60.5	60.0	60.2
F_2	68.1	68.1	65.4	65.4
F_3	74.7	76.5	78.9	79.5

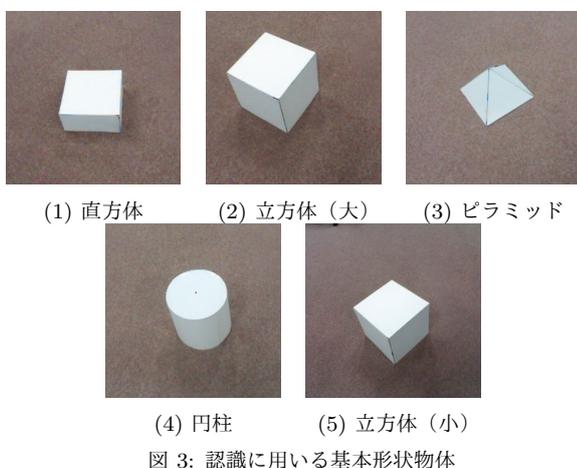


図 3: 認識に用いる基本形状物体

SVM においては、線形 SVM と非線形 SVM(多項式カー

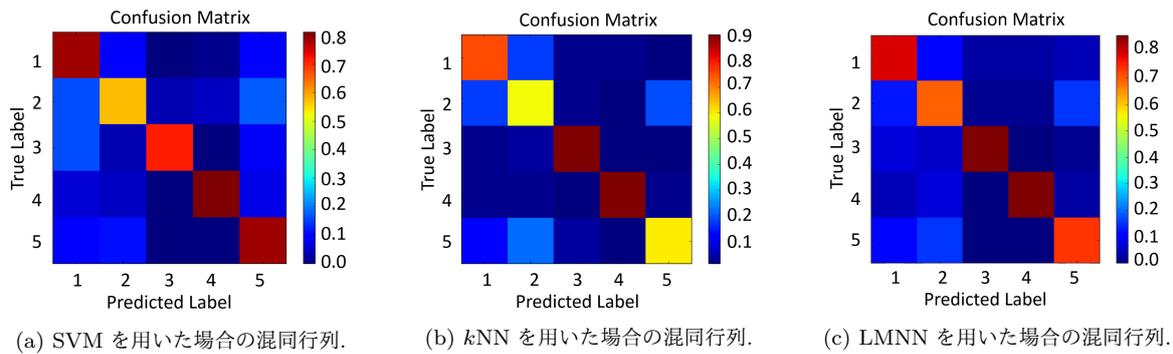


図 4: 各識別器における混同行列の比較

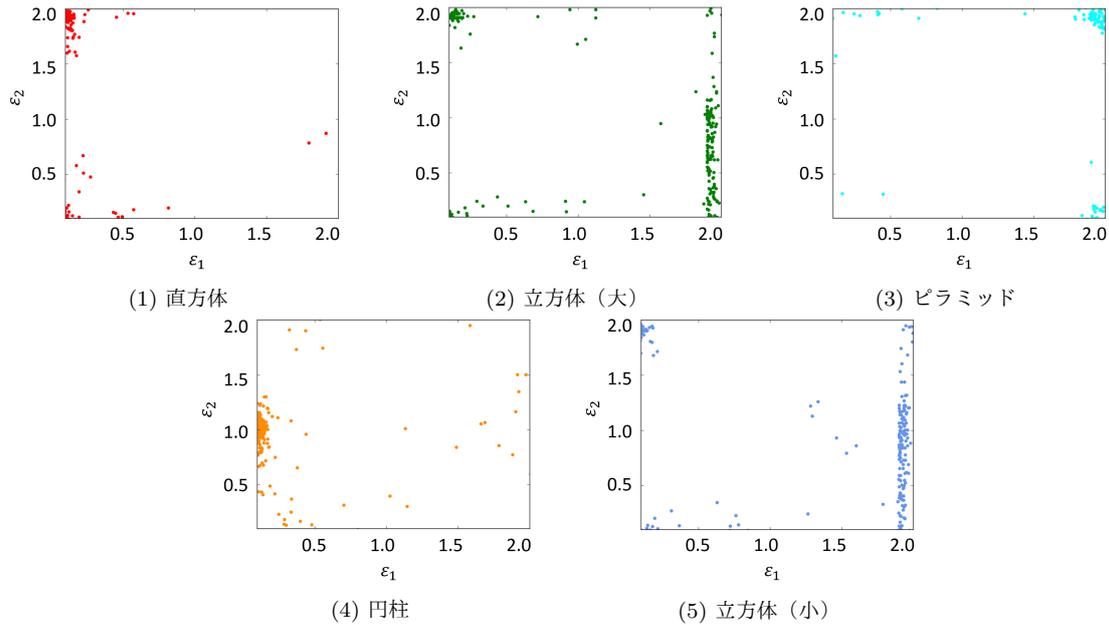


図 5: 各基本形状物体の推定された形状パラメータの散布図

次に、 $k = 3, 5$ の時の k NN と LMNN を用いた時の認識率の結果を表 2 に示し、 $k = 3$ の時の k NN と LMNN の混同行列を図 4 (b), (c) にまとめる。図 5 は推定されたパラメータ (ϵ_1, ϵ_2) の散布図を示している。図 5 (4) は円柱形の物体に対してのパラメータ推定結果の散布図を示している。この図を見ると、点が ($\epsilon_1 = 0.1, \epsilon_2 = 1.0$) あたりに集中しており、図 1 により超二次曲面は ($\epsilon_1 = 0.1, \epsilon_2 = 1.0$) のときに円柱形を成すため推定結果は妥当であるといえる。また、図 5 (2), (5) により大きい立方体と小さい立方体はスケールが異なるだけで同じ形状をしているので似た散布図となっていることがわかる。そして、図 5 (3) はピラミッド系の物体に対する推定結果の散布図を示している。この図によると、ピラミッド形の物体に対しての推定結果は ($\epsilon_1 = 2.0, \epsilon_2 = 0.1$) や ($\epsilon_1 = 2.0, \epsilon_2 = 2.0$) 付近に散布していることがわかる。これは図 1 によると、八面体を表していることがわかる。この理由としては超二次曲面は四面体 (ピラミッド型) の形状を成すことができないためであると考えられる。

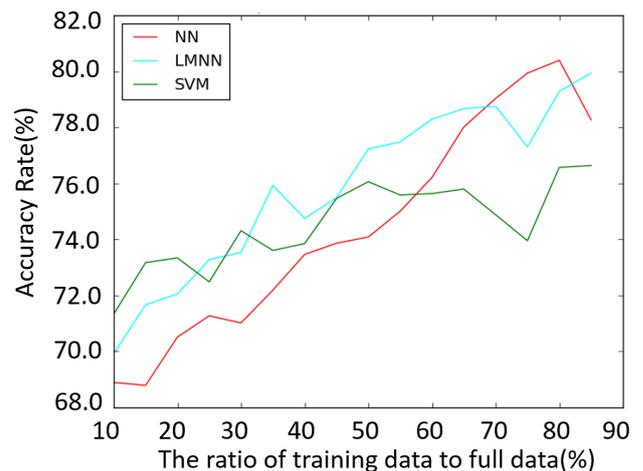


図 6: 学習用データの全体に対する割合

図 6 は全体のデータに対する学習用データに使うデータの割合を 10% から 85% まで変えたときに各識別器 (SVM, k NN, LMNN) でどのように認識率が変化したかを示している。3 つの識別器を比較したときに、SVM は過学習に強いいため学習データの割合が 10%, 20% といったように小さい

ときは k NN と LMNN に比べて SVM が良い認識率となっていることがわかる。しかし、学習用データの割合が大きくなるほど LMNN は SVM と比べて安定して高い認識率を保持していることがわかる。また、 k NN と LMNN を比較した際に LMNN は学習用データの割合が 75%と 80%の時は k NN のほうがよりよい認識率となっているが、85%の時は LMNN のほうが高い認識率となったことから全体的にデータ数が不足していたということがわかる。

4. まとめ

本論文では基本形状で物体を認識する重要性について述べ、基本形状表現として超二次関数を用いて超二次関数パラメータを距離計量学習手法の 1 つである LMNN を用いることで識別した。具体的には、基本形状実物体を撮影して得られた距離画像から、物体形状を表現する超二次関数パラメータを推定しそれらを特徴ベクトルとして LMNN を用いて基本形状物体認識を行った。また、実験としては実物体に対する超二次関数パラメータ推定の妥当性を検証するとともに、得られたパラメータを LMNN によって識別する有効性も確認した。

参考文献

- [1] Zhang, Z.: Microsoft kinect sensor and its effect, *IEEE multimedia*, Vol. 19, No. 2, pp. 4–10 (2012).
- [2] Biederman, I.: Recognition-by-components: a theory of human image understanding., *Psychological review*, Vol. 94, No. 2, p. 115 (1987).
- [3] Marr D, V.: A Computational Investigation into the Human Representation and Processing of Visual Information (1982).
- [4] Barr, A. H.: Superquadrics and angle-preserving transformations, *IEEE Computer graphics and Applications*, Vol. 1, No. 1, pp. 11–23 (1981).
- [5] Raja, N. S. and Jain, A. K.: Recognizing geons from superquadrics fitted to range data, *Image and vision computing*, Vol. 10, No. 3, pp. 179–190 (1992).
- [6] Xing, W., Liu, W. and Yuan, B.: Superquadric-based geons recognition utilizing support vector machines, *Signal Processing, 2004. Proceedings. ICSP'04. 2004 7th International Conference on*, Vol. 2, IEEE, pp. 1264–1267 (2004).
- [7] Weinberger, K. Q. and Saul, L. K.: Distance metric learning for large margin nearest neighbor classification, *Journal of Machine Learning Research*, Vol. 10, No. Feb, pp. 207–244 (2009).
- [8] 小篠裕子, 八馬遼, 斎藤英雄: SVM を用いた超二次関数モデル表現に基づく三次元物体認識 (情報論的学習理論と機械学習), 電子情報通信学会技術研究報告 = IEICE technical report : 信学技報, Vol. 116, No. 209, pp. 215–219 (2016).
- [9] Pentland, A. P.: Perceptual organization and the representation of natural form, *Artificial Intelligence*, Vol. 28, No. 3, pp. 293–331 (1986).
- [10] Press, W. H., Flannery, B. P., Teukolsky, S. A. and Vetterling, W. T.: Numerical recipes: the art of scientific computing, *Cambridge U. Press, Cambridge, MA* (1986).