

# 正準相関分析を用いた三次元物体のアノテーション

立間 淳司<sup>1,a)</sup> 青野 雅樹<sup>1,b)</sup>

**概要:**近年、産業・医療・映画など、様々な分野で三次元物体が利用されている。それにともない、検索や分類など、三次元物体の利活用を目的とした技術が注目されている。三次元物体のアノテーションは、三次元物体に対して、形状にあわせたラベルを付与する技術である。三次元物体にラベルを付与することで、文字列による管理・検索が可能となる。本研究では、三次元物体のアノテーションに関する基礎的な手法を示すとともに、三次元物体とラベルとの関連性を、正準相関分析により学習する手法を提案する。評価実験から、正準相関分析により学習することで、三次元物体の特徴量のみを用いた場合よりも、適合率の向上が見られた。

## 1. はじめに

三次元物体は、産業・医療・エンターテインメントなど、様々な分野で利用されている。さらに、近年、スマートデバイスの普及や Web ブラウザの機能向上により、三次元物体による豊かな表現を利用したアプリケーションやサービスが増加している。スマートデバイスには、三次元コンピュータグラフィックス API である OpenGL for Embedded Systems [16] が搭載され、拡張現実アプリケーションをはじめ、三次元物体が多く利用されている。また、Web ブラウザ上で三次元コンピュータグラフィックスを扱うための標準仕様である WebGL [17] が策定され、特別なプラグインをインストールすることなく、三次元物体を表示できるようになった。このような状況から、文書・画像データなどと同様に、三次元物体データが、爆発的に増加することが予想される。

大量にある三次元物体データの利活用を目的として、形状に基づいて検索を行う、三次元物体の形状類似検索技術が注目されている [14], [28]。目的とする形状の三次元物体を、検索することができれば、既存の三次元物体を再利用することも容易となり、制作コストの削減につながる [7]。

三次元物体の形状類似検索で問題となるのは、検索質問の提示方法である。一般に、三次元物体の形状類似検索では、検索質問として三次元物体そのものを提示する。そのため、利用者は、検索したいと考える三次元物体と形状が類似したものを、事前に所有している必要がある。

この問題に対して、Min らは、Web ページ検索などと同様に、文字列により三次元物体を検索する手法を提案した [22]。Min らの手法では、ファイル名やコメントなど、三次元物体に付与されている文字列情報から作成した、ベクトル空間モデルによる文書ベクトルを比較することで検索を行う。しかし、三次元物体に意味ある文字列情報が付与されていることは少なく、また、大量にある三次元物体に人手でラベル付けを行うことは、膨大な労力を必要とするため現実的ではない。

このような状況から、三次元物体に対して、その形状を表現するラベルを自動的に付与する、三次元物体のアノテーション技術が必要とされている。しかし、これまで三次元物体のデータとしての利活用を目的とした研究は、検索技術が主流であったため、アノテーション技術に関する研究は、いまだ事例が少ない [9], [24], [33]。そこで、本研究では、三次元物体の形状特徴量と  $k$ -近傍識別器によるベースライン手法を提案する。さらに、正準相関分析により形状とラベルとの関連性を学習する手法を提案する。

## 2. 関連研究

三次元物体のアノテーションに関する研究は、文字列による三次元物体の検索を目的として進められてきた。

三次元物体の検索技術の研究では、三次元物体の形状類似検索 (3D Object Retrieval, 3DOR) が主流である [14], [28]。3DOR では、三次元物体から、形状情報を表現した特徴量を抽出し、その特徴量を比較することで検索を行う。これまでも、様々な特徴抽出手法が提案されており、現在も研究が進められている。

3DOR において問題となるのが、検索質問の提示方法で

<sup>1</sup> 豊橋技術科学大学 情報・知能工学系  
Toyohashi University of Technology, Toyohashi, Aichi 441-8580, Japan

a) tatsuma@cs.tut.ac.jp

b) aono@tut.jp

ある。一般に、3DORの研究では、三次元物体そのものを検索質問とする。これは、利用者が、検索したいと考えている三次元物体と形状が類似したものを、事前に所有していることを仮定している。そのため、利用者は、三次元物体の作成などを行なっている専門家に限られてしまう。

これに対して、二次元スケッチ画像や文字列を検索質問とした、三次元物体の検索技術が研究されている [6], [15], [22]. 特に、文字列を検索質問とした場合、既存の Web ページ検索などと同様のインターフェースとなるため、利用者にとって扱い易いものとなる。

Min らは、ファイル名やコメントなど、三次元物体に付与された文字列情報から、ベクトル空間モデルによる文書ベクトルを作成することで、文字列による三次元物体の検索を実現した [22]. しかし、比較実験では、文書ベクトルによる検索は、形状特徴量による検索よりも、優れた検索性能を得ることができなかった。

Goldfeder らは、Min らの研究をふまえ、文書ベクトルの作成に必要な文字列情報に、三次元物体のアノテーションにより推定したラベルを用いる手法を提案した [9].

Goldfeder らのアノテーション手法は、特徴空間上での確率分布に従ってラベルを推定する。まず、三次元物体から、形状特徴量として 3D Zernike Descriptor [23] を抽出する。次に、テストデータと訓練データとの間で、形状特徴量の相違度を計算し、特徴空間上でテストデータの近傍にある訓練データを選び出す。そして、近傍にある訓練データに付与されたラベルが、テストデータに適しているかの確率を計算し、その値が閾値を超えたラベルを、推定結果として出力する。

Goldfeder らは、Google 3D Warehouse [10] から収集した 192,343 個の三次元物体のうち、ラベルが付与されたものを訓練データセットとして、Princeton Shape Benchmark (PSB) [26] に含まれる三次元物体に対してアノテーションを行った。そして、推定したラベルを用いて文書ベクトルを作成し、その検索性能を評価した。結果として、文書ベクトルを用いた検索は、3D Zernike Descriptor を用いたものと、同等の検索性能となった。

Ohbuchi らは、Goldfeder らと同様にして、アノテーションにより推定したラベルからなる文書ベクトルを用いた、三次元物体の検索を提案した [24].

Ohbuchi らのアノテーション手法では、まず、形状特徴量として Surflet Pair Relation Histograms [32] を抽出する。次に、テストデータと訓練データとを合わせたデータセットで、形状特徴量間の相違度によるグラフラプシアン行列を計算する。そして、グラフラプシアン行列で表される、特徴空間上のデータ間の関係をもとに、Manifold Ranking [34] を用いて、訓練データに付与されたラベルとテストデータとの適合度を計算する。この適合度が閾値を超えたラベルを、推定結果として出力する。PSB を用いた

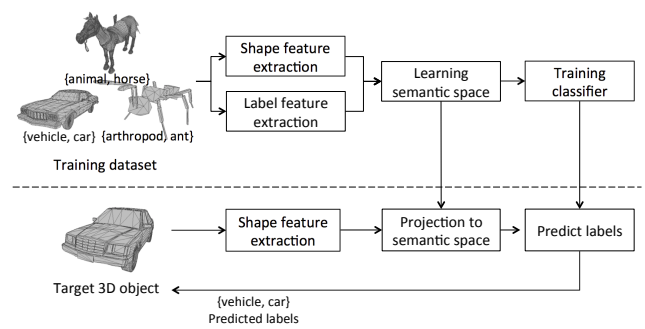


図 1 提案システムの概要

Fig. 1 Overview of the our 3D object annotation system.

実験では、推定したラベルによる文書ベクトルは、形状特徴量 Light-Field Descriptor [2] よりも、優れた検索性能となった。

この他に、Zhang らは、能動学習による三次元物体のアノテーション手法を提案している。Zhang らは、アノテーションによる内容に基づく情報検索の枠組みを考案し、その応用例として三次元物体データセットを用いた [33].

以上の研究では、文字列による三次元物体の検索を目的としており、推定されたラベルによる検索の適合性や網羅性を評価している。そのため、推定したラベル自体の正確性などは評価していない。このように、三次元物体のアノテーションの研究は事例が少なく、未だ初期段階にあると言える。

### 3. 提案システムの概要

図 1 は、本研究で提案する、三次元物体のアノテーションシステムの概要図である。システムの処理は、形状とラベルとの関連性を学習する訓練フェーズと、与えられたデータのラベルを推定するテストフェーズに分けられる。

訓練フェーズでは、まず、複数ラベルが付与された三次元物体からなる訓練データセットを用意する。そして、訓練データセットの各三次元物体から、形状特徴量を抽出する。また同時に、形状とラベルとの関連性を学習するため、ラベル特徴量を抽出する。次に、形状特徴量とラベル特徴量から、形状とラベルとの関連性を反映した部分空間を学習する。最後に、部分空間に射影した形状特徴量を用いて、ラベルを推定するための識別器を訓練する。

テストフェーズでは、利用者から、ラベルを推定したい三次元物体が与えられると、訓練フェーズと同様にして形状特徴量を抽出する。そして、得られた形状特徴量を部分空間に射影し、識別器を用いてラベルを推定する。

本研究では、形状特徴量と  $k$ -近傍識別器によりラベルを推定するベースライン手法と、正準相関分析により形状とラベルとの関連性を学習する手法を提案する。

## 4. 三次元物体とラベルからの特徴抽出

### 4.1 形状特徴量

三次元物体のアノテーションを行うためには、三次元物体の形状を数値で表した、形状特徴量を抽出する必要がある。形状特徴量の抽出手法は、これまでに、三次元物体の形状類似検索の分野において様々なものが提案されている [14], [28]。その内、本研究では、Bag-of-Visual-Words (BoV) に基づく手法を用いる。

BoV は、もともと、画像認識の分野で提案された特徴量である [3], [27]。BoV は、統計的言語処理における Bag-of-Words モデルのアナログであり、画像を局所特徴量の集合で表現する。この BoV の枠組みを、形状特徴量の抽出に応用した手法が提案されている [1], [8], [18], [19]。

図 2 は、BoV による形状特徴量の抽出手法の概要図である。まず、訓練データセットに含まれる三次元物体から、局所特徴量を抽出する。次に、抽出した局所特徴量群に対してベクトル量子化を行い、代表ベクトルを得る。この代表ベクトルを、文書の単語になぞらえて Visual Word と呼ぶ。テストデータが与えられると、同様にして局所特徴量を抽出し、最近傍の Visual Word に割り当てる。最後に、各 Visual Word に割り当てられた局所特徴量をカウントすることで、ヒストグラムを作成する。

本研究では、形状特徴量に、三次元物体よりレンダリングした Depth Buffer 画像から、局所特徴量として Dense SIFT (DIFT) [31] を抽出した、BoV に基づく特徴量を用いる。以降、この特徴量を、Depth Buffered Bag-of-Visual-Words (DBoV) と呼ぶ。

DBoV では、まず前処理として、Point SVD [29] を用いて三次元物体の姿勢正規化を行う。三次元物体では、製作者やソフトウェアにより、位置・大きさ・回転が任意となる。これら姿勢の任意性は、特徴量にも影響を与えるため、正規化する必要がある。Point SVD は、三次元物体の面上にランダムな点群を発生させ、点群の平均が三次元空間の原点となるよう平行移動することで位置の任意性を、点群を特異値分解することで求まる主軸が三次元空間の  $x$  軸・ $y$  軸・ $z$  軸に沿うように回転することで回転の任意性を、各頂点座標を原点からの最大距離で割ることで大きさの任意性を解決する。

次に、DBoV では、複数視点からレンダリングした Depth Buffer 画像より、局所特徴量を抽出する。Depth Buffer 画像は、視点から三次元物体の各面までの距離を、グレースケール画像で表現したものである。DBoV では、図 3 のように、正八面体を分割してできた Geodesic Sphere の頂点を視点として、計 18 方向から、 $256 \times 256$  ピクセルの大きさを、Depth Buffer 画像をレンダリングする。そして、各 Depth Buffer 画像から、局所特徴量として DIFT を抽出

する。

局所特徴量の抽出は、特徴点の検出と、検出した特徴点周辺での特徴抽出の二段階からなる。特徴点の検出手法には、特徴点検出アルゴリズムを用いて、輝度値の変化に応じた位置と周辺領域の大きさで特徴点を配置する Sparse Sampling と、輝度値の変化に関係なく、一定の間隔と大きさを格子状に特徴点を配置する Dense Sampling がある。DIFT は、この Dense Sampling により特徴点を配置し、SIFT 記述子 [20] を抽出したものである。DBoV では、8 ピクセル間隔、 $64 \times 64$ ,  $80 \times 80$ ,  $96 \times 96$  ピクセルの大きさを特徴点を配置した。

最後に、抽出した局所特徴量を Visual Word に割り当て、ヒストグラムを生成する。いま、訓練データセットの三次元物体から抽出した局所特徴量群に対してベクトル量子化を行い、 $p$  個の Visual Word  $V = \{\mathbf{v}_1, \dots, \mathbf{v}_p\}$  を得たとする。テストデータの三次元物体が与えられると、訓練と同様に局所特徴量を抽出し、最近傍の Visual Word に割り当てる。すなわち、ある局所特徴量  $\mathbf{s}$  と、それに対応する visual word  $\mathbf{v}(\mathbf{s})$  との間には、以下の関係がある。

$$\mathbf{v}(\mathbf{s}) = \min_{i=1, \dots, p} \|\mathbf{s} - \mathbf{v}_i\|^2$$

そして、各 Visual Word に割り当てられた局所特徴量の数を、Visual Word の出現頻度としてヒストグラムを作成する。ある Visual Word  $\mathbf{v}_i$  の出現頻度を  $tf_i$  とすると、DBoV による形状特徴量  $\mathbf{x}$  は、以下のように定義される。

$$\mathbf{x} = [tf_1, tf_2, \dots, tf_p]^T$$

### 4.2 ラベル特徴量

三次元物体の形状と付与されたラベルとの関連性を学習するため、ラベル特徴量を定義する。ラベル特徴量には、各ラベルの有無を表す二値ベクトルを用いる。

いま、三次元物体の訓練データセットがあり、各三次元物体に、データセット全体で  $q$  個のラベル  $w_1, \dots, w_q$  が付与されているとする。このとき、ある三次元物体のラベル特徴量  $\mathbf{y} = [y_1 \dots y_q]^T$  の要素  $y_i$  を、以下のように定義する。

$$y_i = \begin{cases} 1 & \text{ラベル } w_i \text{ が付与されている} \\ 0 & \text{その他の場合} \end{cases}$$

ラベル特徴量の要素  $y_i$  は、ラベル  $w_i$  の有無を表している。三次元物体に付与されているラベルに対応する要素は 1 となり、付与されていないラベルに対応する要素は 0 となる。

## 5. 正準相関分析による形状とラベルとの関連性の学習

正準相関分析 (Canonical Correlation Analy-

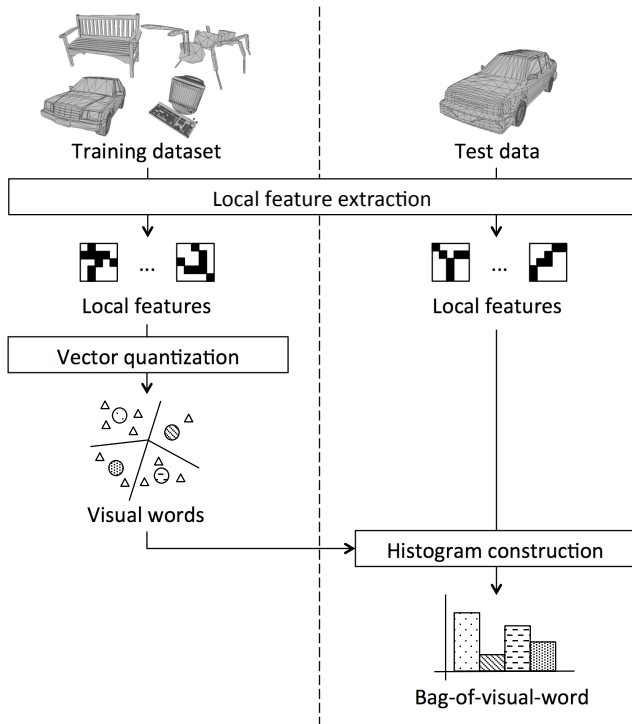


図 2 Bag-of-Visual-Word による特徴抽出処理

Fig. 2 Feature extraction process using Bag-of-Visual-Word model.

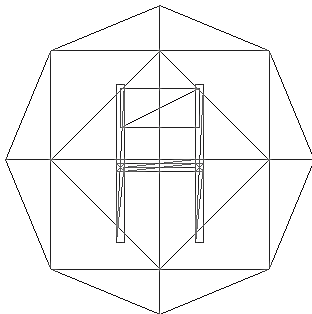


図 3 Geodesic Sphere 内に位置する椅子の三次元物体

Fig. 3 3D chair object in geodesic sphere.

sis, CCA) [13] は、二種類の  $n$  個のサンプル対  $X = \{\mathbf{x}_i\}_{i=1}^n \in \mathbb{R}^{p \times n}$  と  $Y = \{\mathbf{y}_i\}_{i=1}^n \in \mathbb{R}^{q \times n}$  との相関係数が、できるだけ大きくなるような部分空間を求める手法である。本研究では、形状特徴量を  $\mathbf{x}$ 、ラベル特徴量を  $\mathbf{y}$  として正準相関分析を行い、形状とラベルとの関連性を反映した部分空間を学習する。

ここでは、簡単のため、 $X$  と  $Y$  の平均は 0 であり、部分空間の次元数は 1 であると仮定する。  $\mathbf{x}$  と  $\mathbf{y}$  とを、それぞれ、射影ベクトル  $\mathbf{w}_x$  と  $\mathbf{w}_y$  により、部分空間に射影することを考えると、CCA における相関係数の最大化問題は、以下のように表される。

$$J = \arg \max_{\mathbf{w}_x, \mathbf{w}_y} \frac{\mathbf{w}_x^T C_{xy} \mathbf{w}_y}{\sqrt{\mathbf{w}_x^T C_{xx} \mathbf{w}_x \mathbf{w}_y^T C_{yy} \mathbf{w}_y}}$$

ここで、

$$C_{xy} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{y}_i^T$$

$$C_{xx} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T$$

$$C_{yy} = \frac{1}{n} \sum_{i=1}^n \mathbf{y}_i \mathbf{y}_i^T$$

である。

大きさの自由度を制限するために、以下の制約条件を設けると、

$$\mathbf{w}_x^T C_{xx} \mathbf{w}_x = 1$$

$$\mathbf{w}_y^T C_{yy} \mathbf{w}_y = 1$$

CCA の最大化問題は、ラグランジュの未定乗数法から、ラグランジュ関数が、

$$\mathcal{L} = \mathbf{w}_x^T C_{xy} \mathbf{w}_y - \frac{\sigma_x}{2} (\mathbf{w}_x^T C_{xx} \mathbf{w}_x - 1) - \frac{\sigma_y}{2} (\mathbf{w}_y^T C_{yy} \mathbf{w}_y - 1)$$

となるような最適化問題となる。ここで、 $\sigma_x, \sigma_y$  はラグランジュ乗数である。

極値を求めるため、 $\mathcal{L}$  を  $\mathbf{w}_x$  と  $\mathbf{w}_y$  について、それぞれ偏微分して 0 と置くと、以下ようになる。

$$\frac{\partial \mathcal{L}}{\partial \mathbf{w}_x} = C_{xy} \mathbf{w}_y - \sigma_x C_{xx} \mathbf{w}_x = 0 \quad (1)$$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{w}_y} = C_{yx} \mathbf{w}_x - \sigma_y C_{yy} \mathbf{w}_y = 0$$

この二式をまとめて、

$$\sigma_y \mathbf{w}_y^T C_{yy} \mathbf{w}_y - \sigma_x \mathbf{w}_x^T C_{xx} \mathbf{w}_x = 0$$

となる。制約条件から  $\sigma_y - \sigma_x = 0$  であるので、 $\sigma = \sigma_x = \sigma_y$  とおき、 $C_{yy}$  が正則であると仮定すると、

$$\mathbf{w}_y = \frac{1}{\sigma} C_{yy}^{-1} C_{yx} \mathbf{w}_x$$

となる。これを、式 (1) に代入すると、

$$\frac{1}{\sigma} C_{yy}^{-1} C_{yx} \mathbf{w}_x - \sigma C_{xx} \mathbf{w}_x = 0$$

となることから、改めて  $\lambda = \sigma^2$  とおいて式をまとめると、CCA における射影ベクトルを求める問題は、以下の一般固有値問題に帰着する。

$$C_{xy} C_{yy}^{-1} C_{yx} \mathbf{w}_x = \lambda C_{xx} \mathbf{w}_x \quad (2)$$

同様に、 $\mathbf{w}_y$  についても一般化固有値問題に帰着する。

$$C_{yx} C_{xx}^{-1} C_{xy} \mathbf{w}_y = \lambda C_{yy} \mathbf{w}_y$$

$\mathbf{x}$  に対する  $d$  次元の部分空間への射影行列  $W_x \in \mathbb{R}^{p \times d}$  は、式 (2) の一般固有値問題から得られる、 $d$  個の固有値  $\lambda_1 \geq \dots \geq \lambda_d$  に対応する固有ベクトルからなる。ラベルとの関連性を反映した形状特徴量  $\mathbf{z}$  は、射影行列  $W_x$  を用いて、以下のようにして得られる。

$$\mathbf{z} = W_x^T \mathbf{x}$$

## 6. $k$ -近傍識別器によるアノテーション

### 6.1 $k$ -近傍識別器

$k$ -近傍識別器 ( $k$ -Nearest Neighbor Classifier,  $k$ -NN) は、特徴空間上で、テストデータと訓練データとの比較を行い、テストデータの  $k$ -近傍にある訓練データが属するクラスの内、最多数を占めたクラスを、識別結果として出力する分類器である。

いま、テストデータの  $k$ -近傍にある訓練データを  $t_1, \dots, t_k$  とする。さらに、各訓練データは、 $a$  個のクラス  $c_1, \dots, c_a$  に分類されているとすると、 $k$ -NN によるテストデータのクラス  $c$  の推定は、以下のように表される。

$$c = \arg \max_{\theta \in \{c_1, \dots, c_a\}} \sum_{i=1}^k \delta(t_i, \theta)$$

ここで、 $\delta(t_i, \theta)$  は、訓練データ  $t_i$  がクラス  $\theta$  に属している場合に 1、それ以外では 0 となる関数である。

$k$ -NN をアノテーションに用いる場合は、訓練データにクラスではなく複数ラベルを付与しておき、テストデータの  $k$ -近傍にある訓練データのラベルを、アノテーション結果として出力する。 $k$ -NN は、シンプルなアルゴリズムでありながら、画像の自動アノテーションでは優れた性能を示している [4], [11], [21], [30]。

### 6.2 Label Transfer 法

$k$ -NN によるアノテーションでは、 $k$  の値を大きくするに従って、出力ラベル数が増加する。ここで問題となるのが、 $k$  の値を大きくすると、正解ラベル数だけでなく、不正解ラベル数も増加してしまうことである。この問題に対して、Makadia らは、Label Transfer 法を提案した [21]。

Label Transfer 法では、 $b$  個のラベルを、テストデータの  $k$ -近傍にある訓練データから、出現頻度を基準に出力する。いま、テストデータの  $k$ -近傍にある訓練データを  $t_1, \dots, t_k$  とする。ここで、 $k$ -近傍は、テストデータと距離が近いものから順位付けられているとする。さらに、訓練データ  $t_i$  に付与されているラベル数を  $|t_i|$  とすると、Label Transfer 法のアルゴリズムは以下のとおりである。

- (1) 最近傍の訓練データ  $t_1$  に付与されているラベルを、訓練データセット内での出現頻度で順位付けする。
- (2) 上位  $b$  個のラベルをテストデータのラベルとして出力する。ここで、 $|t_1| < b$  であれば (3) に進む。
- (3)  $t_2$  から  $t_k$  に付与されているラベルの内、(2) で出力したラベルと共起しているものを、 $t_1$  を除く  $k$ -近傍での出現頻度で順位付けを行い、上位  $b - |t_1|$  個のラベルを出力する。

## 7. 評価実験

### 7.1 実験環境

提案手法の有効性を確認するために、評価実験を行った。

三次元物体のアノテーションを目的としたデータセットは、これまでに提案されていない。そこで、三次元物体の形状類似検索の評価を目的としたデータセットである Princeton Shape Benchmark (PSB) [26] を用いて新たに作成した。

PSB には、1,814 個の三次元物体が、例えばウサギの三次元物体であれば natural/animal/quadruped/rabbit のように階層的に分類されている。三次元物体の形状類似検索の研究では、最下層の 90 個のクラスが評価に用いられている。これに対して、本研究では、PSB の各階層のクラスを一つのラベルと考え、新たに三次元物体に複数ラベルが付与されたデータセットを作成した。結果として、188 個のラベルからなるデータセットとなった。

訓練データとテストデータの分割には、Leave-one-out 法を用いた。すなわち、PSB を用いた実験であれば、1,813 個の三次元物体を訓練データセットとして、残り 1 個の三次元物体に対してアノテーションを行う。この操作を、全ての三次元物体が、訓練データとテストデータに用いられるよう繰り返す。

評価尺度には、画像のアノテーションの研究にならい、ラベルに対する再現率 (Recall)、適合率 (Precision)、調和平均 (F-Measure)、さらに、正しくアノテーションできたラベルの数 ( $N_+$ ) を用いる。

いま、あるラベル  $w_i$  について、 $w_i$  を正解とする三次元物体数を  $N_r$  とし、そのうち正しく  $w_i$  を付与できた三次元物体数を  $N_w$  とすると、Recall は以下のように定義される。

$$\text{Recall}(w_i) = N_w / N_r$$

また、システムが  $w_i$  を付与した三次元物体数を  $N_p$  とすると、Precision は以下のように定義される。

$$\text{Precision}(w_i) = N_w / N_p$$

実験では、これら Recall と Precision の全ラベルでの平均値を用いる。さらに、Recall と Precision はトレードオフの関係にあるため、総合的な評価尺度として F-Measure を用いる。全ラベルでの Recall の平均値を  $R$ 、Precision の平均値を  $P$  とすると、F-Measure は以下のように定義される。

$$\text{F-Measure} = 2 \cdot R \cdot P / (R + P)$$

本研究で用いる三次元物体の特徴量 DBoV は、Visual Word の数がパラメータとなる。この Visual Word の数は、DBoV の次元数と一致する。実験に用いるデータセットは、いずれも小規模なデータセットであることから、十

分な学習が行えるように、Visual Word の数  $p$  は 500 とした。Visual Word の生成に必要な訓練データセットには、SHREC'11 Generic 3D Warehouse Dataset [5] を用いた。また、DBoV の類似度にはコサイン類似度を用いた。

## 7.2 実験結果

正準相関分析 (Canonical Correlation Analysis, CCA) では、部分空間の次元数がパラメータとなる。そこで、部分空間の次元数を変化させて、調和平均の値が最大となる次元数を確認した。また、形状とラベルとの関連性を学習することの効果を確認するため、主成分分析 (Principal Component Analysis, PCA) [12], [25] による部分空間に射影したものとも比較を行う。

図 4 は、部分空間の次元数に対する、各手法の調和平均の値をプロットしたものである。 $k$ -近傍識別器における近傍数は  $k = 1$  とした。

DBoV は、アノテーションに DBoV を用いたもの、DBoV-PCA と DBoV-CCA は、それぞれ、PCA と CCA による部分空間に射影した DBoV を用いたものである。ここで、CCA では、あらかじめ、PCA を用いて、DBoV を 100 次元に、ラベル特徴量を 140 次元に次元削減している。DBoV とラベル特徴量の次元削減数は、ともに、寄与率が 98% を超える値である。

図 4 を見ると、次元数が 30 以上で、DBoV-CCA は、もとの DBoV よりも調和平均の値が大きくなっている。また、次元数が 60 のときに、調和平均の値が最大となっている。一方、DBoV-PCA は、次元数が大きくなるほど、調和平均の値も大きくなるが、もとの DBoV の値を超えることはない。以上から、CCA により、形状とラベルとの関連性を学習することが、有効であることがわかる。

表 1 は、 $k$ -近傍識別器における近傍数  $k$  の変化に対する、各手法の評価尺度の値をまとめたものである。各手法で、LT は Label Transfer を用いることを示す。LT における出力ラベル数  $l$  は 4 とした。また、PCA と CCA における部分空間の次元数は、PCA は 100 次元、CCA は 60 次元とした。

表 1 を見ると、DBoV では、近傍数が大きくなるにつれ、再現率の値が大きくなっている。しかし、近傍数を大きくすると、不正解となるラベルも多く出力されるため、適合率の値は低下する。一方、DBoV-LT では、Label Transfer の効果により、近傍データのなかで出現頻度の高い上位のラベルのみ出力されるため、近傍数が大きくなっても、適合率の値は低下しない。

DBoV-PCA-LT では、 $k = 1$  の場合に調和平均の値が最大の 55.0% となっているが、DBoV-LT の最大値 56.2% を超えることはない。これに対して、DBoV-CCA-LT では、 $k = 10$  の場合に調和平均の値が最大の 59.8% となり、その他の手法と比較しても、最も優れたラベル推定性能であ

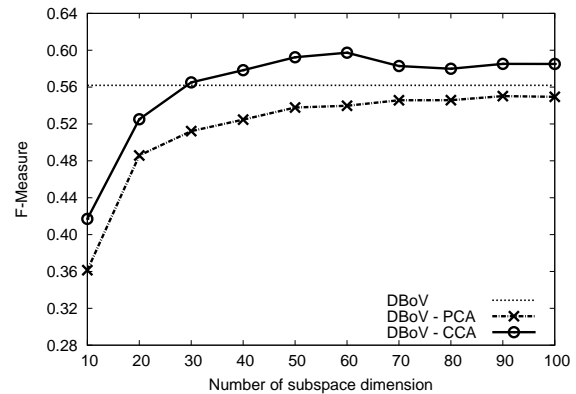


図 4 各手法の部分空間の次元数に対する性能比較

Fig. 4 Comparison of performance to number of subspace dimension of each method.

ることがわかる。

図 5 は、各手法でのアノテーション結果の例をまとめたものである。DBoV における近傍数は  $k = 1$  であり、DBoV-CCA-LT における近傍数は  $k = 10$  である。

m334.off は、人の手の三次元物体であるが、ポリゴン数が少なく直線的な形状となっている。そのため、DBoV では、desktop や display\_device といった機械製品のラベルが推定されている。一方、DBoV-CCA-LT では、CCA によりラベル情報を形状特徴量に反映するため、正しく hand のラベルが推定されている。

m871.off は、長方形の机の三次元物体である。しかし、形状が椅子に類似しているため、DBoV では、seat や dining\_chair といったラベルが推定されている。これに対して、DBoV-CCA-LT では、CCA により形状とラベルとの関連性を学習するため、table や rectangular といったラベルを推定できている。

m1803.off は、スロットマシンの三次元物体であるが、DBoV、DBoV-CCA-LT とともに、誤ったラベルが推定されている。 $k$ -近傍識別器は、訓練データが多いほど、特徴空間上でのデータ分布を正しく捉えることができる。PSB に含まれるスロットマシンの三次元物体は少ないため、正しいラベルを推定できなかったと考える。

m1158.off は、旅客機の三次元物体である。DBoV は正しく commercial のラベルを推定しているが、DBoV-CCA-LT は fighter\_jet のラベルを誤って推定している。CCA による部分空間では、共通のラベルを多く有する三次元物体同士は、近距離に位置する。旅客機と戦闘機の三次元物体には、ともに、manmade, vehicle, winged\_vehicle のラベルが付与されている。このため、DBoV-CCA-LT では、ラベル情報を反映することで、形状特徴量の差異が失われ、誤ったラベルを推定したと考える。

## 8. おわりに

本稿では、三次元物体のアノテーションに関するベース

表 1 各手法の近傍数  $k$  に対する性能比較

Table 1 Comparison of performance to number of the neighborhood  $k$  of each method.

$k$	DBoV				DBoV-LT				DBoV-PCA-LT				DBoV-CCA-LT			
	P%	R%	F%	N+	P%	R%	F%	N+	P%	R%	F%	N+	P%	R%	F%	N+
1	57.5	54.9	56.2	168	57.5	54.9	56.2	168	56.4	53.6	55.0	169	60.0	59.5	59.7	172
2	47.7	63.1	54.3	176	56.3	55.9	56.1	170	54.0	54.4	54.7	170	57.6	60.5	59.0	173
3	40.5	67.5	50.6	180	56.2	55.2	55.7	168	55.1	53.9	54.5	169	58.7	60.1	59.4	172
4	34.7	70.4	46.5	182	56.6	55.1	55.8	168	55.4	53.7	54.5	169	59.0	59.8	59.3	172
5	31.0	72.6	43.4	182	57.1	55.1	56.1	168	56.8	53.7	54.7	169	59.2	59.8	59.5	172
6	28.0	74.5	40.7	183	57.2	55.0	56.1	168	56.0	53.7	54.8	169	59.7	59.7	59.7	172
7	25.7	76.3	38.5	183	57.3	55.0	56.1	168	56.0	53.6	54.8	169	59.7	59.7	59.7	172
8	23.9	77.8	36.5	185	57.3	55.0	56.1	168	56.0	53.6	54.8	169	59.8	59.7	59.7	172
9	22.4	78.9	34.9	186	57.1	55.0	56.0	168	55.9	53.6	54.7	169	59.9	59.7	59.8	172
10	20.9	79.9	33.2	186	57.3	55.0	56.1	168	55.9	53.7	54.8	169	59.9	59.7	59.8	172

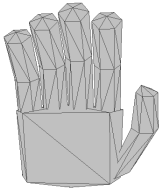
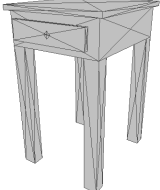
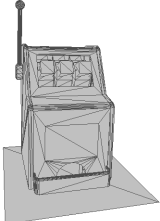
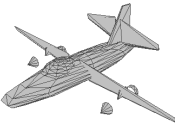
				
	m334.off	m871.off	m1803.off	m1158.off
Ground Truth	natural,hand	manmade,furniture,table rectangular	manmade,slot_machine	manmade,vehicle winged_vehicle,commercial
DBoV	manmade,household, desktop,display_device	manmade,furniture,seat dining_chair	manmade,computer_monitor, household,display_device	manmade,vehicle winged_vehicle,commercial
DBoV-CCA-LT	natural,hand	manmade,furniture,table rectangular	manmade,furniture, seat,desk_with_hutch	manmade,vehicle winged_vehicle,fighter_jet

図 5 各手法で推定されたラベルの例

Fig. 5 Predicted labels of each method for sample 3D objects.

ライン手法の提案と、正準相関分析により形状とラベルとの関連性を学習する手法を提案した。

ベースライン手法では、Bag-of-Visual-Word による形状特徴量と  $k$ -近傍識別器を用いて、特徴空間においてテストデータの  $k$ -近傍にある訓練データのラベルを、テストデータに付与する。さらに、ラベルの有無による二値ベクトルを用いて、正準相関分析により、形状特徴量にラベルによる形状の意味を反映した。Princeton Shape Benchmark を用いた実験では、調和平均の値が、ベースライン手法では 56.2%、正準相関分析を用いた手法では 59.8% となり、正準相関分析により形状とラベルとの関連性を学習することの有効性を確認できた。

今後の課題は、三次元物体のアノテーションを目的とした、大規模データセットの作成と実験である。本稿の実験で使ったデータセットは、含まれる三次元物体が 2,000 個程度の小規模なものである。また、三次元物体に付与されているラベルの数も少ない。そのため、実用化を想定した評価には、形状やラベルの種類が十分でない。これに対しては、Google 3D Warehouse [10] など、三次元物体に

タイトルやコメントが付与されたりポジトリを利用して、データセットを作成することが考えられる。

**謝辞** 本研究の実施にあたり、科研基盤 (C) 23500119 と A-step 課題番号 AS242Z01713H の助成を受けました。記して感謝致します。

#### 参考文献

- [1] Bronstein, A. M., Bronstein, M. M., Guibas, L. J. and Ovsjanikov, M.: Shape google: Geometric words and expressions for invariant shape retrieval, *ACM Transactions on Graphics*, Vol. 30, No. 1, pp. 1–20 (2011).
- [2] Chen, D.-Y., Tian, X.-P., Shen, Y.-T. and Ouhyoung, M.: On Visual Similarity Based 3D Model Retrieval, *Computer Graphics Forum*, Vol. 22, No. 3, pp. 223–232 (2003).
- [3] Csurka, G., Dance, C. R., Fan, L., Willamowski, J. and Bray, C.: Visual categorization with bags of keypoints, *Proceedings of the Workshop on Statistical Learning in Computer Vision, ECCV '04*, pp. 1–22 (2004).
- [4] Cui, C., Ma, J., Lian, T., Wang, X. and Ren, Z.: Ranking-oriented nearest-neighbor based method for automatic image annotation, *Proceedings of the 36th international ACM SIGIR conference on Research and*

- development in information retrieval*, SIGIR '13, New York, NY, USA, ACM, pp. 957–960 (2013).
- [5] Dutagaci, H., Godil, A., Daras, P., Axenopoulos, A., Litos, G. C., Manolopoulou, S., Goto, K., Yanagimachi, T., Kurita, Y., Kawamura, S., Furuya, T. and Ohbuchi, R.: SHREC '11 Track: Generic Shape Retrieval, *Proceedings of Eurographics Workshop on 3D Object Retrieval 2011*, pp. 65–69 (2011).
- [6] Eitz, M., Hildebrand, K., Boubekur, T. and Alexa, M.: Sketch-based 3D shape retrieval, *ACM SIGGRAPH 2010 Talks*, SIGGRAPH '10, New York, NY, USA, ACM, pp. 5:1–5:1 (2010).
- [7] Funkhouser, T., Kazhdan, M., Shilane, P., Min, P., Kiefer, W., Tal, A., Rusinkiewicz, S. and Dobkin, D.: Modeling by example, *ACM SIGGRAPH 2004 Papers*, SIGGRAPH '04, New York, NY, USA, ACM, pp. 652–663 (2004).
- [8] Furuya, T. and Ohbuchi, R.: Dense sampling and fast encoding for 3D model retrieval using bag-of-visual features, *Proceedings of the ACM International Conference on Image and Video Retrieval*, CIVR '09, No. 26, ACM, pp. 1–8 (2009).
- [9] Goldfeder, C. and Allen, P.: Autotagging to improve text search for 3d models, *Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries*, JCDL '08, New York, NY, USA, ACM, pp. 355–358 (2008).
- [10] Google: Google 3D Warehouse, Google (online), available from (<http://sketchup.google.com/3dwarehouse/>) (accessed 2013-08-15).
- [11] Guillaumin, M., Mensink, T., Verbeek, J. and Schmid, C.: TagProp: Discriminative metric learning in nearest neighbor models for image auto-annotation, *International Conference on Computer Vision*, ICCV '09, pp. 309–316 (2009).
- [12] Hotelling, H.: Analysis of a complex of statistical variables into principal components, *Journal of Educational Psychology*, Vol. 24, pp. 417–441 (1933).
- [13] Hotelling, H.: Relations Between Two Sets of Variates, *Biometrika*, Vol. 28, No. 3-4, pp. 321–377 (1936).
- [14] Iyer, N., Jayanti, S., Lou, K., Kalyanaraman, Y. and Ramani, K.: Three-dimensional shape searching: state-of-the-art review and future trends, *Computer-Aided Design*, Vol. 37, No. 5, pp. 509–530 (2005).
- [15] Kanai, S.: Content-based 3D mesh model retrieval from hand-written sketch, *International Journal on Interactive Design and Manufacturing*, Vol. 2, No. 2, pp. 87–98 (2008).
- [16] Khronos Group: OpenGL ES - The Standard for Embedded Accelerated 3D Graphics, Khronos Group (online), available from (<http://www.khronos.org/opngles/>) (accessed 2013-08-15).
- [17] Khronos Group: WebGL - OpenGL ES 2.0 for the Web, Khronos Group (online), available from (<http://www.khronos.org/webgl/>) (accessed 2013-08-15).
- [18] Lavoué, G.: Bag of words and local spectral descriptor for 3D partial shape retrieval, *Proceedings of the Eurographics Workshop on 3D Object Retrieval*, 3DOR '11, Aire-la-Ville, Switzerland, Switzerland, Eurographics Association, pp. 41–48 (2011).
- [19] Lian, Z., Godil, A. and Sun, X.: Visual Similarity Based 3D Shape Retrieval Using Bag-of-Features, *Proceedings of the 2010 Shape Modeling International Conference*, Washington, DC, USA, IEEE Computer Society, pp. 25–36 (2010).
- [20] Lowe, D. G.: Distinctive Image Features from Scale-Invariant Keypoints, *International Journal Computer Vision*, Vol. 60, No. 2, pp. 91–110 (2004).
- [21] Makadia, A., Pavlovic, V. and Kumar, S.: A New Baseline for Image Annotation, *Proceedings of the 10th European Conference on Computer Vision*, ECCV '08, No. 3, pp. 316–329 (2008).
- [22] Min, P., Kazhdan, M. and Funkhouser, T.: A Comparison of Text and Shape Matching for Retrieval of Online 3D Models, *European Conference on Digital Libraries*, pp. 209–220 (2004).
- [23] Novotni, M. and Klein, R.: 3D zernike descriptors for content based shape retrieval, *Proceedings of the eighth ACM symposium on Solid modeling and applications*, pp. 216–225 (2003).
- [24] Ohbuchi, R. and Kawamura, S.: Shape-Based Auto-tagging of 3D Models for Retrieval, *Proceedings of the 4th International Conference on Semantic and Digital Media Technologies: Semantic Multimedia*, SAMT '09, Berlin, Heidelberg, Springer-Verlag, pp. 137–148 (2009).
- [25] Pearson, K.: On lines and planes of closest fit to systems of points in space, *Philosophical Magazine*, Vol. 5, pp. 559–572 (1901).
- [26] Shilane, P., Min, P., Kazhdan, M. and Funkhouser, T.: The Princeton Shape Benchmark, *Proceedings of the Shape Modeling International 2004*, SMI '04, Washington, DC, USA, IEEE Computer Society, pp. 167–178 (2004).
- [27] Sivic, J. and Zisserman, A.: Video Google: A Text Retrieval Approach to Object Matching in Videos, *Proceedings of the Ninth IEEE International Conference on Computer Vision*, ICCV '03, Vol. 2, Washington, DC, USA, IEEE Computer Society, pp. 1470–1477 (2003).
- [28] Tangelder, J. W. and Veltkamp, R. C.: A survey of content based 3D shape retrieval methods, *Multimedia Tools Applications*, Vol. 39, No. 3, pp. 441–471 (2008).
- [29] Tatsuma, A. and Aono, M.: Multi-Fourier spectra descriptor and augmentation with spectral clustering for 3D shape retrieval, *The Visual Computer*, Vol. 25, No. 8, pp. 785–804 (2009).
- [30] Torralba, A., Fergus, R. and Freeman, W. T.: 80 Million Tiny Images: A Large Data Set for Nonparametric Object and Scene Recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 30, No. 11, pp. 1958–1970 (2008).
- [31] Uijlings, J. R., Smeulders, A. W. and Scha, R. J.: Real-Time Visual Concept Classification, *IEEE Transactions on Multimedia*, Vol. 12, No. 7, pp. 665–681 (2010).
- [32] Wahl, E., Hillenbrand, U. and Hirzinger, G.: Surflet-Pair-Relation Histograms: A Statistical 3D-Shape Representation for Rapid Classification, *Proceedings of International Conference on 3D Digital Imaging and Modeling*, pp. 474–482 (2003).
- [33] Zhang, C. and Chen, T.: An Active Learning Framework for Content-Based Information Retrieval, *IEEE TRANSACTIONS ON MULTIMEDIA*, Vol. 4, No. 2, pp. 260–268 (2002).
- [34] Zhou, D., Weston, J., Gretton, A., Bousquet, O. and Schölkopf, B.: Ranking on Data Manifolds, *Advances in Neural Information Processing Systems 16* (Thrun, S., Saul, L. and Schölkopf, B., eds.), MIT Press, Cambridge, MA (2004).