

距離空間ピラミッドを用いた LLC による 3次元物体認識

堀 貴博[†] 滝口 哲也^{††} 有木 康雄^{††}

[†] 神戸大学大学院システム情報学研究科 〒 657-8501 兵庫県神戸市灘区六甲台町 1-1

^{††} 神戸大学自然科学系先端融合研究環 〒 657-8501 兵庫県神戸市灘区六甲台町 1-1

E-mail: [†]horitaka@me.cs.scitec.kobe-u.ac.jp, ^{††}{takigu,ariki}@kobe-u.ac.jp

あらまし 近年、高精度な RGB-D カメラの登場により、高品質な 3次元情報（色、奥行き情報）が容易に測定可能となった。これを用いた従来の物体認識手法は、奥行き情報を局所の特徴の抽出にしか使用していない。つまり、奥行き情報を取得することで物体の全体的な形状を把握することが可能であるにも関わらず、部分的な利用に留まっている。そこで、提案手法では、奥行き情報に基づく距離空間ピラミッドによって、全体的な物体形状を表現する手法を提案する。具体的には、距離空間ピラミッドでの特徴点の座標位置によって、奥行きの位相情報を含んだ特徴表現を実現する。また、距離画像から抽出する 3次元局所特徴量として、HONV (Histogram of Oriented Normal Vectors) を用い、特徴量のコード化には、特徴空間座標系での近傍制限を利用した LLC (Locality-constrained Linear Coding) を適用した。3次元情報を含む画像データセットを用いた評価実験により、提案手法は従来手法と比較して認識精度が向上することを示す。

キーワード 3次元物体認識, HONV, LLC, 空間ピラミッド, 距離空間ピラミッド

RGB-D Based 3D-Object Recognition by LLC Using Depth Spatial Pyramid

Takahiro HORI[†], Tetsuya TAKIGUCHI^{††}, and Yasuo ARIKI^{††}

[†] Graduate School of System Informatics, Kobe University 1-1 Rokkodai, Nada-ku, Kobe, 657-8501 Japan

^{††} Organization of Advanced Science and Technology, Kobe University 1-1 Rokkodai, nada, Kobe 657-8501, Japan

E-mail: [†]horitaka@me.cs.scitec.kobe-u.ac.jp, ^{††}{takigu,ariki}@kobe-u.ac.jp

Abstract Recently introduced high-accuracy RGB-D cameras are capable of providing high quality three dimension information (color and depth information) easily. The overall shape of the object can be understood by acquiring depth information. However, conventional methods adopted this camera uses depth information only to extract the local feature. To deal with this problem, in the proposed method, the overall object shape is expressed by the Depth Spatial Pyramid based on depth information. In more detail, multiple features within each sub-region of the Depth Spatial Pyramid are pooled. As a result, the feature representation including the depth topological information is constructed. We use HONV (Histogram of Oriented Normal Vectors) designed to capture local geometric characteristics as 3D local features and apply LLC (Locality-constrained Linear Coding) utilized the locality constraints to project each descriptor into its local-coordinate system. As a result of image dataset, the proposed method has improved the recognition rate compared with conventional methods.

Key words 3D-Object Recognition, HONV, LLC, spatial pyramid, depth spatial pyramid

1. ま え が き

物体認識とは、制約のない実世界シーンを撮影した画像に含まれる物体を、計算機がその名称で認識することを指し、コンピュータビジョンの分野において困難な研究課題の一つである。

一般的な名称であるカテゴリ名を認識する一般物体認識と、各物体ごとに認識する特定物体認識があり、機械による人間の高度視覚機能の実現という観点から、ロボットビジョンへ応用される。近年、そのロボット工学分野において、高品質な 3次元情報（色、奥行き情報）を測定する RGB-D カメラが用いられ

るようになった。また、ロボットに限らず、今後3次元センサやステレオカメラの普及に伴い、3次元情報を持つデジタル画像が増加すると考えられる。そこで計算機の3次元物体認識による画像の検索や分類が求められている。3次元物体認識における入力情報は図1に示されるような、物体の色、奥行き情報である。

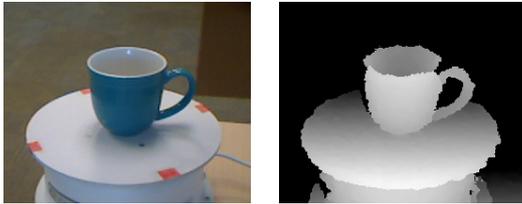


図1 入力情報
Fig.1 Input Information

最先端の2次元画像分類システムは大きく分けて3つの工程で構成される。SIFT(Scale-Invariant Feature Transform)による特徴抽出[1]、BoF(Bag of Features)によるコード化[2]、SPM(Spatial Pyramid Matching)によるプーリング処理である[3]。特徴抽出では、多くの局所特徴量が提案されているが、SIFTが最も良く用いられている。

BoFは図2のように、局所特徴の出現頻度ヒストグラムとして画像を表現する。これは、特徴の空間的な変形に対して特に頑健であり、画像分類タスクにおいて良い性能を示す。しかし、それゆえにBoFでは特徴の空間的な配置や関係性が失われ、物体の形状や位置の情報を得ることができない。

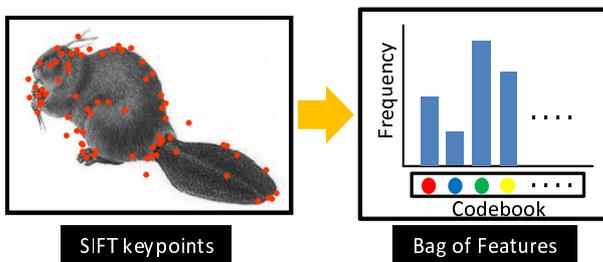


図2 Bag of Features
Fig.2 Bag of Features

そこで、その拡張としてSPMを使用する。SPMは画像を段階的に細かい部分領域に区切り、その各部分領域から局所特徴のヒストグラムを計算する。具体的には、図3のように、 $2^l \times 2^l$ の部分領域、 $l = 0, 1, 2$ が用いられる。この位置的な制約を与える空間ピラミッドは、多くの画像分類タスクに関して良い性能を示している。これらの2次元画像に用いられていた手法が3次元物体認識に適用されている。しかし、このため従来手法は、奥行き情報を基に作成した距離画像上で特徴抽出を行うといった形でしか、3次元情報を使用していない。

本研究ではこの問題に対処するために、奥行き情報に基づく距離空間に制約を与えて、全体的な物体形状を表現する手法を提案する。具体的には、2次元画像を階層的に分割する空間ピラミッドを3次元拡張した、距離空間ピラミッドを用いる。ま

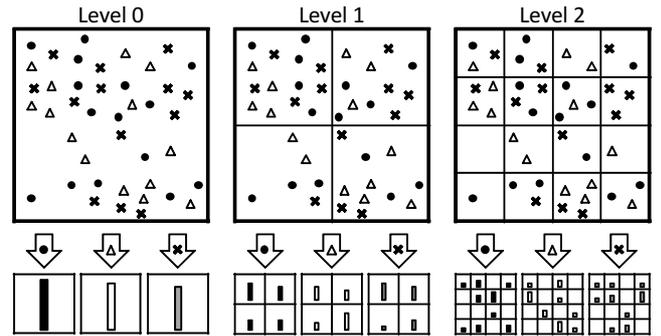


図3 空間ピラミッドマッチング
Fig.3 Spatial Pyramid Matching

た、距離画像から抽出する3次元局所特徴量として、局所的な3次元情報の幾何学特性を表すHONV(Histogram of Oriented Normal Vectors)を用い、特徴量のコード化には、特徴空間座標系での近傍制限を利用したLLC(Locality-constrained Linear Coding)を適用した。

本論文の構成は以下の通りである。2章で、提案手法の流れについて述べ、3章でHONV特徴量について述べる。4章ではコード化処理としてLLCを適用し、5章でプーリング処理である距離空間ピラミッドを提案する。6章では、提案手法の有効性を画像データセットで評価した結果について述べ、7章で、本論文のまとめと今後の方針について述べる。

2. 提案手法の流れ

提案手法の流れを図4に示す。まず、奥行き情報から距離画像、色情報からRGBの各チャンネルの画像(カラー画像)を作成する。そして、各画像上に等間隔で格子状に点を配置し、このグリッド点を特徴点とする。その特徴点に関して、距離画像ではHONV特徴、カラー画像ではSIFT特徴を抽出する。画像ごとに、抽出された各特徴量に対してLLCによるコード化処理を行う。次に、得られた特徴を空間ピラミッドの部分領域ごとにプーリング処理で統合する。この際、距離画像に対しては距離空間ピラミッドによるプーリング処理も行う。こうして得られた各画像の特徴を並べて一つの特徴ベクトルとする。その特徴ベクトルを用いて学習データより分類器を学習する。テストデータに対しては、学習した分類器を用いてクラス分類して認識結果を出力する。

次に、提案手法の各処理について、それぞれ詳細に述べる。

3. HONV(Histogram of Oriented Normal Vectors)

HONVは、奥行きセンサによる物体認識を目的として、局所的な3次元情報の幾何学特性を表すために設計された特徴量である[4]。具体的には、物体の表面点における法線ベクトル方向の局所分布として、局所的3次元形状の特性を表現する。

奥行きセンサによって得られた立体情報は、距離画像に変換される。この距離画像の各画素を $p = (x, y, d(x, y))$ と表す。この $d(x, y)$ が画素位置とセンサの間の距離であり、

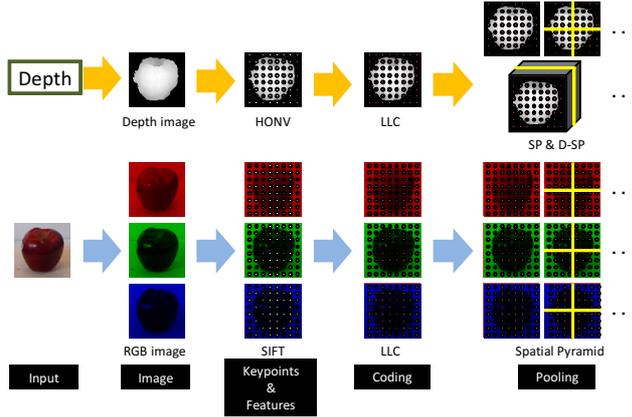


図 4 提案手法の流れ
Fig. 4 System overview

奥行き情報となる．この各点 p における法線ベクトルは $\mathbf{N} = (-\frac{\partial d(x,y)}{\partial x}, -\frac{\partial d(x,y)}{\partial y}, 1)$ となる． $\frac{\partial d(x,y)}{\partial x}$ と $\frac{\partial d(x,y)}{\partial y}$ は、それぞれ差分近似を使用することで、計算される．法線ベクトル方向として、球面座標系における法線ベクトルの天頂角 θ と方位角 φ を用いる．法線ベクトルの 3 次元目が定数であるので、2つの角度で法線ベクトルを十分に説明できる．図 5 に法線ベクトルと天頂角 θ 、方位角 φ との関係を示す．それぞれ以下のように計算される．

$$\theta = \tan^{-1} \left(\left(\frac{\partial d(x,y)}{\partial x} \right)^2 + \left(\frac{\partial d(x,y)}{\partial y} \right)^2 \right)^{\frac{1}{2}} \quad (1)$$

$$\varphi = \tan^{-1} \left(\frac{\partial d(x,y)}{\partial y} / \frac{\partial d(x,y)}{\partial x} \right) \quad (2)$$

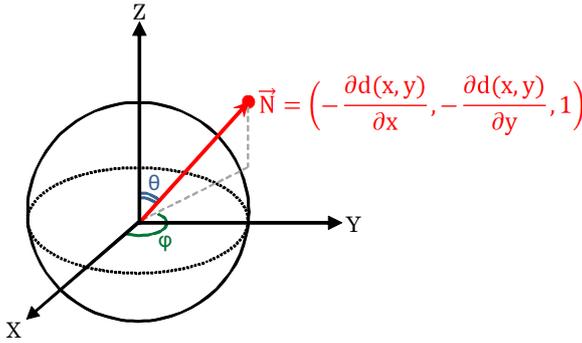


図 5 法線ベクトルの天頂角 θ と方位角 φ
Fig. 5 Zenith angle θ and azimuthal angle φ of a normal vector

この 2 つの角度に対して、HOG (Histograms of Oriented Gradients) などのヒストグラムベースの特徴と同様の処理を行う [5]．まず、各特徴点を中心とするセルによって、画像を分割する．その各セルにおいて、各画素の法線ベクトル方向 θ と φ を 2 次元ヒストグラムに量子化、投票する． I 個のビンと J 個のビンがそれぞれ θ と φ に使用されると仮定すると、この処理により各セルから $I \times J$ 次元特徴ベクトルが作成される．空間的エッジの影響を抑制するために、隣接するセルヒストグラム上でガウシアンフィルタによる平滑化を行う．各セルのヒ

ストグラムをベクトルとし、L2 ノルムで割って正規化を行う．この各セルごとに作成された特徴ベクトルを対応する特徴点の HONV 特徴量とする．

4. コード化処理

コード化とは、特徴量を性能の高い別の特徴量 (コード) に変換する処理である．本論文では、コード化処理に LLC を適用する [6]．画像から抽出された N 個の D 次元局所特徴量 \mathbf{x}_i の集合を $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \in \mathbb{R}^{D \times N}$ とする．また、全学習画像の特徴量を k-means 法 [7] でクラスタリングして作成した、 M 個のコードワード \mathbf{b}_j の集合であるコードブックを $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_M] \in \mathbb{R}^{D \times M}$ とする．これにより、各特徴量は最終的に $[0, 1]$ を要素にもつ M 次元のコードに変換される．図 6 に示すように、2 つの既存のコード化処理と比較して、LLC を説明する．

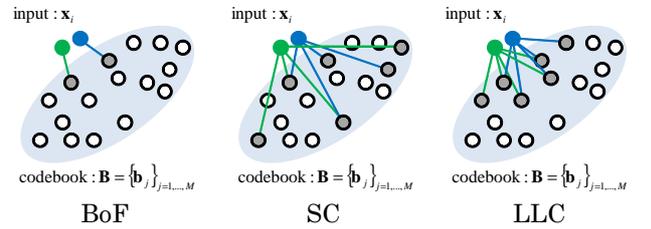


図 6 BoF, SC, LLC の比較
Fig. 6 Comparison between BoF, SC and LLC

4.1 BoF (Bag of Features)

BoF の出現頻度ヒストグラムを求める内容は、以下の制約付き最小二乗問題を解くことと同義である．

$$\arg \min_C \sum_{i=1}^N \|\mathbf{x}_i - \mathbf{B}\mathbf{c}_i\|^2 \quad (3)$$

$$s.t. \|\mathbf{c}_i\|_{l_0} = 1, \|\mathbf{c}_i\|_{l_1} = 1, \mathbf{c}_i \geq 0, \forall i$$

ここで、 $\mathbf{C} = [\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_N]$ は \mathbf{X} におけるコードの集合である．定数制約 $\|\mathbf{c}_i\|_{l_0} = 1$ は各 \mathbf{c}_i に 0 でない要素が 1 つのみあることを示す．これは、 \mathbf{x}_i の量子化インデックスに対応している． $\|\mathbf{c}_i\|_{l_1} = 1, \mathbf{c}_i \geq 0$ は、 \mathbf{x} へのコード重みが 1 であることを意味する．実際には、ただ一つの 0 でない要素は最近傍探索によって見つけることができる．

4.2 SC (Sparse coding)

図 6 に示されるように、BoF は特徴量が単一のコードワードにしか所属できないので、量子化による情報損失が起こる．その損失を改善するために、式 (3) の定数制約 $\|\mathbf{c}_i\|_{l_0} = 1$ の代わりに、スパースな L1 ノルム正則化項を用いる [8]．この緩和により、少数のコードワードへの所属が許容される．その結果、各局所特徴量 \mathbf{x}_i をコード化する処理は、以下のように標準の Sparse coding 問題となる [9]．

$$\arg \min_C \sum_{i=1}^N \|\mathbf{x}_i - \mathbf{B}\mathbf{c}_i\|^2 + \lambda \|\mathbf{c}_i\|_{l_1} \quad (4)$$

このスパース正則化項の導入により、量子化誤差を大幅に減

小さくすることができる。

4.3 LLC (Locality-constrained Linear Coding)

LCC (Local Coordinate Coding) [10] によって示されるように、位置情報はスパース性よりも有用である。よって、LLC では式 (4) でのスパース制約の代わりに位置制約を取り入れる。つまり、入力特徴量をその近傍のコードワードで表現する。LLC によるコード化は以下の式を解くことで行われる。

$$\arg \min_C \sum_{i=1}^N \|\mathbf{x}_i - \mathbf{B}\mathbf{c}_i\|^2 + \lambda \|\mathbf{d}_i \odot \mathbf{c}_i\|^2 \quad (5)$$

$$\text{s.t. } \mathbf{1}^\top \mathbf{c}_i = 1, \forall i$$

ここで、 \odot は要素同士の乗算を示す。 $\mathbf{d}_i \in \mathbb{R}^M$ は、以下の式で表される位置情報アダプタである。

$$\mathbf{d}_i = \exp\left(\frac{\text{dist}(\mathbf{x}_i, \mathbf{B})}{\sigma}\right) \quad (6)$$

ここで、 $\text{dist}(\mathbf{x}_i, \mathbf{B}) = [\text{dist}(\mathbf{x}_i, \mathbf{b}_1), \dots, \text{dist}(\mathbf{x}_i, \mathbf{b}_M)]^\top$ であり、 $\text{dist}(\mathbf{x}_i, \mathbf{b}_j)$ は \mathbf{x}_i と \mathbf{b}_j 間のユークリッド距離である。 σ は、位置情報アダプタの減衰速度を調節する重みとして用いられる。 $\text{dist}(\mathbf{x}_i, \mathbf{B})$ の最大値を用いることで、 \mathbf{d}_i を $(0, 1]$ に正規化する。

良いクラス分類性能を達成するために、コード化処理は類似する特徴量において同様のコードを発生させるべきである。この要件において、式 (5) の位置情報正規化項は優れた特性を示す。図 6 に示されるように、SC では、スパース性のために、類似する特徴が異なったコードワードを選択する可能性がある。その結果、コード間の相関関係は失われる。一方、LLC での明確な位置情報アダプタは、類似したパッチが同様のコードを持つように処理する。LLC コードがコードワードを共有することによって、類似する特徴量の間相関関係を得ることができる。実際の実験では、式 (5) の \mathbf{B} 代わりに、 \mathbf{x}_i の K 近傍 ($K < D < M$) コードワードの集合 \mathbf{B}_i を用いる近似 LLC を使用する。これにより、 $O(M^2)$ から $O(M + K^2)$ に計算の複雑性を軽減できる。

5. プーリング処理

プーリング処理とは、画像を部分領域に分割し、その領域ごとに特徴量を統合する処理である。本論文では、各画像の 2 次元上の領域分割を空間ピラミッドで行う。距離画像においては、それに加えて、奥行き情報による領域分割を距離空間ピラミッドによって行う。空間ピラミッドについては、前述したように、図 3 に示す $2^l \times 2^l$ の部分領域に分割する。

5.1 距離空間ピラミッド

距離空間ピラミッドとは、奥行き情報から作成した距離空間座標における空間ピラミッドである。つまり、奥行き値を座標として距離空間を作成し、部分領域に分割する。ただし、通常の空間ピラミッドの座標と異なり、奥行き値は測定値なので、値にばらつきがある。そのため、空間ピラミッドのように空間を単純に等分すると、各部分領域に属する特徴点の個数に偏りが生じる。よって、座標による等分割ではなく、各部分空間

の特徴点の個数が等しくなるように分割を行う。具体的には、図 7 のように、 m の部分領域、 $m = 0, 1, 2$ が用いられる。この距離空間ピラミッドと通常の空間ピラミッドを併用することで、3 次元空間における空間制約を与える。それにより、局所特徴量のみでは表現できない、物体の全体的な 3 次元形状を表現することができる。

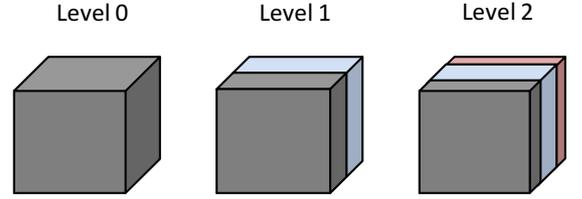


図 7 距離空間ピラミッド

Fig. 7 Depth Spatial Pyramid

5.2 プーリング手法

空間ピラミッドの各層で、各部分領域の特徴量 (LLC コード) を統合する。各部分領域で統合された特徴は、最終的な画像特徴表現として連結され、正規化される。プーリングには、以下の式で表される最大プーリングを用いる。

$$\mathbf{c}_{out1} = \max(\mathbf{c}_{in1}, \dots, \mathbf{c}_{inH}) \quad (7)$$

ここで、 H は部分領域に含まれる特徴点の個数である。この \max 関数は、列で機能し、 \mathbf{c}_{in1} と同じサイズのベクトルを返す。これらの各領域ごとに統合された特徴 \mathbf{c}_{out1} を、画像ごとに連結したベクトル \mathbf{c}_{in} は以下の式で正規化される。

$$\mathbf{c}_{out2} = \mathbf{c}_{in} / \|\mathbf{c}_{in}\|_2 \quad (8)$$

この \mathbf{c}_{out2} が、最終的な画像特徴表現となる。ただし、距離画像については、空間ピラミッドと距離空間ピラミッドで別々に正規化を行い、その結果のベクトルを結合する。

6. 評価実験

本章では、ベンチマークデータセットを用いた物体の分類実験を行い、提案手法と従来手法を比較することで、その有効性を評価する。

6.1 RGB-D オブジェクトデータセット

評価実験には、RGB-D オブジェクトデータセットを用いる [17]。このデータセットは、300 物体、51 カテゴリの奥行き情報と色情報を持ち、奥行き情報を用いて背景と対象物体の分離が行われている。撮影は、各物体をターンテーブルに乗せ、それを 1 メートル離れた位置から、通常のカメラと奥行き情報を観測するのに用いる RGB-D カメラにより、3 つの高さから行われた。1 メートル離すのは、RGB-D カメラが奥行き情報を正確に得るために、最低限必要な距離だからである。つまり、各物体につき、表面を一周撮影した 3 つの動画像が用意されている。実験では、この動画から 5 フレームごとに画像を取り出し、全体で約 42000 枚の RGB-D 画像を用いる。各カテゴリの物体の例を図 8 に示す。このデータセットの詳細は、[11] で紹



図 8 RGB-D オブジェクトデータセットの物体
Fig. 8 Objects from the RGB-D Object Dataset

介されている。

6.2 実験条件

本論文では、RGB-D オブジェクトデータセットを使用して、従来手法と同様の方法で、一般物体認識と特定物体認識の実験を行う。一般物体認識は物体のカテゴリを認識し、特定物体認識は個別に物体（インスタンス）を認識する。まず、一般物体認識は、51 クラスの分類実験である。各カテゴリにつき 1 物体を選択し、その物体の画像をテスト画像、それ以外を学習画像とする。1 物体の選択はランダムで行い、10 回の実験の平均認識率と標準偏差を最終的な実験結果とする。学習画像が約 35000 枚、テスト画像が約 7000 枚となる。次に、特定物体認識は、300 物体の分類実験である。3 つの撮影角度のうち、中間の角度の画像をテスト画像、それ以外を学習画像とする。学習画像が約 28000 枚、テスト画像が約 14000 枚となる。一般物体認識と特定物体認識の両方について、使用する情報を変えて実験を行う。カラー情報のみ (RGB)、奥行き情報のみ (Depth)、両方使用 (RGB-D) の 3 種類であり、それぞれ実験結果を従来手法と比較する。

6.3 実験概要

実験内容について、一般物体認識と特定物体認識の違いはない。実験は、まず 2 章で説明したように、全てのカラー画像、距離画像にグリッド点を配置し、SIFT 特徴量と HONV 特徴量を抽出する。この際、グリッド点は 4 ピクセル間隔で配置し、SIFT のスケール値と HONV のセルの大きさは、8,12,16 ピクセルのマルチスケールを用いる。次に、5 章で述べたように、それぞれの特徴量を LLC を用いてコード化する。コードブックのサイズは 1024 とし、特徴空間座標系での 200 近傍のコードワードを用いて線形システムを解く。そして、空間ピラミッドと距離空間ピラミッドを用いてプーリング処理を行う。特徴ベクトルの次元を揃えるために、カラー画像のみの場合は空間ピラミッド 2 階層、奥行き情報のみの場合は空間ピラミッド 3 階層と距離空間ピラミッド 3 階層、両方の場合は空間ピラミッド 2 階層と距離画像のみそれに加えて距離空間ピラミッド 3 階層とする。ただし、距離空間ピラミッドの 1 階層目は、空間ピラミッドの 1 階層目と同一なので用いない。最後に、画像の各部分領域で統合された特徴ベクトルを全て結合し、物体を表現する特徴ベクトルとして、識別器で学習、認識を行う。本実験で

は、作成されたベクトルをクラス分類するための識別器として SVM (Support Vector Machine) [12] を用いる。SVM は基本的には 2 クラスの識別にしか対応していないが、複数の SVM を組み合わせることでマルチクラスに対応できる。今回は、マルチクラス SVM (線形カーネル linear) を用いる。また、上記の実験に加えて、奥行き情報のみを用いる場合において、距離空間ピラミッドを使用しない実験を行う。これにより、距離空間ピラミッドの有効性を確かめる。

6.4 実験結果と考察

まず、奥行き情報のみを用いる場合における比較実験結果を表 1 に示す。空間ピラミッドのみの場合を SP、距離空間ピラミッドを追加で用いる場合を D-SP とする。

表 1 奥行き情報のみを用いた場合の認識結果 (%)
Table 1 Recognition result by depth information(%)

	Category	Instance
HONV + SP	79.5 ± 1.3	46.6
HONV + D-SP	80.4 ± 1.5	47.1

全ての実験に関する認識率と、従来手法との比較を表 2 に示す。

表 2 認識結果と従来手法との比較 (%)
Table 2 Recognition result and Comparisons with the baseline(%)

	Category			Instance		
	RGB	Depth	RGB-D	RGB	Depth	RGB-D
ICRA11 [11]	74.3 ± 3.3	53.1 ± 1.7	81.9 ± 2.8	59.3	32.3	73.9
Kernel desc [13]	80.7 ± 2.1	80.3 ± 2.9	86.5 ± 2.1	90.8	54.7	91.2
CKM desc [14]	N/A	N/A	86.4 ± 2.3	82.9	N/A	90.4
HMP [15]	74.7 ± 2.5	70.3 ± 2.2	82.1 ± 3.3	75.8	39.8	78.9
ISER12 [16]	82.4 ± 3.1	81.2 ± 2.3	87.5 ± 2.9	92.1	51.7	92.8
Proposed	83.1 ± 3.4	80.4 ± 1.6	88.6 ± 1.8	93.4	47.1	94.2

色情報と奥行き情報の両方を用いた一般物体認識の結果のコンフュージョンマトリックスを図 9 に示す。

表 1 より、提案した距離空間ピラミッドを導入することで、認識率が向上した。これより、奥行き情報を用いて物体の全体的な形状を表現する、距離空間ピラミッドの有効性を示すことができた。また、表 2 より、HONV と LLC を適用し、距離空間ピラミッドを用いた提案手法を用いることで、色情報のみの

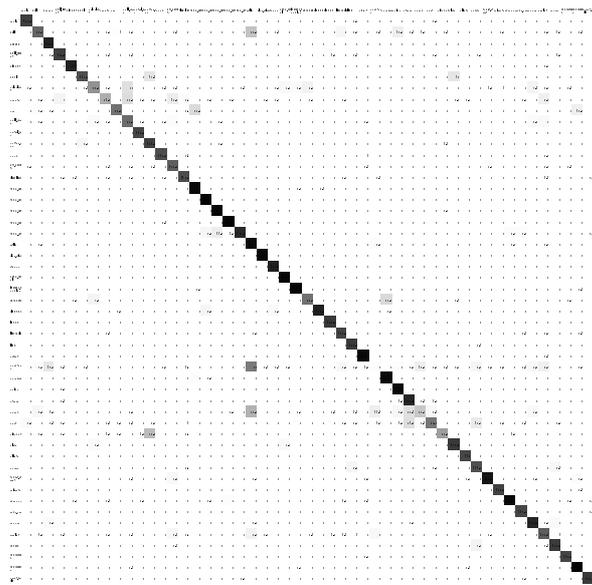


図9 コンフュージョンマトリックス
Fig.9 Confusion matrix

場合と、色情報と奥行き情報の両方の場合において、従来手法よりも認識率が向上した。奥行き情報のみを用いた場合も、他の手法と比較して、最良ではないが高い認識率を示している。よって、3次元物体認識に対するHONVとLLCの適用の有効性が示された。

このように、従来手法よりも認識率が向上した理由について、以下に考察する。本研究で提案した距離空間ピラミッドは、通常の空間ピラミッドと組み合わせることで、物体を3次元的部分空間で区切り、その全体的な形状を各部分空間の局所特徴量の分布として表現する。これは、距離画像から抽出する局所的な表面形状の特徴量だけでは表せない情報である。この奥行き情報から得られた新たな全体形状情報が、カラー画像や距離画像から得られる情報を補い、物体認識の認識率を向上させたと考えられる。

最後に、提案手法で認識に失敗したものについて考察する。ただし、提案手法は実験において、一般物体認識では認識率約88%、特定物体認識では認識率約94%を達成しているの、大半の物体は認識できている。図9の一般物体認識において、ほぼ認識できなかったカテゴリは、マッシュルームクラスである。このカテゴリには3つしか物体がなく、それぞれ形状が異なる。また、物体が小さいので奥行き情報が荒くなり、適切な特徴が取れない。これらの要因により、認識できなかったと考えられる。また、特定物体認識において、認識率の低かったカテゴリは野菜や果物である。これらは人工物と異なり、同じカテゴリ内に色、形状ともに非常に似ている物体が含まれている。よって、物体を特定できず、類似する物体同士で誤認識が起こった。

7. まとめ

本論文では、HONVとLLCを適用し、距離空間ピラミッドを用いた、新しい3次元物体認識手法を提案した。全体的な物

体形状を表現した提案手法は、3次元情報を含むデータセットでの実験において、従来手法に比べて認識率が向上した。今後は、距離空間ピラミッドの部分領域分割法の改良や、それに対応するプーリング手法を作成する予定である。また、部分領域同士に関係性を持たせることで、物体の形状をより正確に表現することも検討していく。

文 献

- [1] D. G. Low, "Distinctive image features from scale-invariant keypoints," *Journal of Computer Vision*, Vol. 60, No. 2, pp. 91–110, 2004.
- [2] G. Csurka, C.R. Dance, L. Fan, J. Willamowski, C. Bray, "Object Recognition as Machine Translation: Learning aLexicons for a Fixed Image Vocabulary," *ECCV Workshop on Statistical Learning in Computer Vision*, pp. 1–22, 2004.
- [3] S. Lazebnik, C. Schmid and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp. 2169–2178, 2006.
- [4] S. Tang, X. Wang, X. Lv, T. X. Han, J. Keller, Z. He, M. Skubic, and S. Lao, "Histogram of Oriented Normal Vectors for Object Recognition with a Depth Sensor," *The Asian Conference on Computer Vision*, 2012.
- [5] N. Dalal, and B. Triggs, "Histograms of oriented gradients for human detection," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 886–893, 2005.
- [6] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang and Y. Guo, "Locality-constrained Linear Coding for Image Classification," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3360–3367, 2010.
- [7] J. HARTIGAN, "A K-means clustering algorithm," *Applied Statistics*, pp. 100–108, 1979.
- [8] J. Yang, K. Yu, Y. Gong and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1794–1801, 2009.
- [9] H. Lee, A. Battle, R. Raina, and A. Ng, "Efficient sparse coding algorithms," *Advances in Neural Information Processing Systems*, MIT Press, pp. 801–808, 2006.
- [10] K. Yu, T. Zhang, and Y. Gong, "Nonlinear learning using local coordinate coding," *Neural Information Processing Systems (NIPS)*, 2009.
- [11] K. Lai, L. Bo, X. Ren, and D. Fox, "A Large-Scale Hierarchical Multi-View RGB-D Object Dataset," *IEEE International Conference on Robotics and Automation*, pp. 1817–1824, 2011.
- [12] C. Cortes and V. Vapnik, "Support-Vector Networks," *Machine Learning*, Vol. 20, No. 3, pp. 273–297, 1995.
- [13] L. Bo, X. Ren and D. Fox, "Depth Kernel Descriptors for Object Recognition," *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 821–826, 2011.
- [14] M. Blum, J. Springenberg, J. Wlfling, and M. Riedmiller, "A Learned Feature Descriptor for Object Recognition in RGB-D Data," *IEEE International Conference on Robotics and Automation*, pp. 1298–1303, 2012.
- [15] L. Bo, X. Ren, and D. Fox, "Hierarchical Matching Pursuit for Image Classification: Architecture and Fast Algorithms," *Neural Information Processing Systems (NIPS)*, 2011.
- [16] L. Bo, X. Ren, and D. Fox, "Unsupervised Feature Learning for RGB-D Based Object Recognition," *In International Symposium on Experimental Robotics*, 2012.
- [17] RGB-D Object Dataset, <http://www.cs.washington.edu/rgbd-dataset/>