

距離画像を用いた空間のカテゴリ識別

○Oscar Martinez Mozos** 水谷 仁* 蔡 現旭** 倉爪 亮** 長谷川 勉**

* 九州大学大学院システム情報科学府

** 九州大学大学院システム情報科学研究所

E-mail: {omozos,mizutani, huchae}@irvs.is.kyushu-u.ac.jp

E-mail: {kurazume,hasegawa}@ait.kyushu-u.ac.jp

1. はじめに

人と同様に周囲環境の状況を瞬時に認識・判断する機能、あるいはロボットと人との認識結果を共有する仕組みは、人と共生するサービスロボットを実現するうえで極めて重要である [1]。その機能の一つに、オフィスや廊下など、ロボット自身によるロボットの置かれた空間の認識 (place recognition) がある。空間の認識が実現すれば、トポロジカル地図や知覚地図の作成 [2, 3]、あるいは物体同定のためのコンテキストとしての利用など様々なアプリケーションが考えられる。

空間の認識は2つの問題に分けて考えることができる。すなわち、特定の空間の認識 (*instance classification*) と空間の識別 (*categorization*) である。特定の空間の認識は、ロボットが以前に訪れたある特定の空間 (例えば 965 号室など) を、センサ情報から認識するものである。一方、空間の識別は、ロボットがそれぞれの空間のカテゴリ (クラス) を識別する問題であり、例えば先ほどの 956 号室は、他の部屋と同様により一般的な概念である “オフィス” として識別される。本稿では、後者の問題を取り扱う。

空間の認識に対する従来手法では、これまでに様々なセンサが用いられている。例えば、[2] では 2D レーザスキャナが入力として用いられ、[4, 5, 6] では画像センサが用いられた。また、これらの2つのセンサを統合した手法も提案されている [7, 8]。一方、ごく最近では様々なロボットアプリケーションにおいて、Kinect センサ [9] が盛んに用いられつつある。このセンサは低価格で、かつ高速に面状の距離情報を得ることができ、我々も今回、このセンサを用いて空間の識別を行うことを試みる。

本稿では、ロボットによる空間の識別の最初のステップとして、Kinect センサにより集められた、様々な空間の距離画像群からなる新たな距離画像データセットを紹介する。さらに局所二値パターン [10, 5] と空間ピラミッドマッチング [11] に基づく空間の識別手法を提案し、構築した距離画像データセットの一部を用いた識別実験の結果を示す。

2. 距離画像データセット

まず、移動台座の上部、高さ 125cm の位置に Kinect センサを固定し、様々な空間で連続的に計測を行い、距離画像データセットを構築した。対象とした空間のカテゴリは “corridor”, “kitchen”, “laboratory”, “office”, “study room” および “meeting room” の6つである。ま



図 1 左: “kitchen” に対応する距離画像。暗い部分ほどカメラに近いことを示している。右: 対応する RGB 画像

category	total places	total sequences
corridor	6	9
kitchen	4	11
laboratory	4	20
office	2	3
study room	2	5
meeting room	2	4

表 1 距離画像データセットのデータ数

ず、それぞれのカテゴリ毎に、いくつかの異なる空間 (部屋) で距離画像の取得を行った。具体的には、それぞれの空間で Kinect センサを異なる軌道を移動させ、1つあるいは複数の距離画像列を得た。例えば、カテゴリ “laboratory” は4つの異なる laboratory の情報を含み、それぞれの laboratory 毎に複数の距離画像列、合計で20の距離画像列からなる。表1はデータセットに含まれるカテゴリ毎の空間数、および距離画像列数を示す。またデータセットに含まれる “kitchen” の画像の一例を図1に示す。

3. 距離画像の特徴表現

距離画像から空間のカテゴリを識別するために、まずそれぞれの距離画像を、局所二値パターン (Local Binary Patterns, LBP) [10] のヒストグラムとして表現する。

LBP は近傍画素の情報を用いて注目画素に新たなラベルを付加するものであり、今回は距離画像に対して以下のように適用した。まず、距離画像 D 中のそれぞれの画素 i に対して、その距離値を周囲 8 近傍の画素

8	163	208	1	0	0
28	100	35	1		1
131	62	51	0	1	1

→ (11101100)₂ = 236

図2 中心画素(太字の100)に対するLBPオペレータの計算例. 236がセットされる

$j = \mathcal{N}(i)$ の距離値 $D(j)$ と比較する. 次に, それぞれの近傍画素ごとに二値 $\{0, 1\}$ を以下の式で決定した.

$$v(j) = \begin{cases} 1 & \text{if } D(i) > D(j); \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

得られた周囲画素の二値 $v(j)$ は決められた順番で並べられ, 図2に示すように0から255までのラベル d に変換される. 注目画素の距離値は, このラベル d により置き換えられ, 新たな特徴量画像 T が生成される. なお同様の方法は, 濃淡画像を用いた空間の識別手法として [5] でも用いられている.

Kinect カメラは約0.5mから3.5mの範囲を計測可能であり, 計測範囲外は距離値に nan がセットされる. そこで, 距離値が nan である画素, あるいは上記のLBPの計算で周囲画素に nan が含まれる場合には, LBPの出力を強制的に256とした. 従って, 変換された特徴量画像 T の画素値は以下のように計算される.

$$T(i) = \begin{cases} 256 & \text{if } \delta_{nan}(I(i)) \vee \exists \delta_{nan}(\mathcal{N}(j)); \\ d & \text{otherwise,} \end{cases} \quad (2)$$

ここで $\delta_{nan}(\cdot)$ は, もし距離値に nan が設定されていた場合には1を, それ以外は0を返す関数であり, d はLBPにより得られた特徴量である(図2).

次に, [5] と同様に特徴量画像 T をヒストグラム H で表現する. ここでヒストグラム H のそれぞれのピンは, 対応する画素値 $d \in [0, 255]$, および例外である画素値256をもつ特徴量画像 T 中の画素数である.

次に, 特徴量画像 T に対して空間ピラミッド [11] を生成する. まず解像度のレベル l を図3に示すように定義する. すなわち, 初期画像を $2^l \times 2^l$ 個の領域に分割したものをレベル l の画像とする. 次に分割したそれぞれの領域で局所ヒストグラム H^l を計算する. 最終的なヒストグラム H は, 各レベルの局所ヒストグラムを結合したものと表される. さらに [11] と同様に, それぞれのヒストグラムにはそれぞれの重要度 w が定義されている.

4. 識別方法

空間のカテゴリの識別には, 空間マッチング法 [11] を用いる. 我々の実装では, 各特徴量画像 T のそれぞれの画素値 f は $F = 257 (= 256 + non)$ 段階を有する.

特徴量画像 T_x, T_y に対し, それぞれのヒストグラムが H_x, H_y で与えられたとき, レベル l の空間ピラミッドにおける, ある画素値 f のヒストグラム間類似度 $I_f^l(H_x(f), H_y(f))$ は以下のように定義される [12].

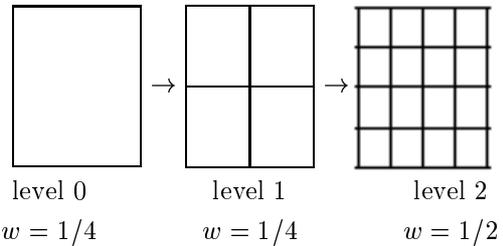


図3 3つのレベルからなる空間ピラミッドとそれぞれの重要度

$$I_f^l(H_x(f), H_y(f)) = \sum_{i=1}^C \min(H_x^i(f), H_y^i(f)), \quad (3)$$

ここで $C = 2^l \times 2^l$ はレベル l の分割領域数(図3)であり, $H_x^i(f)$ はレベル l での分割領域 i の局所ヒストグラムにおける画素値 f の度数を表す. また, 特徴量画像 T_x, T_y 間の特徴カーネルは以下の式で定義される [11].

$$\kappa_f^L(T_x, T_y) = \frac{1}{2^L} I_f^0 + \sum_{l=1}^L \frac{1}{2^{L-1+l}} I_f^l, \quad (4)$$

ただし $I_f^l(H_x(f), H_y(f))$ を I_f^l とした. 最後に, それぞれの特徴カーネル $\kappa_f^L(T_x, T_y)$ の和として, 特徴量画像間のカーネル $K^L(T_x, T_y)$ が求められる [11].

$$K^L(T_x, T_y) = \sum_{f=1}^F \kappa_f^L(T_x, T_y) \quad (5)$$

得られたカーネル $K^L(T_x, T_y)$ を, Support Vector Machine (SVM) により学習, 識別する.

5. 識別実験

本章では, 構築した距離画像データセットを用いた空間のカテゴリ識別実験の結果を示す. 本実験の目標は, 屋内環境の異なる空間における提案手法の識別性能を調べることであり, その第一歩として, 距離画像データセットの一部を用いた実験を行った. 具体的には, “corridor”, “kitchen”, “lab”, “study room” の4カテゴリを用いた. なお, それぞれのカテゴリには, 例えば “lab” は4つの異なる実験室のデータからなるなど, 同一のカテゴリには属するが実際には異なる空間(部屋)の情報が含まれている. さらに, それぞれの距離画像は一続きの画像列から構成されている. 実験で用いたデータセットの詳細は表2を参照されたい.

実験では, leave-one out cross-validation 法により, 提案手法の識別性能を評価した. それぞれのカテゴリに含まれる空間数が異なることから, 学習データと評価データの N 個の組を以下のように決定した. まず各カテゴリに含まれるある一つの空間を評価データとする. この空間はそれぞれのカテゴリ内でランダムに決定される. また残りの空間のデータは, 学習データとして使用される. 評価データは, 例えば $\{c1, k2, l2, s1\}$ あるいは $\{c1, k1, l4, s1\}$ などである. もし選択された

place id	category	images
c1	corridor	30
c2	corridor	142
c3	corridor	29
k1	kitchen	22
k2	kitchen	23
l1	laboratory	80
l2	laboratory	76
l3	laboratory	32
l4	laboratory	31
s1	study room	29
s2	study room	30

表2 実験で用いた距離画像データセット．“images”列は含まれる画像数を示す．

%	corridor	kitchen	lab	study
corridor	90.0	4.0	5.0	0.0
kitchen	5.2	77.4	17.4	0.0
lab	2.5	10.9	86.6	0.0
study	0.0	31.0	37.3	31.7

表3 空間の識別実験の結果，“study”は“study room”を表す

評価データの組が既に評価に使用されていた場合には、それを破棄し、それ以前とは異なる学習、評価データの組が得られるまで選択処理を続ける。これにより、評価データは常に学習データには含まれない空間となる。なお、本実験では $N = 10$ とした。また、それぞれの評価時には、SVM を学習データを式5で示したカーネルにより学習させ、評価データをそれぞれ学習後のSVMで評価して識別率を求めた。またSVNの学習には *libsvm* [13] を用いた。

実験結果を表3に示す。それぞれの値は実験を10回繰り返した時の識別精度の平均値である。実験の結果、“corridor”、“kitchen”、“lab”については77.4～90%の高い識別率が得られているが、“study room”のカテゴリに属する距離画像は、他のカテゴリに比べて識別精度が低いことがわかった。これは、“study room”には椅子や机のみが多く含まれ、“kitchen”や“lab”との差が少ないことが原因として考えられる。一方、“corridor”は平面が多く、他の物体が少ないなど、他のカテゴリに対して明確な差異が存在し、識別が容易であることがわかる。

6. 結論と今後の課題

本稿では、Kinect センサにより様々な屋内環境を撮影し、得られたデータから構築した距離画像データセットを紹介した。また初期実験として、距離画像データセットの一部を用いて、空間のカテゴリ識別実験を行った。実験の結果、概ね高い識別率が得られているが、

“study room”など識別が難しいカテゴリも存在することが明らかとなった。今後は、より適当な記述子や識別器について検討するとともに、学習、評価データのカテゴリ数やデータ数を増やすことを目標とする。加えて、構築したデータセットの公開についても検討する予定である。

謝辞

本研究の一部は、日本学術振興会外国人特別研究員制度(00243890)の支援を受けた。また実験にご協力いただいた九州大学内田誠一教授、谷口倫一郎教授に感謝いたします。

参考文献

- [1] H. I. Christensen, G.J. Kruijff, and J. Wyatt, eds.: Cognitive Systems, COSMOS, Springer Verlag, 2010.
- [2] Oscar Martinez Mozos and Wolfram Burgard: “Supervised Learning of Topological Maps using Semantic Information Extracted from Range Data.”, IEEE/RSJ International Conference on Intelligent Robots and Systems, Beijing, China, 2006.
- [3] Hendrik Zender, Oscar Martinez Mozos, Patric Jensfelt, Geert-Jan M. Kruijff, and Wolfram Burgard: “Conceptual Spatial Representations for Indoor Mobile Robots”, Robotics and Autonomous Systems, 56, 6, pp. 493–502, 2008.
- [4] Antonio Torralba, Kevin P. Murphy, William T. Freeman, and Mark A. Rubin: “Context-based Vision System for Place and Object Recognition”, International Conference on Computer Vision, Nice, France, 2003.
- [5] Jianxin Wu and James M. Rehg: “CENTRIST: A Visual Descriptor for Scene Categorization”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 33, 8, pp. 1489–1501, 2011.
- [6] Ananth Ranganathan: “PLISS: Detecting and Labeling Places Using Online Change-Point Detection”, Robotics: Science and Systems VI, 2010.
- [7] Axel Rottmann, Oscar Martinez Mozos, Cyrill Stachniss, and Wolfram Burgard: “Semantic Place Classification of Indoor Environments with Mobile Robots using Boosting.”, National Conference on Artificial Intelligence, Pittsburgh, PA, USA, 2005.
- [8] Andrzej Pronobis, Oscar Martinez Mozos, Barbara Caputo, and Patric Jensfelt: “Multi-modal Semantic Place Classification”, International Journal of Robotics Research, 29, 2–3, pp. 298–320, 2010.
- [9] Microsoft Kinect: <http://www.xbox.com/en-us/kinect>.
- [10] Timo Ojala, M Pietikainen, and David Harwood: “A Comparative Study of Texture Measures with Classification Based on Featured Distributions”, Pattern Recognition, 29, 1, pp. 51–59, 1996.
- [11] Svetlana Lazebnik, Claudia Schmid, and Jean Ponce: “Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories”, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2006.
- [12] Kristen Grauman and Trevor Darrell: “The pyramid match kernel: Discriminative classification with sets of image features”, IEEE International Conference on Computer Vision, pp. 1458–1465, 2005.
- [13] Chih-Chung Chang and Chih-Jen Lin: “LIBSVM: A library for support vector machines”, ACM Transactions on Intelligent Systems and Technology, 2, pp. 27:1–27:27, 2011.