

RGB-D カメラを用いた屋内環境のカテゴリ識別

水谷 仁^{*1}, マルティネス モゾス オスカル^{*2}, 倉爪 亮^{*2},
岩下 友美^{*2}, 長谷川 勉^{*2}

Indoor Space Categorization Using RGB-D Camera

Hitoshi MIZUTANI^{*1}, Oscar MARTINEZ MOZOS^{*2}, Ryo KURAZUME^{*2}, Yumi
IWASHITA^{*2}, and Tsutomu HASEGAWA^{*2}

^{*1} Graduate School of Information Science and Electrical Engineering, Kyushu University
744 Motoooka, Nishi-ku, Fukuoka 819-0395 Japan

^{*2} Faculty of Information Science and Electrical Engineering, Kyushu University
744 Motoooka, Nishi-ku, Fukuoka 819-0395 Japan

An important capability for service robots working and interacting with humans in indoor environments is their ability for categorizing the different places where they are located. In this paper, we present a supervised method to categorize different areas in indoor environments by a mobile robot. Our approach uses RGB and depth images provided by a Kinect camera. All RGB images are converted into gray-scale images. Each gray-scale and depth images at same time are transformed into respective histograms of local binary patterns. Both histograms are combined to integrate gray and depth information. The final combined histogram is categorized using support vector machines. We apply this method to distinguish five different place categories: corridor, laboratory, office, kitchen, and study room. The classification results show that we can categorize these places with high accuracy using both gray and depth information.

Key Words : Place Categorization, Local Binary Pattern, Mobile Robot, RGB-D Camera

1. はじめに

サービスロボットに求められる重要な機能として、ロボットが置かれた屋内環境の空間の認識 (*place recognition*) が挙げられる。空間の認識は高度な位置推定問題であり、ロボットに対し正確な座標の推定ではなく、より抽象度の高い空間の種類を認識する能力が求められる。ロボットは空間の種類に関する情報を得ると、次に取るべき行動を決定できる⁽¹⁾⁽²⁾。さらに、空間の種類に関する情報は、典型的な物体の認識において事前情報またはコンテキストとして用いることができ⁽³⁾⁽⁴⁾、またセマンティックマップ⁽⁵⁾⁽⁶⁾や空間の概念的表現⁽⁷⁾⁽⁸⁾などにも利用できる。また空間の認識において、ロボットは自身が置かれた、または見ている空間に対しラベルを割り当てる。そのとき、“office”や“laboratory”のように、人が一般的に用いる名前をラベルとして用いれば、人とロボットの直感的なコミュニケーションが実現できる⁽⁸⁾⁽⁹⁾。

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。

空間の識別では、これまでに様々なセンサが用いられている。例えば、レーザレンジファインダ⁽⁵⁾⁽¹⁰⁾⁽¹¹⁾や画像センサ⁽³⁾⁽¹²⁾⁽¹³⁾、あるいはこれらの 2 つのセンサを統合した手法も提案されている⁽¹⁴⁾⁽¹⁵⁾。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⁽¹⁶⁾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る。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、Kinect センサにより集められた RGB および距離画像を用いて、空間の識別を行う手法を提案する。提案手法では、まず RGB 画像を濃淡画像に変換し、その濃淡画像および同時刻に得られた距離画像

^{*1} 九州大学大学院システム情報科学府 (〒 819-0395 福岡市西区元岡 744 番地) mizutani@irvs.ait.kyushu-u.ac.jp

^{*2} 九州大学大学院システム情報科学府 (〒 819-0395 福岡市西区元岡 744 番地) omozos@irvs.ait.kyushu-u.ac.jp {kurazume, yumi, hasegawa}@ait.kyushu-u.ac.jp

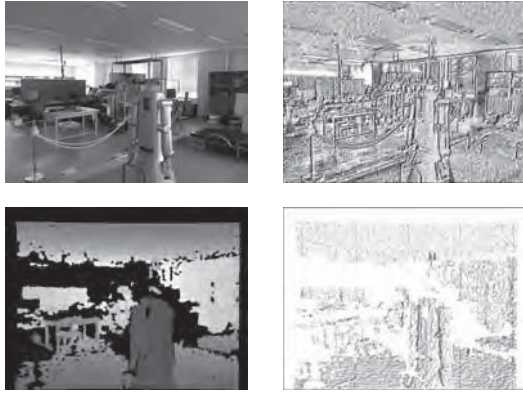


Fig. 1 The top-left picture shows a gray-scale image of a laboratory. The corresponding depth image is shown in the bottom-left image (gray values have been equalized for a better visualization). In the depth image completely black pixels correspond to undefined depth values delivered by the Kinect camera. The outputs of the local binary pattern transformation of the gray-scale and depth images are shown in the top-right and bottom-right images, respectively.

を、それぞれ局所二値パターン⁽¹⁷⁾と局所ヒストグラムを用いて、近傍画素間の関係を表した特徴ベクトルに変換する。濃淡画像および同時刻に得られた距離画像と、それらの局所二値パターンの例を図1に示す。その後、得られた濃淡および距離情報の特徴ベクトルを結合し、両者を統合した特徴ベクトルとする。この特徴ベクトルを用いて、サポートベクトルマシンにより空間を識別する。

第2章から第4章では、それぞれ関連手法、濃淡および距離画像を用いた空間の識別、識別実験について述べる。

2. 関 連 手 法

近年、移動ロボットによる空間の識別技術が向上し、ロボットは様々なセンサにより典型的な屋内空間が識別できるようになった。例えば、レーザレンジファインダにより得られる2D画像は、屋内環境で異なる空間を識別するのに用いることができる。Mozosらは2D画像をいくつかの幾何学的特徴を含むベクトルに変換し、それらをAdaBoostを用いて分類した⁽⁵⁾⁽¹⁸⁾。さらに、Friedmanらは、ボロノイ図の各点に条件付き確率場を適用した。各点はそれぞれの地点でレーザ計測により得られる空間の幾何的特徴も表している。さらに、BrunskillらはSpectral Clusteringを用いて、屋内環境の2D下位マップを作成した⁽¹¹⁾。その下位マップは局所的な2Dスキャンから得られる特徴ベクトルを用いて分類される。また、Shiらは全体をスキャン

せず、それぞれの独立したスキャンを意味的ラベルへ分類する手法を提案している⁽¹⁹⁾。

一方、視覚センサも空間の認識に多く用いられてきた。例えば、WuとRehgによる手法では、家の中の複数の空間の識別にCENTRIST記述子を用いた⁽¹²⁾。この記述子は、Ojalaらが提案している局所二値パターンと同様である⁽¹⁷⁾。さらに、Ranganathanは空間の識別のためのPLISシステムを提案している⁽¹³⁾。この手法では変化点を見つけることにより、一続きの画像列を分割する。そのあと、分割された画像はそれぞれ視覚カテゴリに分類される。Torralbaらはウェアラブルカメラから得られる画像の局所特徴量と大域特徴量を用いて、隠れマルコフモデルにより屋内・屋外環境の特定の空間の認識および空間の識別を行った⁽³⁾。

空間を識別するために、異なる種類のセンサを組み合わせる試みも行われてきた。Rottmannらは2Dレーザスキャンと視覚による物体検出を組み合わせ、ブースティングを用いて屋内環境の空間を識別した⁽¹⁴⁾。Pronobisらは複数の視覚とレーザによる手がかりを組み合わせ、サポートベクトルマシンを用いて、屋内環境の空間を識別した⁽¹⁵⁾。

3. 濃淡および距離画像を用いた空間の識別

3.1 局所二値パターンを用いた濃淡画像からの特徴量画像生成

局所二値パターン (Local Binary Patterns, LBP) 演算子⁽¹⁷⁾は、濃淡画像のテクスチャ分析と分類を目的にOjalaにより提案された。本稿ではそれを濃淡画像と距離画像に適用し、典型的な屋内空間の識別を行う。

濃淡画像を用いた空間の識別では、まずKinectセンサから得られるRGB画像を濃淡画像に変換¹し、その濃淡画像をLBPを用いて特徴量画像に変換する。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、今回は濃淡画像に対して以下のように適用した。まず、濃淡画像のそれぞれの画素*i*に対して、その画素値 p_i を周囲8近傍の画素値 $p_j \in \mathcal{N}(p_i)$ 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。

$$b_j = \begin{cases} 1, & p_i > p_j \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

得られた周囲画素の二値 b_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 v_i に変換される。注目画素の値は、このラベル v_i により

¹R, G, Bチャンネルのそれぞれに適用することも可能。ただし什器や壁は、形状は同一でも色が異なるものもあり、色情報の有効性は今後の検討課題である

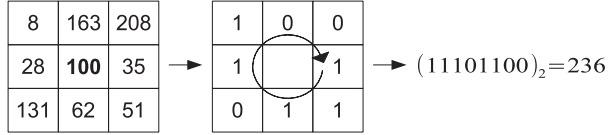


Fig. 2 Toy example for the LBP operator for the center pixel (marked in bold). The resulting value 236 is assigned to that pixel.

置き換えられ、LBPに変換された特徴量画像 T_{gray} が生成される。

3.2 局所二値パターンを用いた距離画像からの特徴量画像生成 同様に Kinect センサから得られる距離画像に、LBPを適用して特徴量画像を得ることができる。しかし、Kinect センサでは、計測不能な箇所および計測範囲外の距離値に nan がセットされるため、距離画像に対して単純に LBP の計算ができない。そこで、距離値が nan である画素、あるいは前章の 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 256 とした。従って、LBP を用いて得られた特徴量画像 T_{depth} の画素値 v_i^+ は以下のように計算される。

$$v_i^+ = \begin{cases} 256, & \delta_{\text{nan}}(p_i) \vee \exists \delta_{\text{nan}}(\mathcal{N}(p_i)) \\ v_i, & \text{otherwise} \end{cases} \quad (2)$$

ここで $\delta_{\text{nan}}(\cdot)$ 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、 v_i は LBP により得られた特徴量である (図 2)。

3.3 特徴量ベクトルの生成 3.1, 3.2 項の方法で得られた特徴量画像 T_{gray} , T_{depth} を以下の方法でそれぞれ特徴ベクトル \mathbf{x}_{gray} , $\mathbf{x}_{\text{depth}}$ に変換する。

まず、特徴量画像 T (T_{gray} および T_{depth}) に対して空間ピラミッド⁽²⁰⁾を生成する。ただし、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 $p = 2^l \times 2^l$ 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_p^l を計算する。ここで局所ヒストグラム H_p^l のそれぞれのビンは、レベル l の p 番目の画像において対応する画素値の画素数である。最後に、特徴ベクトル \mathbf{x}_T 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。

$$\mathbf{x}_T = \{\{H_1^0\}, \{H_1^1, \dots, H_{2^1 \times 2^1}^1\}, \dots, \{H_1^L, \dots, H_{2^L \times 2^L}^L\}\} \quad (3)$$

3.4 濃淡情報と距離情報の統合 濃淡情報と距離情報を統合するため、前項の方法で得られた特徴ベクトル \mathbf{x}_{gray} と $\mathbf{x}_{\text{depth}}$ を結合し、最終的な特徴ベクトル

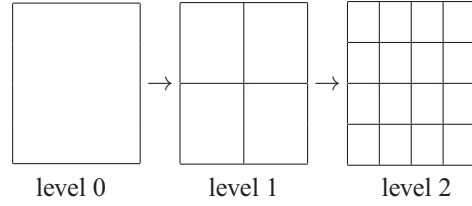


Fig. 3 Spatial pyramid with 3 levels.

\mathbf{x} とする。この特徴ベクトル \mathbf{x} が識別器の入力として用いられる。

$$\mathbf{x} = \{\mathbf{x}_{\text{gray}}, \mathbf{x}_{\text{depth}}\} \quad (4)$$

3.5 識別方法 各画像を特徴ベクトルに変換し、Support Vector Machine (SVM)⁽²¹⁾⁽²²⁾を用いて識別を行う。SVMには M 対の濃淡および距離画像の組から得られた M 個の m 次元特徴ベクトル $\mathbf{x}_i \in \mathfrak{R}^m (i = 1, \dots, M)$ とともに、ラベル $y_i \in \{1, -1\}$ (すなわちそのカテゴリに属するか、属さないか) が入力される。SVM の目的は 2 クラス $\{1, -1\}$ の学習データとの距離が最大となる超平面を見つけることであり、次の最適化問題で表される。

$$\min_{\mathbf{w}, b, \xi} C \sum_{i=1}^M \xi_i + \frac{1}{2} \|\mathbf{w}\|^2 \quad (5)$$

$$\text{subject to } y_i (\mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}_i) + b) \geq 1 - \xi_i \quad (6)$$

ここで、 C はマージンパラメータ、 \mathbf{w} は超平面に対する垂線、 $\xi_i \geq 0$ は \mathbf{x}_i の誤分類に伴う超平面からの逸脱量、 $\boldsymbol{\phi}(\cdot)$ は非線形写像、 b はバイアス項を表している。実験では次のラジアル基底関数 (Radial Basis Function) カーネルを用いた。

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2), \quad \gamma > 0 \quad (7)$$

ここで、 γ は分布の半径を制御するパラメータである。

多クラス問題の識別は一対一方式⁽²³⁾で行い、実験では LIBSVM ライブラリ⁽²⁴⁾を用いた。パラメータ C および γ は、 $C \in [2^{-5}, 2^{-2}, \dots, 2^{16}]$ および $\gamma \in [2^{-12}, 2^{-9}, \dots, 2^3]$ の範囲でクロスバリデーションを用いた総当たり探索により決定した⁽²⁵⁾。

4. 識別実験

4.1 データセット 提案手法の性能評価のために、屋内環境で実験を行った。実験では、まず移動台車の上部、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、RGB および距離画像の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” および “study room” の 5 つである。まず、それぞれの

Table 1 Number of image sets by place.

category	place	number of image sets
corridor	corridor 1	68
	corridor 2	42
	corridor 3	70
	corridor 4	99
kitchen	kitchen 1	73
	kitchen 2	65
	kitchen 3	53
laboratory	laboratory 1	99
	laboratory 2	99
	laboratory 3	81
	laboratory 4	78
study room	study room 1	71
	study room 2	70
	study room 3	49
	study room 4	62
office	office 1	57
	office 2	45
	office 3	47

カテゴリ毎に、いくつかの異なる空間（部屋）で RGB および距離画像の取得を同時に行った。例えば、表 1 ではカテゴリ“laboratory”は 4 つの異なる laboratory の情報を含んでおり、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。また、それぞれの空間で Kinect センサを異なる軌道で移動させ、RGB および距離画像列の対を複数得た。このあと、各 RGB 画像列とそれに対応する距離画像列の同期をとり、同時刻の RGB および距離画像を得た。表 1 は実験で用いた空間毎の画像の組（RGB 画像および距離画像）の数を示す。また、データセットに含まれる各カテゴリの RGB および距離画像の一例を図 4 に示す。

4.2 識別性能 構築した RGB および距離画像のデータセットを用い、空間のカテゴリ識別実験を行った。

実験では、leave-one out cross-validation 法を N 回適用し、提案手法の識別性能を評価した。ただし、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。また同じカテゴリの残りの空間のデータは、学習データとして使用される。もし選択された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、学習データを式 (7) で示したカーネルにより SVM に学習させ、評価データを

Table 2 Classification results for experiment 1.

Confusion matrix for $L = 0$, average correct **89.37%**

%	corridor	kitchen	laboratory	study room	office
corridor	95.66	2.02	2.32	0.00	0.00
kitchen	2.45	76.85	3.21	12.40	5.10
laboratory	0.00	0.12	99.02	0.23	0.63
study room	0.00	8.86	5.57	84.64	0.93
office	0.00	3.51	5.79	0.00	90.70

Confusion matrix for $L = 1$, average correct **85.68%**

%	corridor	kitchen	laboratory	study room	office
corridor	97.58	1.01	0.40	1.01	0.00
kitchen	1.13	56.04	8.87	32.45	1.51
laboratory	0.00	0.20	98.88	0.66	0.26
study room	0.00	9.43	7.57	81.86	1.14
office	0.00	2.81	1.93	1.23	94.04

Confusion matrix for $L = 2$, average correct **84.59%**

%	corridor	kitchen	laboratory	study room	office
corridor	100.00	0.00	0.00	0.00	0.00
kitchen	0.94	51.51	2.83	41.32	3.40
laboratory	0.00	0.00	95.17	4.83	0.00
study room	0.00	8.00	9.14	81.71	1.14
office	0.00	4.91	0.53	0.00	94.56

それぞれ学習後の SVM で評価して識別率を求めた。

実験 1 では提案手法の識別性能を評価した。すなわち、濃淡および距離画像から LBP を用いてそれぞれ特徴ベクトルを生成し、それらを組み合わせた特徴ベクトルにより識別を行った。空間ピラミッドのレベルは 3 種類 ($L = 0; 1; 2$) とした。得られた識別性能は表 2 に示す。表 2 より、すべてのレベルにおいて“corridor”、“laboratory”の識別精度が最も高かった。これは、これらのカテゴリの空間は一般的に構造が似ているため、識別が容易であるためと考えられる。一方、“kitchen”は最も識別精度が低かった。これは特に“kitchen”と“study room”の識別の難しかったことが挙げられる。クラスごとの識別精度の平均は、 $L = 0$ のとき 89.37%、 $L = 1$ のとき 85.68%、 $L = 2$ のとき 84.59% であった。

一方、比較のため、濃淡および距離画像を個々に用いた場合の結果を実験 2, 3 に示す。実験 2 では、濃淡画像のみを用いて識別性能を評価した。表 3 は、実験 2 で最も識別精度が高かった $L = 2$ のときの混同行列である。また、実験 3 では、距離画像のみを用いて識別性能を評価した。表 4 は、実験 3 で最も識別精度が高かった $L = 1$ のときの混同行列である。

実験 4, 5, 6 では、実験 1, 2, 3 との比較のため、LBP を用いない場合の結果を示す。

実験 4 では、距離画像を 257 (= 256 + nan) 段階に離散化したあと、濃淡および距離画像からヒストグラムを直接抽出し、それぞれ特徴ベクトルを生成し、そ

Table 3 Classification results for experiment 2.

Confusion matrix for $L = 2$, average correct **82.18%**

%	corridor	kitchen	laboratory	study room	office
corridor	97.27	2.73	0.00	0.00	0.00
kitchen	10.19	53.43	2.26	34.12	0.00
laboratory	0.00	0.20	86.24	11.95	1.62
study room	1.43	9.86	11.16	76.43	1.13
office	0.00	2.46	0.00	0.00	97.55

Table 4 Classification results for experiment 3.

Confusion matrix for $L = 1$, average correct **85.11%**

%	corridor	kitchen	laboratory	study room	office
corridor	99.50	0.10	0.40	0.00	0.00
kitchen	0.00	55.10	24.15	20.19	0.57
laboratory	0.37	0.50	94.46	4.03	0.64
study room	0.00	7.14	6.42	85.58	0.86
office	0.00	2.99	5.40	0.70	90.90

Table 5 Classification results for experiment 4.

Confusion matrix for $L = 2$, average correct **72.19%**

%	corridor	kitchen	laboratory	study room	office
corridor	91.63	2.94	3.86	1.57	0.00
kitchen	35.54	35.90	18.93	9.62	0.00
laboratory	8.14	8.94	69.76	12.42	0.74
study room	1.57	5.86	19.72	71.28	1.57
office	0.00	4.35	2.22	1.05	92.38

Table 6 Classification results for experiment 5.

Confusion matrix for $L = 2$, average correct **70.26%**

%	corridor	kitchen	laboratory	study room	office
corridor	95.18	0.00	4.82	0.00	0.00
kitchen	12.62	40.09	33.14	10.00	4.15
laboratory	4.64	5.08	79.78	9.76	0.74
study room	0.29	5.43	17.81	76.48	0.00
office	0.00	1.59	26.24	12.39	59.78

れらを組み合わせた場合の識別性能を評価した。表5は、実験4で最も識別精度が高かった $L = 2$ のときの混同行列である。

実験5では、濃淡画像のみを用いて識別性能を評価した。表6は、実験5で最も識別精度が高かった $L = 2$ のときの混同行列である。

実験6では、距離画像のみを用いて識別性能を評価した。表7は、実験6で最も識別精度が高かった $L = 2$ のときの混同行列である。

実験1~6の識別結果の比較を表8に示す。各レベルの最も高い識別精度を太字で示した。表8において、Gray-LBPは濃淡画像に対するLBP、Depth-LBPは距離画像に対するLBP、Grayは濃淡画像に対する256

Table 7 Classification results for experiment 6.

Confusion matrix for $L = 2$, average correct **54.05%**

%	corridor	kitchen	laboratory	study room	office
corridor	84.46	2.93	3.44	3.72	5.46
kitchen	25.36	29.28	11.40	33.96	0.00
laboratory	5.45	3.52	70.85	17.36	2.83
study room	18.07	3.99	48.69	27.11	2.14
office	20.40	8.69	8.25	4.13	58.54

Table 8 Summary of results in %.

exp.	descriptor	$L = 0$	$L = 1$	$L = 2$
1	combination of 2 and 3	89.37	85.68	84.59
2	Gray-LBP	76.60	79.47	82.18
3	Depth-LBP	80.72	85.11	83.14
4	combination of 5 and 6	60.94	66.77	72.19
5	Gray	45.15	50.79	54.05
6	Depth discretized	70.24	67.34	70.26

段階の離散値、Depth discretizedは距離画像に対する257 (= 256 + nan)段階の離散値を表している。表8から、距離画像と濃淡画像を同時に用いた場合が最も識別精度が高く、次いで距離画像を用いた場合、濃淡画像のみを用いた場合の順であった。また、LBPを用いた場合は用いない場合と比べて、全ての場合で識別精度が高かった。これより、距離画像の奥行き情報およびLBPは空間の識別に有効であることがわかる。最も識別精度が高かったのは、 $L = 0$ においてGray-LBPとDepth-LBPを組み合わせた場合であった。なお、本実験では $L = 0$ の時の識別精度が総じて高かったが、局所ヒストグラムは天井付近、床付近など、各カテゴリでの局所的な特徴の違いを表現するために有効であると考えられ、今後は画像の水平方向のみの分割などにより、局所ヒストグラムの有効性をより詳細に検討する予定である。

最後に、実験1において誤識別された濃淡および距離画像の例を図5に示す。

5. 結論と今後の課題

本稿では、Kinectセンサから得られたRGBおよび距離画像を用いて、空間を識別する手法を提案した。提案手法では、濃淡および距離画像からLBPを用いてそれぞれ特徴ベクトルを抽出し、それらを組み合わせた特徴ベクトルをSVMにより識別した。識別実験の結果、提案手法は5つの空間のカテゴリにおいて高い識別精度を示すことが確認された。

今後は、より適当な記述子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。



Fig. 5 Examples of misclassified pairs of RGB and depth images in experiment 1. In the upper row a scene of “kitchen 3” is categorized as a corridor. The middle row show an scene of “laboratory 4” categorized as an office. The bottom row shows a scene in “study room 2” categorized as a kitchen.

謝辞

本研究の一部は、文部科学省科学研究費補助金基盤研究(B) (課題番号 23360115) 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、諸岡健一准教授に感謝いたします。

参考文献

- (1) C. Stachniss, O. M. Mozos, and W. Burgard, “Speeding-up multi-robot exploration by considering semantic place information,” in *Proc. of the IEEE International Conference on Robotics and Automation (ICRA)*, Orlando, FL, USA, 2006, pp. 1692–1697.
- (2) H. Zender, P. Jensfelt, and G.-J. Kruijff, “Human and situation-aware people following,” in *Proc. of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, August 2007, pp. 1131–1136.
- (3) A. Torralba, K. P. Murphy, W. T. Freeman, and M. A. Rubin, “Context-based vision system for place and object recognition,” in *International Conference on Computer Vision (ICCV)*, Nice, France, 2003.
- (4) T. Kollar and N. Roy, “Utilizing object-object and object-scene context when planning to find things,” in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Kobe, Japan, 2009.
- (5) O. M. Mozos, C. Stachniss, and W. Burgard, “Supervised learning of places from range data using AdaBoost,” in *IEEE International Conference on Robotics and Automation (ICRA)*, Barcelona, Spain, 2005.
- (6) A. Pronobis, P. Jensfelt, K. Sjöo, H. Zender, G.-J. M. Kruijff, O. M. Mozos, and W. Burgard, “Semantic modelling of space,” in *Cognitive Systems*, ser. Cognitive Systems Monographs, H. I. Christensen, A. Sloman, G.-J. M. Kruijff, and J. Wyatt, Eds. Germany: Springer, 2010, pp. 165–221.
- (7) H. Zender, O. M. Mozos, P. Jensfelt, G.-J. M. Kruijff, and W. Burgard, “Conceptual spatial representations for indoor mobile robots,” *Robotics and Autonomous Systems*, Vol. 56, No. 6, pp. 493–502, June 2008.
- (8) H. I. Christensen, G.J. Kruijff, and J. Wyatt, Eds., *Cognitive Systems*, ser. COSMOS. Springer Verlag, May 2010.
- (9) H. Zender, P. Jensfelt, O. M. Mozos, G.-J. M. Kruijff, and W. Burgard, “An integrated robotic system for spatial understanding and situated interaction in indoor environments,” in *Proceedings of the Conference on Artificial Intelligence (AAAI)*, Vancouver, British Columbia, Canada, 2007.
- (10) S. Friedman, H. Pasula, and D. Fox, “Voronoi random fields: Extracting the topological structure of indoor environments via place labeling,” in *In Proc. of the International Joint Conference on Artificial Intelligence (IJCAI)*, Hyderabad, India, 2007.
- (11) E. Brunskill, T. Kollar, and N. Roy, “Topological mapping using spectral clustering and classification,” in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, San Diego, October 2007.
- (12) J. Wu, H. I. Christensen, and J. M. Rehg, “Visual place categorization: Problem, dataset, and algorithm,” in *Proc of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, St. Louis, MO, USA, 2009.
- (13) A. Ranganathan, “PLISS: Detecting and Labeling Places Using Online Change-Point Detection,” in *Robotics: Science and Systems VI*, 2010.
- (14) A. Rottmann, O. M. Mozos, C. Stachniss, and W. Burgard, “Semantic place classification of indoor environments with mobile robots using boosting,” in *National Conference on Artificial Intelligence (AAAI)*, Pittsburgh, PA, USA, 2005.
- (15) A. Pronobis, O. M. Mozos, B. Caputo, and P. Jensfelt, “Multi-modal semantic place classification,” *International Journal of Robotics Research*, Vol. 29, No. 2–3, pp. 298–320, February-March 2010.
- (16) Microsoft Kinect, <http://www.xbox.com/en-us/kinect>.
- (17) T. Ojala, M. Pietikainen, and D. Harwood, “A comparative study of texture measures with classification based on featured distributions,” *Pattern Recognition*, Vol. 29, No. 1, pp. 51–59, 1996.

- (18) O. M. Mozos and W. Burgard, "Supervised learning of topological maps using semantic information extracted from range data," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (ICRA)*, Beijing, China, 2006.
- (19) L. Shi, S. Kodagoda, and G. Dissanayake, "Laser range data based semantic labeling of places," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Taipei, Taiwan, October 2010.
- (20) S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2006.
- (21) C. Cortes and V. Vapnik, "Support-vector network," *Machine Learning*, Vol. 20, pp. 273–297, 1995.
- (22) C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
- (23) S. Knerr, L. Personnaz, and G. Dreyfus, "Single-layer learning revisited: a stepwise procedure for building and training a neural network," in *Neurocomputing: Algorithms, Architectures and Applications*, J. Fogelman, Ed., Springer-Verlag, 1990.
- (24) C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, Vol. 2, pp. 27:1–27:27, 2011, software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- (25) C.-W. Hsu, C.-C. Chang, and C.-J. Lin, "A practical guide to support vector classification," <http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>, 2010.

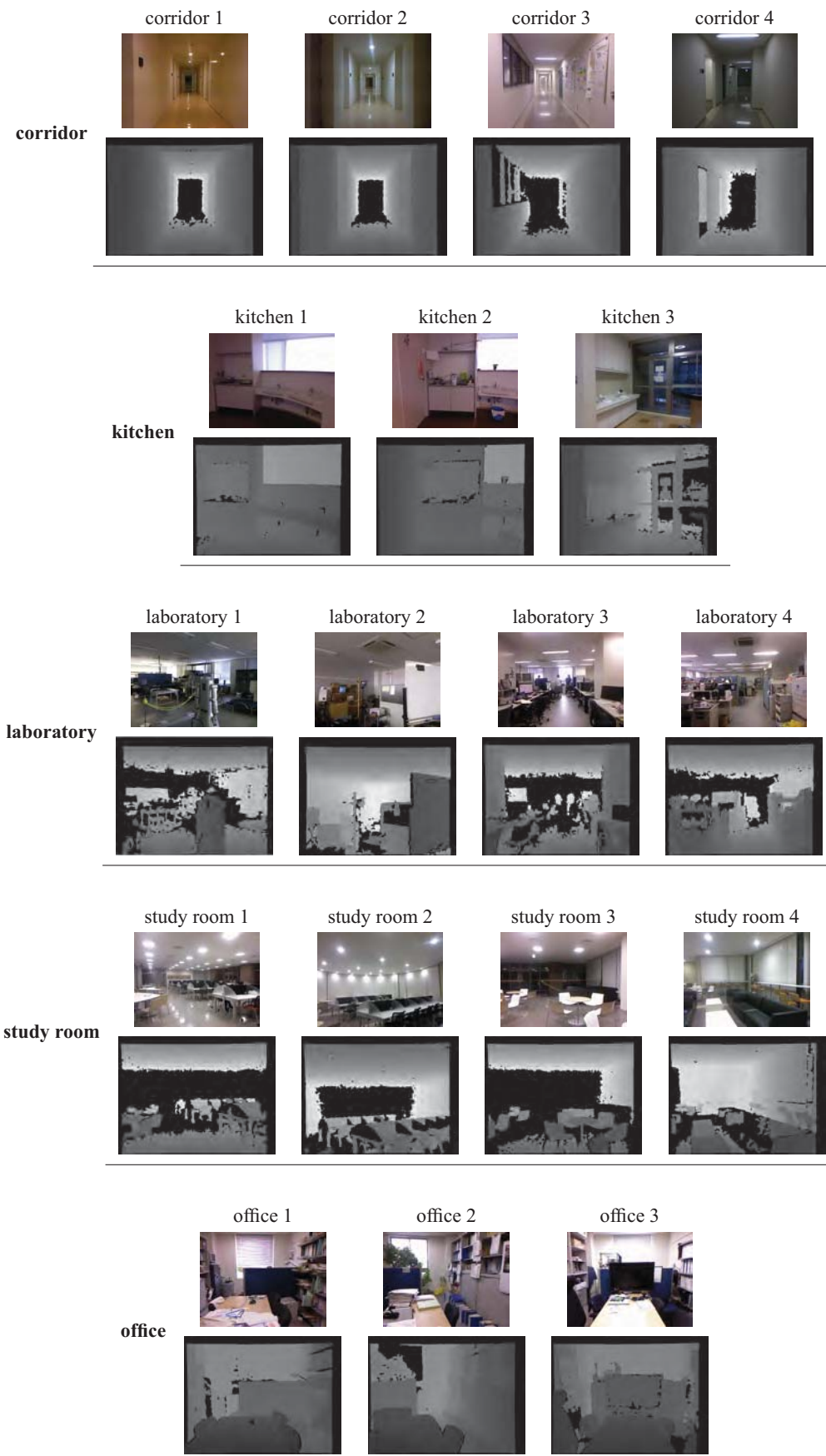


Fig. 4 Examples of different pairs of RGB and depth images for the five categories: corridor, kitchen, laboratory, study room, and office. We show one example for each place inside the category.