距離画像を用いたオフィス機器の探索
− Bag-of-keypoints と SIFT の利用−

稲浦 雄哉† 鈴木 昌人† 高橋 智一† 青柳 誠司†

†関西大学 システム理工学部 機械工学科 〒564-8680 大阪府吹田市山手町 3-3-35

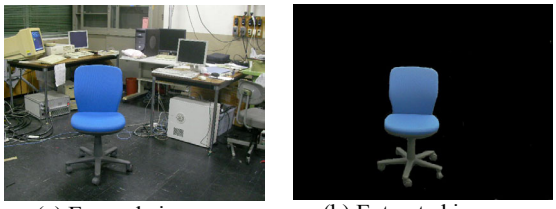
E-mail: ina19@iemec01.iecs.kansai-u.ac.jp, {m.suzuki, t.taka, aoyagi}@kansai-u.ac.jp

あらまし Time-of-Flight (TOF) カメラを使用して距離画像を取得し，室内のオフィス機器を探索する手法を提案する．オフィス機器の一例として椅子を使用し，さまざまな背景の中から椅子の領域抽出と物体認識を行う．まず TOF カメラで取得した距離画像をもとに距離が近い領域を抽出する．その後抽出された画像ごとに格子点状に SIFT を抽出して Bag-of-keypoints を行い，AdaBoost により椅子であるかどうかの 2 値分類機を作成する．ここでの SIFT とは濃淡画像で用いられる変換と同一の変換を濃淡画像に施したものである．距離に基づく領域抽出とそれへの SIFT 特徴量の適用にオリジナリティーがある．

キーワード 特定物体認識，TOF カメラ，距離画像，Bag-of-keypoints

1. 研究背景

提案手法では図 1(a)に例を示すような複雑な背景を有する室内において，Time-of-Flight (TOF) カメラを使用して距離画像を取得し，その情報をもとにオフィス機器の領域抽出と物体認識を行う．本研究ではオフィス機器の一例として椅子の認識を行う（**図 1(b)**に抽出された椅子の領域を示す）．椅子の認識には一般物体認識の分野で使用される Bag-of-keypoints を使用し AdaBoost により学習を行うことで椅子かどうかの判定を行う．将来的には一般的な椅子を認識することを目標とするが，本項ではその予備段階として，特定の椅子を認識すること（特定物体認識）を試みる．



(a) Example image (b) Extracted image

Fig.1 Target chair

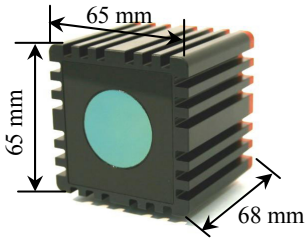


Fig.2 MESA SR-4000

Table 1 TOF camera spec

Pixel array	176(H)×144(V)
Viewing angle	69°(H)×56°(V)
Measuring range	0.3〜10 m
Frame rate	Max 50fps

2. 提案手法

2.1 距離画像における領域分割

本研究では距離情報を用いて各物体ごとのの領域を抽出してからその領域がどのような物体であるかの認識を行う．距離情報の取得には TOF カメラを使用する．TOF カメラは，アレイ状に並んだ多数の赤外 LED から周波数変調を施した赤外線を投光し，視野内の対象物から反射してきた変調信号の位相を計測することで，対象までの距離を計測することができる[1]．TOF カメラには MESA 社の SR-4000 を使用した[2]．図 2 にその外観を，表 1 にその仕様を示す．TOF カメラはステレオカメラに比べて照明の変化に強く（理由については後述する），リアルタイムに距離情報を取得できることが利点である．このカメラを 1 m の高さに，光軸が床面と平行になるように設置して，室内の距離画像の取得を解像度 176×144 pixel で行った．また室内の大きさの都合上 5 m 以内の距離データを使用する．図 3 に計測結果の一例を示す．

次に取得した距離情報を用いて物体間の領域を分割する．1 画像ごとに周囲 8 画素との間の距離データの差をとり，閾値 100 mm 以下の場合同一の物体とし，閾値以下だった画素にラベルを振る（ラベリング処理）[3]．この結果として，図 4 に示すような距離ごとに分割された画像を得ることができる．濃淡画像でのラベリングでは輝度値が似ている領域を抽出するのに対して，距離画像のラベリングでは距離が似ている領域を抽出することになる．輝度値を利用していないので，照明等の影響を受けずに領域の抽出を行えることが期待できる．

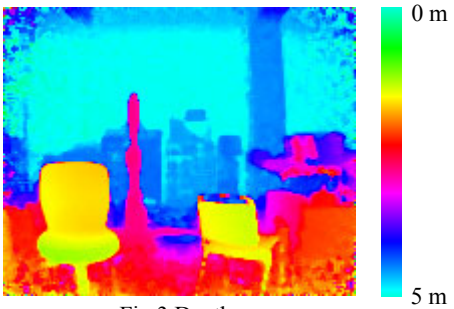


Fig.3 Depth map

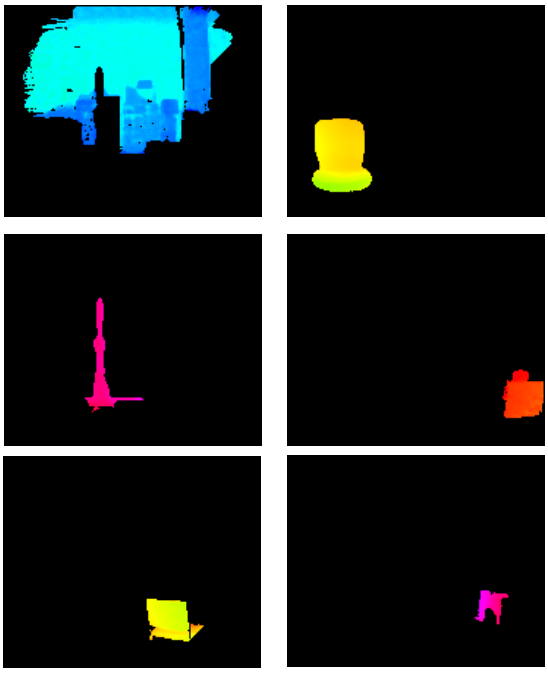


Fig.4 Extracted region (refer to color in Fig.3)

2.2 距離画像への SIFT の適用

距離ごとに分割された（抽出された）各画像が，探索対象である椅子であるかどうかの判別（認識）を行う．本稿ではその手法として Bag-of-keypoints を利用する[4]．この手法は Bag-of-words [5] の考え方を画像処理に適用したものである．Bag-of-words は，文書内の単語の語順を無視し，どのような単語が頻出するかを評価することで，その文書がどのようなカテゴリの文書であるのかを分類する手法である．この考え方を適用し，画像を局所特徴の集合と考えることで，様々な画像をカテゴリに分類することが可能となる．

Bag-of-keypoints で用いる特徴量として，SIFT（Scale-Invariant Feature Transform）[6] を使用する．SIFT は位置，スケール，方向の変化に頑健な局所特徴である．SIFT は，位置，スケール，方向，輝度勾配ヒストグラム（128 次元．以下特徴量ベクトルと呼称する）の 4 個の特徴量から成るが，Bag-of-keypoints では特徴量ベクトルのみを使用する．

本稿では，濃淡画像で用いられる SIFT と同一の変換を濃淡画像に施して，濃度勾配ヒストグラムの特徴量を得るところにオリジナリティーがある．なお，藤吉らは距離画像を用いて，HOG，距離ヒストグラム特徴量 [7, 8] を特徴量とした人物検出を行っている．

2.3 Bag-of-keypoints の具体的手順

まず TOF カメラで得られた距離画像から SIFT を抽出する．Dog 画像を用いて画像をぼかし，輝度勾配が大きい場所を探索して，SIFT keypoints を抽出することが一般的に行われる[6]．しかし，Bag-of-keypoints を用いた物体認識においては，格子状に強制的に抽出点を設け（この時点で位置の特徴量を無視していることになる），スケールのパラメータを適当に与え，そこでの輝度勾配ヒストグラムの特徴量を抽出することが有効であることが報告されている（なお，特徴量ベクトルの計算に方向，スケールのパラメータは必要である）[9]．本稿でもこれに従う．

なお，格子点が背景にあるときは，SIFT を抽出しない．格子点の間隔は 5pixel，スケールは 10，15，20，25 に変化させて SIFT 特徴量ベクトルの計算を行う．次に抽出された SIFT 特徴量ベクトルを k-means 法により k 個のクラスタにクラスタリングする．得られた k 個のクラスタの代表点（重心）を visual word と呼び，特徴量を量子化する際に使用する．図 5 にこの様子を模式的に示す．k の値は事前実験により経験的に，$k=100$ とした．クラスタの各分割画像から得られた SIFT 特徴量を visual words に基づきヒストグラム化を行う．このヒストグラムにはカテゴリ（探索物体の種別．例えば「椅子」とか「机」）ごとの特徴が現れ，あるカテゴリの画像とそれ以外の画像とを分類することが可能となる．

2.4 AdaBoost

visual words を用いた識別器の重みの学習には AdaBoost [10, 11] を使用する．本研究では 2 値分類判定に AdaBoost を用いる．AdaBoost とは boosting 学習と呼ばれる教師あり学習の一種であり，弱識別器と呼ばれる一つ一つはあまり高精度ではないが，50%以上の識別が行える識別器を多数組み合わせることにより，最終的に高い精度で識別を行うことが可能なアルゴリズムである．

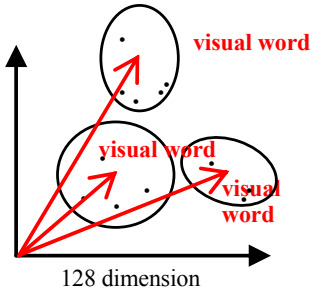


Fig. 5 Calculating visual words by k-means method

以下に学習・識別の流れを記載する．

1)まず m セットの学習データに対してラベル{0,1}を割り当てる（0:不正解，1:正解）．Bag-of-keypoints の場合，入力 x_i は各画像から得られた SIFT 特徴量を visual words に基づき量子化を行ったヒストグラムである．

$$S=\langle (x_1,y_1),...,(x_m,y_m)\rangle \tag{1}$$

$$y_i \in Y=\{0,1\} \tag{2}$$

2)各データに重要度 $D_t(i)$ を設定する．初期重要度は均等に設定する．

$$D_1(i)=1/m \quad \text{for all } i \quad (i=1\sim m) \tag{3}$$

3)弱識別器を選択．

以下の処理を $t=1,2,...,T$ （T は識別器の総数．本研究では $T=300$）について行う：

重要度 $D_t(i)$ に基づき，式(4)に示す誤り率 ε_t を設定し，最も誤り率の低くなる弱識別器 $h_t()$を求める．ここで $h_t()$ は図 6 に示すようにあるビンに着目し，その値と設定した閾値を評価し，1 か 0 を出力するものである．一つ弱識別器を選択するごとに重要度が更新されるため，あるビンが何度も弱識別器の評価に用いられることや，弱識別器の評価に全く用いられないビンがある可能性もある．

$$\varepsilon_t=\sum_{i=0}^{m}\begin{cases}0 & \text{if } h_t(x_i)=y_i\\ D_t(i) & \text{otherwise}\end{cases} \tag{4}$$

誤ったデータの重要度が大きくなるように $D_t(i)$ を更新する．

$$\beta_t=\varepsilon_t/(1-\varepsilon_t) \tag{5}$$

$$D_{t+1}(i)=\frac{D_t(i)}{Z_t}\times\begin{cases}\beta_t & \text{if } h_t(x_i)=y_i\\ 1 & \text{otherwise}\end{cases} \tag{6}$$

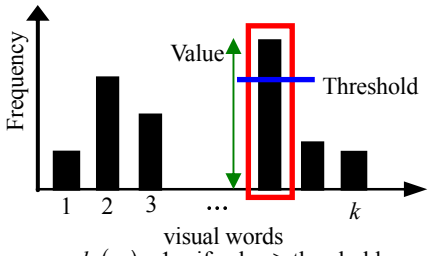
ここで，Z_t は重要度の総和を 1 とするための正規化定数である．

4)最終的な識別は式(8)に示すように全弱識別器を連ねたものとなる．α_t は弱識別器 h_t の重みであり，β_t の逆数であるので，正しく識別が行えた場合に大きくなる．

$$\alpha_t=1/\beta_t \tag{7}$$

$$h_{fin}(x)=\begin{cases}1 & \text{if } \sum_{t=1}^{T}h_t(x)\alpha_t\geq\frac{1}{2}\sum_{t=1}^{T}\alpha_t\\ 0 & \text{otherwise}\end{cases} \tag{8}$$

1 と判定した弱識別器の重みの総和が，全弱識別器の重みの総和の半分以上であれば最終的な判定は 1（正解）となる．



$h_t(x_i)=1$ if value ≥ threshold

$h_t(x_i)=0$ if value < threshold

Fig.6 Histogram for an image based on the appearance frequency of visual words

3. 評価実験

図 7 に実験のフローチャートを示す．まず椅子の領域の学習を行う．学習距離画像には，図 8(a)に示すような椅子だけ存在している領域の画像を正解画像として使用し，図 8(b)に示すようなその他の背景画像を不正解画像として使用する．学習距離画像の枚数は，正解画像 300 枚，不正解画像 500 枚とする．この画像を使用して，Bag-of-keypoints で visual words を作成し，正解画像にラベル 1 を振り，不正解画像にラベル 0 を振り，AdaBoost により学習を行う．

学習が済んだ識別器を用いて，未知画像に対する評価実験を行った．評価実験には図 9 に示すような様々な場所（背景）で取得した，学習に使用していない距離画像を使用する．評価画像は椅子が写っている画像 200 枚と椅子が写っていない画像 100 枚，合計 300 枚の距離画像を使用する．この画像をまず距離ごとの物体に分割して（領域抽出），学習済みの AdaBoost 識別器を用いて，抽出された領域画像が椅子であるの判定を行う．

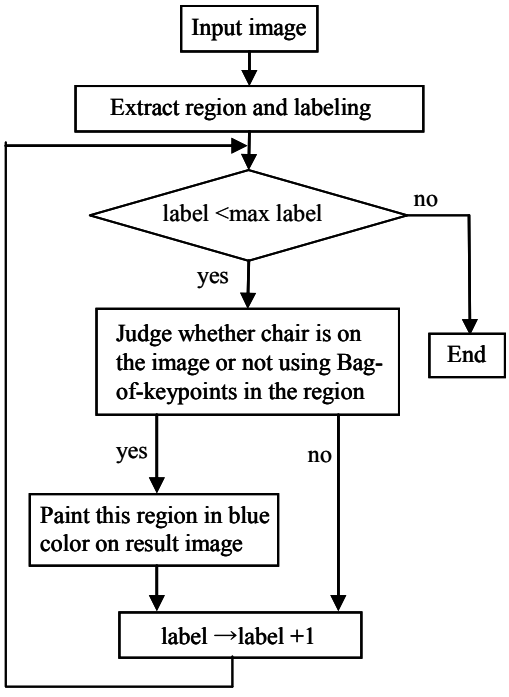
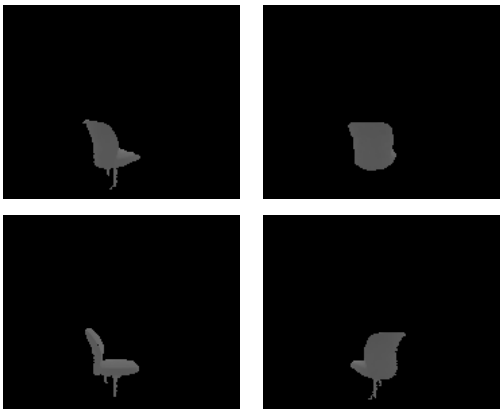


Fig.7 Flow chart



(a)Positive image

(b)Negative image

Fig.8 Depth map for learning AdaBoost classifier

Fig.9 An example of image for verification

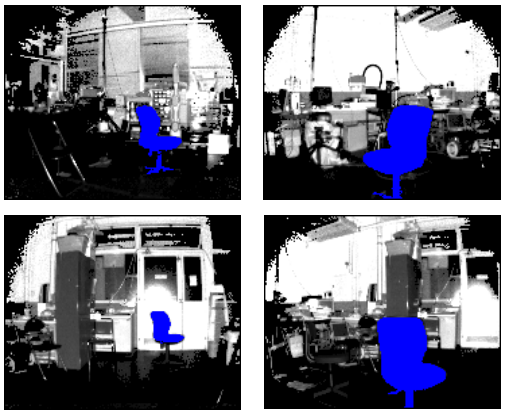
4. 実験結果および考察

椅子の領域を椅子，椅子でない領域を椅子でないと判定した場合を正解とする．また，椅子の領域を椅子でない，椅子でない領域を椅子であると判定した場合を不正解とする．正解，不正解の判定は人間が手動で行う．この判定例を図 10 に示す．図中青く塗りつぶされた部分が椅子であると判定された領域であり，図 10(a)の例では正解，図 10(b)の例では不正解となる．なおこの図の表示において青く塗りつぶされていない領域に関しては，TOF カメラで取得される反射強度情報を表示している．

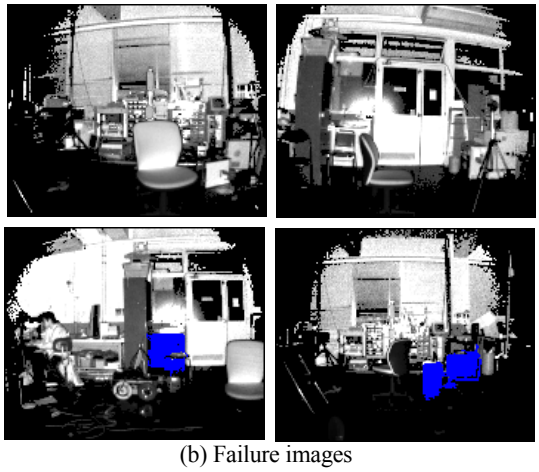
椅子の認識判定結果を表 2 に示す．本稿では特定の椅子の認識を試みたにもかかわらず，認識率（抽出も含めた認識率である）は 124 / 200=62%と低い値である．

原因として，距離情報をもとに認識を行っているため，対象物体の姿勢の変化があると距離情報に変化があり，対応することが難しいためであると考える．また領域を分割する際に対象物体の近くに他の物体が置いてあると，対象物以外の物体も同じ物体としてラベリングされてしまい，正確に領域を抽出することが出来ない．

今後，1) 学習画像の数を増やすことで，物体の姿勢の変化にある程度ロバストに対応できる識別器を構築する，2) TOF カメラが有しているピクセル毎の反射強度情報（例を図 11 に示す）を利用する，3) 他の CCD カメラ等を用いて濃淡画像を取得し，これから得られる情報を併用する，等の工夫を行い，認識率の向上を目指したい．



(a) Success images



(b) Failure images

Fig.10 Example results

Table 2 Recognition result

	Correct image	Wrong image
Number	200	100
Correct	124	25
Wrong	76	75
Recognition rate	62%	75%

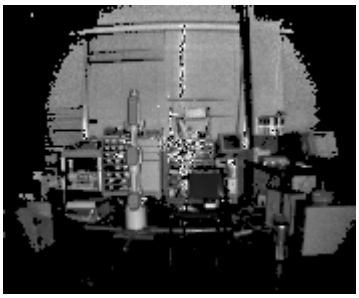


Fig.11 Example of amplitude image supplied by TOF camera

5. 結言

TOF カメラを使用して距離画像を取得し，室内のオフィス機器を探索する手法を提案した．オフィス機器の一例として椅子を使用し，さまざまな背景の中から椅子の領域抽出と物体認識を行った．具体的には，距離画像をもとに距離が近い領域を抽出し，抽出された画像ごとに SIFT を抽出して Bag-of-keypoints を行い，AdaBoost 学習により椅子であるかどうかの 2 値分類機を作成する．ここでの SIFT とは濃淡画像で用いられる変換と同一の変換を濃淡画像に施したものである．特定物体認識の実験を行った結果，認識率は 62%であった．今後この改善をはかりたい．

文　献

[1] T. Oggier, M. Lehmann, R. Kaufmann, M. Schweizer, M. Richter, P. Metzler, G. Lang, F. Lustenberger, and N. Blanc, “An all-solid-state optical range camera for 3D real-time imaging with sub-centimeter depth resolution (SwissRanger),” in Proc. SPIE, vol. 5249, pp.534–545, 2004.

[2] http://www.mesa-imaging.ch/.

[3] 森 涼子, 長坂 保典, 鈴村 宣夫, “ステレオ画像を用いた障害物の抽出と位置推定”, 電子情報通信学会技術研究報告.IE, 画像工学 vol.97 pp.27-32, 1997.

[4] G. Csurka, C. R. Dance, L. Fan, J. Willamowski and C. Bray, “Visual Categorization with Bags of Keypoints,” in Proc. of ECCV Workshop on Statistical Learning in Computer Vision, pp. 59-74, 2004.

[5] C. D. Manning and H. Schtze, “Foundations of Statistical Natural Language Processing,” The MIT Press 1999.

[6] D. G. Lowe, “Object Recognition from Local Scale-Invariant Features,” in Proc. ICCV, pp.1150-1157, 1999.

[7] 藤吉弘亘, 山内悠嗣, 三井相和, 池村翔, 山下隆義, “複数の特徴量間の関連性に着目した Joint-HOG による物体検出,” 電気学会, pp.51-56, 2008.

[8] 池村 翔，藤吉弘亘, “距離情報に基づく局所特徴量によるリアルタイム人検出,” 第 15 回 画像センシングシンポジウム SSII09, IS4-05, 2009.

[9] F. F. Li and P. Perona: “A Bayesian Hierarchical Model for Learning Natural Scene Categories,” in Proc. Computer Vision and Pattern Recognition (CVPR), Vol. 2, pp. 524-53, 2005.

[10] Y. Freund and R. E. Schapire, “Experiments with a new boosting algorithm,” in Proc. of the 13th Intl. Conf. Machine Learning, pp. 148-156, 1996.

[11] 小濱篤，岩本翔太，鈴木昌人，青柳誠司, “ロボットビジョンによる室内環境の認識-SIFT 特徴量とエッジ情報を用いたオフィス機器の画像探索と認識-,” ロボティクス・メカトロニクス講演会 2010　講演論文集，2P1-E19, 2010.