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—

Recognition of Room Environment by Robot Vision

- Image Detection and Recognition of Office Apparatus Using SIFT and Edge Information-

小浜 篤 (関西大) ○学 岩本 翔太 (関西大)
鈴木 昌人 (関西大) 正 青柳 誠司 (関西大)

Atsushi KOHAMA, Kansai University
Shota IWAMOTO, Kansai University
Masato SUZUKI, Kansai University
Seiji AOYAGI, Kansai University, aoyagi@kansai-u.ac.jp

For the purpose of simultaneous localization and mapping (SLAM) of an indoor mobile robot, a method to process only one monocular image (640×480) of the whole view of environment is proposed. This method imitates that a human can grasp momentarily the situation of a room, i.e., the layout of objects/obstacles. Specific object recognition for desks with several view angles are dealt with. The proposed method is concretely organized as follows: 1) Bag of Keypoints is applied, then the existence of object on input image is judged, 2) If object existence is verified, its angle is detected by Bag of Keypoints, 3) The candidates for the projection from template image to input image is obtained using Scale Invariant Feature Transform (SIFT) and edge information, the validity of which is decided by AdaBoost classifier based on various image features, such as Hu moment, Haar-like features, etc.

Key Words: Specific object recognition, Bag of Keypoints, SIFT, AdaBoost classifier

1. 研究背景

ロボットが環境内を移動するためには環境地図の獲得と自己位置の同定が必要である。しかし、自己位置を特定するためには環境地図が必要であり、また環境地図を作成するためには自己位置が定まっていなければならないという問題が同時に生じる。この問題は Simultaneous Localization and Mapping(SLAM)と呼ばれ、従来 Laser Range Finder(LRF)を用いて、未知環境内を走行しながら自己位置特定と環境地図作成を同時に行う手法が提案されている。しかしながら、未知環境内を走行しなければならない危険性や環境地図の獲得までに時間がかかるという問題がある。そこで本研究では、人間が一瞥しただけで室内の環境を認識できるように、視覚センサによる静止画を用いて、室内の環境を認識することを目的とする。本研究ではオフィス機器の一例として特定の机を室内画像から探索する問題（特定物体認識）を扱う。様々な角度から撮影されたテンプレート画像を室内画像から探索する。

本研究で提案する探索手法を以下に記す。1)Bag of Keypointsを用いて画像中に探索物体が写っているかどうかを判定し、どの角度で写っているのか特定を行う。2)Scale Invariant Feature Transform(SIFT)とエッジ情報を利用することで、テンプレート画像を探索画像中に射影する。3)様々な特徴量に基づく弱識別器を用いて AdaBoost による判定を行い、射影領域が対象物体に対応しているかの判定を行う。

2. Bag of Keypoints

Bag of KeypointsとはBag of Words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。この考え方を画像処理に適用することにより、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ではSIFT特徴量を利用する。SIFTとは画像から位置、スケール、方向の変化に頑健な特徴量を局所領域から抽出する処理である[1]。Bag of Keypointsでは画像から抽出されたSIFT特徴点の位置、スケール、方向を無視し、特徴

Fig. 1 Extracted SIFT

Fig. 2 Calculating visual words by k-means

Fig. 3 Histogram of an Image calculating

量のみを用いて、識別器（本研究ではAdaBoost、説明は次章参照）により画像のカテゴリ分類を行うことができる[2]。

まずSIFT特徴量を全学習画像から抽出する（図1）に例を示す。出典：http://www.vision.caltech.edu/Image_Datasets/Caltech101。次に抽出されたSIFT特徴量(128次元)をk-means法によりクラスタリングする（図2）。これにより得られるk個のクラスタの代表点（128次元）をvisual wordsと呼び、特徴量を量子化する際に使用する。次に各画像においてSIFT特徴量を抽出し、先ほど求めたvisual wordsに基づきヒストグラム化を行う（図3）。得られたヒストグラムを用いて識別器の学習を行う。このヒストグラムにはカテゴリごとの特徴が現れ、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。学習の後、未知画像を入力し、学習時と同様に、特徴量を抽出しvisual wordsに基づきヒストグラム作成、2値分類器（AdaBoost, SVMなど、本研究では前者）によりヒストグラ

ムを評価することで、その画像がどのカテゴリに属するものであるかを判断することができる。

Bag of Keypointsは一般物体認識問題における研究で脚光を浴びている手法であるが、本研究では特定の物体が写っているかどうかを判定する特定物体認識にBag of Keypointsを利用する。

3. AdaBoost

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[3]。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。

$$S = \langle (x_1, y_1), \dots, (x_m, y_m) \rangle \quad (1)$$

$$y_i \in Y = \{0, 1\} \quad (2)$$

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

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

3)弱識別器を選択。

以下の処理を $t=1, 2, \dots, T$ (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4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。

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

誤ったデータの重要度が大きくなるように D_t を更新する。

$$\beta_t = \varepsilon_t / (1 - \varepsilon_t) \quad (5)$$

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

ここで、 Z_t は重要度の総和を1とする正規化定数である。

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。

$$\alpha_t = 1 / \beta_t \quad (7)$$

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

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。

本研究ではBag of Keypointsを用いて机が写っているか否かを判定するプロセス、机が写っているとしたら、どのような角度から見た机が写っているのかを判定するプロセス、また重みの大きなヒストグラムのビンに対応するSIFT特徴量を用いて、テンプレート画像全体を探索画像内に射影して、その射影された領域に机が写っているかどうかを様々な画像特徴量を用いて判定するプロセス（これにより机が探索画像内のどの位置に写っているかが判定できる）、の3つのプロセスにおいてAdaBoostの手法を用いる。

Fig. 4 Example of weak classifier

(a) Object in the image

(b) No object in the image

Fig. 5 Scenery Image

(a) 0 deg

(b) 30 deg

(c) 60 deg

(d) 90 deg

Fig. 6 Template Images

Fig. 7 Set grid points

4. 提案手法

4.1 机が写っているかの判定

図5に示すような室内画像から図6に示すような机を探索する。探索画像、テンプレート画像のサイズは640×480 pixelである。学習用に机が写っている画像1,200枚（4方向から撮影されたものが300枚ずつ）、机が写っていない画像1,000枚、学習用とは別に検証用にも同じ枚数の画像（1,200+1,000=2,200枚）を用意した。すなわち、合計4,400枚の画像を用意した。

まず、探索画像中からSIFT特徴量を抽出するのだが、机のようにテクスチャの少ない物体からはSIFT特徴点が抽出されにくいので、本研究では格子状に特徴を抽出する点を設定する。具体的には図7に示すような20 pixel間隔の格子を設定した。また通常ではSIFT Keypointの抽出のために画像を平滑化し、DoG画像の作成を行うが（強度の異なるぼかした画像を利用して位置、スケールを検出する）、本研究ではKeypointを格子状に設定するため、平滑化の必要は無いので、入力画像からそのまま特徴量の抽出を行う。

学習画像から格子状に特徴量を抽出し、2章で述べたvisual wordsを作成する。クラス数は $k=100$ として実験を行った。visual wordsに基づき、各画像のヒストグラム化を行い、机が写っている画像には1、写っていない画像には0とラベルを振り、AdaBoostによる学習を行う（学習にはwekaを用いた。http://www.cs.waikato.ac.nz/~ml/weka）。これにより机が写っているかどうか判定することができる識別器を作成する。

次に、机が写っていることが判ったとして、どのような角度から見た机が写っているのかを判定するプロセスを行う。机が写っている画像のみを用いて先ほどと同様にBag of Keypointsを行う。用意したテンプレート画像の角度の種類だけ識別器を作成する。全学習画像からSIFT特徴量を格子状に抽出し、k-means方によりクラスタリングを行うことでvisual wordsを作成する。各画像から得られるSIFT特徴量をvisual wordsに基づきヒストグラム化する。ここである特定の角度の画像から得られたヒストグラムにはラベル1を割り振り、特定の角度以外の画像から得られたヒストグラムにはラベル0を割り振る。このようにラベルを振りAdaBoostの学習を行うことにより、特定の角度から撮影された机とそれ以外の角度から撮影された机との分類を行うことが可能となる。

ここまでの処理で、入力画像中に机が写っているかどうか、またどの角度から撮影された机が写っているのかを特定することが可能となる。

4.2 テンプレートの射影

机の位置を特定するために射影行列の計算を行う。Bag of Keypointsにより机が写っているか判定を行う際に重みの大きいヒストグラムのビンに対応する特徴量を描画すると図8のようになる。青で描かれた円が特徴量の抽出点を示す。これより、探索物体付近に特徴量が現れることが判る。このことを利用して、テンプレート画像から取得される重みの大きな特徴量と、探索画像中の重みの大きな特徴量との射影関係を求め、テンプレート画像全体を探索画像中に射影する。射影された領域に机が写っているかどうかを様々な画像特徴量を用いて判定する。以下に具体的な手順を述べる。

撮影時のカメラパラメータは既知としているため、カメラ座標 (u, v) からワールド座標 (x, y, z) に射影する際、 $z=0$ という制約を設けることで、 (u, v) から $(x, y, 0)$ への射影関係（並進のみ）を計算できる。テンプレートの脚の部分($z=0$)の射影を行い、実際の脚の間隔と射影により得られた脚の間隔とを比較することにより、射影した後の画像を特徴点を中心として拡大縮小する倍率を決定する。

しかし、重みの大きいヒストグラムに属する特徴量が現れないことや、それのみでは正しい射影関係を満たさない画像も存在する。そこで、本研究では上記射影がうまくいかなかった際に、エッジ情報を利用した射影関係を用いる。探索画像からエッジを求め、そこからハフ変換を用いて直線部を抽出したものを図9に示す。テンプレート画像からも直線部を抽出し、線分の端点同士を射影させていき、その中から正しい射影関係を特定する。端点同士の射影を行う際、図10に示

(a) Scenery image

(b) Template image

Fig. 8 Describe SIFT that belong to weak classifier with heavy weight

(a) Scenery image

(b) Template image

Fig. 9 Extract the straight line with edge

(a) Template image

(b) Scenery image

Fig. 10 Select points that use projection

Fig. 11 Flow chart

すように候補となる直線の端点とペアとなるべき端点までの直線と水平線とのなす角度を計算し、角度の近いもの同士の射影を行う。これにより誤った射影を行う可能性を低くする。

4.3 射影の正否判定（机が探索画像内のどこに写っているかの判定）

人間は物体を探索、認識する際、色や形状など様々な情報を用いていると考えられる。そこで位置特定を行う際の射影の正否判定には様々な特徴量を用いたAdaBoostによる識別を採用する。本研究で用いた特徴量はhuモーメント不変量、位

相限定相関法に基づく相関強度画像の最大輝度, Haar-like特徴量 (予め設定した白と黒の領域の輝度値の平均値の差), 濃淡テンプレートマッチングの類似度, 部分テンプレートマッチングの類似度, エッジ画素の重なり具合, ハフ変換で抽出される直線部の重なり具合を用いる。これらの特徴量による全類似度 (注: 類似度は特徴量によって異なり一個のものもあれば数十個のものもある。Haar-Like特徴量と部分テンプレートマッチングからは多くの類似度が取得される。) を入力としてAdaBoostの学習を行う。このときの弱識別器はある類似度を参照し, その値と閾値との大小関係を比較し, 1 か0を出力するものである。

一つ一つの射影から得られる特徴量に対し, 射影が正しく行われているものと, そうでないものを人が判断してラベルを振り, AdaBoostによる学習を行っておく。このように学習することで, 射影が正しい場合とそうでない場合を判断することが可能となる。

以上述べてきた操作を行うことで, 未知の画像を入力した際に, 探索物体が写っているのかの判定, 物体の角度の特定, 物体が写っている位置の特定を行うことが可能となる。学習結果を用いて未知画像から探索を行う際の処理のフローチャートを図11に示す。なお, 射影については, AdaBoost識別器が最初に1 (正) を出力した射影を採用することにする。

5. 実験結果

机の有無判定を行う際のBag of KeypointsにおけるAdaBoost学習の結果を表1に示す。机の角度特定を行う際のBag of KeypointsにおけるAdaBoost学習の結果を表2に示す。机の位置特定を行う際, 射影の正否判定を行うための様々な特徴量を用いたAdaBoostによる学習結果を表3に示す。いずれも高い精度で学習データを分類できる識別器が作成出来ている。

検証用の未知画像2,200枚を入力した結果を表4に示す。机の位置特定結果の例を図12に示す。白いラインはテンプレート画像におけるエッジを位置特定結果に基づき探索画像に射影したものである。机が写っているかどうか判定に関しては, $(277+278+274+285+978)/2,200 \times 100 = 95\%$ の精度で成功した。机の正しい位置の特定, 机の写っていない画像の識別に関しては, $(248+238+242+266+978)/2,200 \times 100 = 90\%$ の精度で成功した。机が写っている場合の探索平均時間は約11sec, 机が写っていない場合のそれは約2secであった (CPU: Intel® Core™2 Duo 2.0GHz, MEMORY: 2.0GB)。

誤差の要因を表1の右側に掲載しているが, 位置ズレが主たる要因である。位置ズレに関しては学習時も検証時も人が判定を行っているため, 正否の判定が微妙である。

6. 結論

静止画像情報のみを用いて, Bag of Keypoints, AdaBoostの手法を援用し, 特定物体が写っているか否かの判定, 写っている場合の物体の角度の特定, 画像内における物体位置の特定を行うことに成功した。

今後, 机以外に椅子など他のオフィス機器に関しても実験を行うこと, また, 机であれば「ある高さに矩形の領域があり, 脚がある」等の物体に関する概念を用いることで[4], 一般物体認識を行えるようにすることが課題として挙げられる。

文 献

- [1] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", International Journal of Computer Vision, vol. 60, pp. 91-110, 2004.
- [2] 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.

- [3] Y. Freund and R. E. Schapire, "Experiments with a New Boosting Algorithm", in Proc. of the 13th International Conference on Machine Learning, pp. 148-156, 1996.
- [4] 矢口裕明, 岡田慧, 稲葉雅幸, "頭部装着型マルチセンサによる生活環境モデリングの研究", 第27回日本ロボット学会学術講演会講演論文集, 2I2-03, 2009.

Table 1 Decision by AdaBoost classifier for detecting the existence of desk

	Correct	Wrong
Correct data	1200	0
Wrong data	1	999

100.0%

Table 2 Decision by AdaBoost classifier for specifying desk angle

(a) Desk 0[deg]			(b) Desk 30[deg]		
	Correct	Wrong		Correct	Wrong
Correct data	300	0	Correct data	298	2
Wrong data	0	900	Wrong data	0	900
100.0%			99.8%		
(c) Desk 60[deg]			(d) Desk 90[deg]		
	Correct	Wrong		Correct	Wrong
Correct data	300	0	Correct data	296	4
Wrong data	0	900	Wrong data	0	900
100.0%			99.7%		

Table 3 Decision by AdaBoost classifier for detecting desk position

(a) Desk 0[deg]			(b) Desk 30[deg]		
	Correct	Wrong		Correct	Wrong
Correct data	1206	99	Correct data	564	74
Wrong data	123	2381	Wrong data	60	1297
94.2%			93.3%		
(c) Desk 60[deg]			(d) Desk 90[deg]		
	Correct	Wrong		Correct	Wrong
Correct data	1314	28	Correct data	1140	83
Wrong data	45	763	Wrong data	88	2016
96.6%			94.9%		

Table 4 Results of detection

	Input data	Correct data (Desk existence)	Detect desk position	Out of alignment	Desk angle failure
0 [deg]	300	277	248	23	6
30 [deg]	300	278	238	10	30
60 [deg]	300	274	242	31	1
90 [deg]	300	285	266	10	9
no desk	1000	978

Fig. 12 Result Image