

修士論文概要書

Summary of Master's Thesis

Date of submission: 1/31/2012 (MM/DD/YYYY)

専攻名（専門分野） Department	情報理工学専攻	氏 名 Name	藤川哲也	指 導 教 員 Advisor	甲藤二郎教授 印 Seal
研究指導名 Research guidance	画像情報研究	学籍番号 Student ID number	CD 5110B108-5		
研究題目 Title	物体抽出を用いた一般物体認識の性能改善				

1. まえがき

近年、ハードディスクなどの記録媒体の大容量化に伴い、大量のデジタル画像を蓄積できるようになったが、未だに画像の検索や分類に人での介入が不可欠となっている。そこで、画像を一般的なカテゴリとして認識する一般物体認識の要求が高まっている。ここでいう「一般的なカテゴリ」とは、「セダン」や「ワゴン」といった形状の異なる細かいカテゴリのことではなく、形状が違えどより一般的な概念である「車」というカテゴリのことを指す。

2. 関連技術

2.1 Bag-of-Keypoints[1]

Bag-of-Keypoints(BoK)は画像を局所特徴(keypoint)の集合として捉え、k-means 法を用いたベクトル量子化によって一つのヒストグラムとして表現する手法であり、近年多くの一般物体認識に関する研究で用いられている。多くの場合、局所特徴には SIFT 特徴[2]が用いられる。

2.2 Saliency Map[3]

視覚注意を表現する Saliency Map としては複数のモデルが提案されているが、本研究では視覚的注意の向きやすい「物体」を表現できる T. Liu らのモデル[2]を用いて物体抽出を行う。

2.3 Multiple Kernel Learning

Multiple Kernel Learning(MKL)は、複数のカーネルを線形結合することで得られるカーネル関数を用いた学習法であり、各カーネルの最適な重みを学習できる。Varma ら[4]は複数の画像特徴の最適な重みを計算し、それらの特徴を統合して認識させることで高精度な画像認識を行った。本研究では、Varma らの手法と同様に、カテゴリごとに各特徴の最適な重みを、MKL を用いて求めて統合し、分類を行う。

3. 提案手法

物体抽出を用いた一般物体認識手法はすでいくつか提案されている。しかし、これらの手法では物体抽出を行うための学習が必要であり、さらに学習のために各カテゴリのグランドツルース(正解画像)を準備しなければならないという問題点も抱えていた。また、これらの手法は前景の特徴のみを用いる手法であったが、本研究で以前画像の前景領域の特徴のみを用いることで認識精度の改善が可能か検討を行ったところ、背景の特徴が認識に有効な場合があることがわかった[5]。そこで、本手法では学習を伴わない物体抽出手法を用いて画像を前景と背景に分離し、各領域で得られた画像特徴を統合して認識を行うことで精度改善を試みる。具体的な提案手法の流れを Figure 1, Figure 2 に示す。Figure 2 中の赤い枠の部分が従来の BoK との違いである。提案手法では認識対象の物体は視覚的注意が向きやすいと仮定し、入力画像から Saliency Map を生成する。各入力画像の隣接画素との画素値の差とその Saliency Map の値から Graph Cuts[6]ベースの物体抽出を行い、得られた領域に対してバウンディングボックスを構築することで ROI を定義し、ROI 内を前景、ROI 外を背景とする。それぞれの領域から得られた画像特徴を MKL-SVM を用いて重み付け統合することで認識を行う流れとなっている。本研究では画像特徴として SIFT(BoK 表現)と色特徴をそれぞれの領域から抽出する。色特徴は各領域を空間的に 4 分割し、また RGB のそれぞれのチャンネルも 4 分割してヒストグラム化することで、計 $4 \times 4 \times 4 \times 4 = 256$ 次元の特徴ベクトルとして抽出する。

あることがわかった[5]。そこで、本手法では学習を伴わない物体抽出手法を用いて画像を前景と背景に分離し、各領域で得られた画像特徴を統合して認識を行うことで精度改善を試みる。具体的な提案手法の流れを Figure 1, Figure 2 に示す。Figure 2 中の赤い枠の部分が従来の BoK との違いである。提案手法では認識対象の物体は視覚的注意が向きやすいと仮定し、入力画像から Saliency Map を生成する。各入力画像の隣接画素との画素値の差とその Saliency Map の値から Graph Cuts[6]ベースの物体抽出を行い、得られた領域に対してバウンディングボックスを構築することで ROI を定義し、ROI 内を前景、ROI 外を背景とする。それぞれの領域から得られた画像特徴を MKL-SVM を用いて重み付け統合することで認識を行う流れとなっている。本研究では画像特徴として SIFT(BoK 表現)と色特徴をそれぞれの領域から抽出する。色特徴は各領域を空間的に 4 分割し、また RGB のそれぞれのチャンネルも 4 分割してヒストグラム化することで、計 $4 \times 4 \times 4 \times 4 = 256$ 次元の特徴ベクトルとして抽出する。

4. 評価実験

4.1 実験概要

本節では、提案手法の有効性評価のため、次の 6 つの手法について比較を行う。

①従来手法 1(物体抽出なし, BoK のみ)

②従来手法 2(物体抽出なし, BoK+Color)

③提案手法 1(物体抽出あり, 前景 BoK+背景 BoK)

④提案手法 2(物体抽出あり, 前景 BoK+背景 BoK+前景 Color+背景 Color)

⑤比較手法(グランドツルース使用, 前景 BoK+背景 BoK)

⑥比較手法(グランドツルース使用, 前景 BoK+背景 BoK+前景 Color+背景 Color)

グランドツルースは物体領域を境界に沿って抽出した正解画像であり、[7]からダウンロードできる。

4.2 実験環境

Table 1 に示す Caltech-256[8]からランダムに選んだ 10 カテゴリを用いて実験を行う。学習画像には各 50 枚, テスト画像には各 40 枚を用いた。BoK の Visual Word 数は 100 から 1000 まで 100 刻みで実験を行い、もっとも認識率の高いものをその手法の Visual Word 数とした。評価指標には Accuracy(正しく認識された

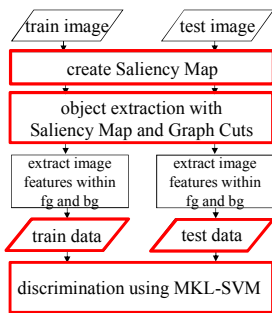


Figure 1: Flowchart of Proposal.
(fg means foreground, and bg means background)

Table 1: 10 categories used in experiment.

ID	Category Name
1	ak47
2	backpack
3	baseball-glove
4	basketball-hoop
5	bathtub
6	beer-mug
7	binoculars
8	cartman
9	cd
10	centipede

画像数/全体の画像数)を用いる。

4.3 結果と考察

実験結果を Table 2, Figure2, Figure 3 に示す。Figure2 は提案手法による物体抽出結果である。Table 2 に示す 10 カテゴリの平均認識率では、BoK のみを画像特徴として用いた場合は 8%, BoK と色特徴を用いた場合には 7.25%の精度向上が確認できた。また、比較手法が提案手法よりも 2~4%精度が高いことから、提案手法の物体抽出の精度を向上できれば、さらに数%程度の精度向上が見込めることがわかった。カテゴリごとの認識率では、5 番の bathtub のように設置場所がある程度決まっており、前景と背景の共起関係が強いカテゴリでは、提案手法によって認識率が大きく改善されていることがわかる。また、一方で、6 番、9 番などどの手法でも認識率の低いカテゴリも存在している。Figure 2 に示した通り、9 番(CD)の画像の自動物体抽出結果の多くは、Saliency Map の特性により縁の部分が背景領域に出てしまっており、また各特徴の重みは {0.21444, 0.69336, 0.08463, 0.00757} となっていることから、CD の縁の部分が認識に重要であったと考えられ、このような形状が重要なカテゴリは Appearance を表現する BoK や色特徴での特徴表現よりも Shape を表現する特徴などが適切であったと考えられる。

5. むすび

本稿では、Saliency Map を用いた自動物体抽出手法と MKL の組み合わせによる一般物体認識の手法を提案し、従来手法と比べ 7%以上の精度向上を確認した。

今後の展望としては、形状を表現する画像特徴など、その他の画像特徴とも組み合わせることで、さらに精度を向上させられると思われる。また、自動物体抽出の精度改善が認識精度の向上につながると考えられるため、物体抽出の精度向上に

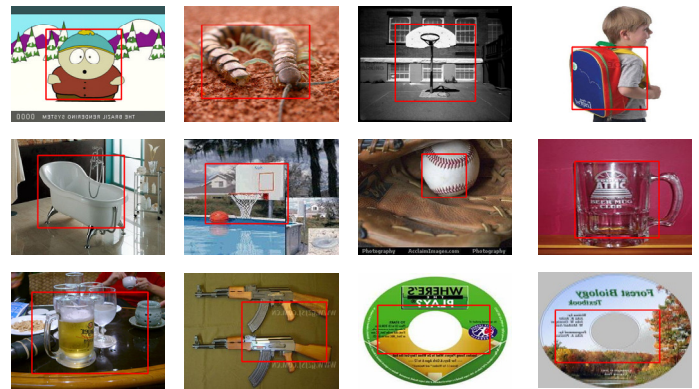


Figure 3: Results of automatic object extraction.

Table 2: Conclusion of experiment.

Method	Conventional1	Conventional2	Proposal1	Proposal2	Groundtruth1	Groundtruth2
Accuracy	0.6475	0.6875	0.7275	0.76	0.7475	0.7975

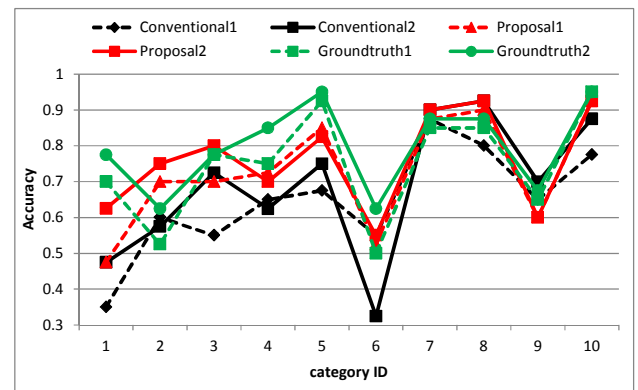


Figure 3: Accuracy of 6 methods for 10 categories.

も取り組む必要がある。

文 献

- [1] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual Categorization with Bags of Keypoints," ECCV Workshop on Statistical Learning in Computer Vision, Prague, Czech, pp. 1-22, 2004.
- [2] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," International Journal of Computer Vision, 2004.
- [3] T. Liu, J. Sun, N.-N. Zheng, X. Tang, and H.-Y. Shum, "Learning to Detect A Salient Object," IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, USA, 2007.
- [4] M. Varma, and D. Ray, "Learning The Discriminative Power-Invariance Trade-Off," IEEE International Conference on In Computer Vision, Rio de Janeiro, Brazil, October 2007.
- [5] 藤川哲也, 甲藤二郎, "物体領域抽出による画像の識別率向上に関する検討," 信学技報, IE2010-70, Sep.2010.
- [6] Y. Y. Boykov, and M. P. Jolly, "Interactive Graph Cuts for Optimal Boundary & Region Segmentation of Objects in N-D Images," International Conference on Computer Vision, Vancouver, Canada, Vol.I, pp. 105-112, July 2001.
- [7] <http://www.katto.comm.waseda.ac.jp/~fujikawa/>
- [8] G. Griffin, A. Holub, and P. Perona, "Caltech 256 object category dataset," Technical Report UCB/CSD-04-1366, California Institute of Technology, 2007.