

領域抽出を利用した一般物体認識に関する研究
A Study on Generic Object Recognition using region extraction

伊東 靖簡[†] 猿田 和樹[†] 寺田 裕樹[†] 武田 和時[†]
Yasuhiro Ito Kazuki Saruta Yuki Terata Kazutoki Takeda

1. はじめに

一般物体認識は対象物を一般的な名称で認識する技術であり、人と計算機の視覚情報処理のギャップを狭める技術として期待されている。従来の物体認識では SIFT, Haar-like 特徴などの形状の局所特徴以外に、大まかな物体形状を表現する特徴量として HOG 特徴量が提案されている。この特徴量は人物検出等、その他の物体に対しても有効性が示されている^[1]。しかし、HOG 特徴量は認識対象となる物体の位置、サイズに依存してしまう。そこで、本研究では前処理により、画像中の物体領域を抽出することで HOG 特徴量による認識精度の向上を検討する。

2. Histograms of Oriented Gradients(HOG)^[2]

HOG は、SIFT と同様に局所領域における輝度の勾配方向をヒストグラム化した特徴量である。

各ピクセルの輝度から勾配強度 m と勾配方向 θ を式(1)~(3)より算出する。HOG 特徴量は算出された m, θ を用いて、 5×5 ピクセルをセルとした領域において輝度の 9 方向の勾配ヒストグラムを作成する。しかし、入力される画像のサイズが異なる場合、生成されるセルの数が異なるため、図 1 のような画像の正規化を必要とする。そのため、図 1 のような黒帯部分を特徴量として抽出してしまい、認識の際に影響を及ぼしてしまうと考えられる。そこで本研究では、すべての画像で縦、横を 10 等分し、 10×10 セルとし、セル数を統一する。さらに、 3×3 セルを 1 ブロックとして式(4)で正規化を行う。 i 行 j 列のセル(i, j)の特徴量を $F_{ij}=[f_1, f_2, \dots, f_9]$ とすると、 k 番目のブロックの特徴量(81次元)は、 $V_k=[F_{i_1j_1}, \dots, F_{i_{10}j_{10}}]$ と表すことができる。

正規化は、ブロックを 1 セルずつ移動するたびに、正規化を行う。本実験では、横方向 8 ブロック、縦方向に 8 ブロック、合計 64 ブロックに対して正規化を行う。画像毎の HOG 特徴量は 64 ブロック \times 81 次元で 5184 次元となる。

$$m(u, v) = \sqrt{f_u(u, v)^2 + f_v(u, v)^2} \quad (1)$$

$$\theta(u, v) = \tan^{-1}(f_v(u, v) / f_u(u, v)) \quad (2)$$

$$\begin{cases} f_u(u, v) = I(u+1, v) - I(u-1, v) \\ f_v(u, v) = I(u, v+1) - I(u, v-1) \end{cases} \quad (3)$$

$$v = f / \sqrt{\|V_k\|_2^2 + \epsilon} \quad (\epsilon = 1) \quad (4)$$



(a)原画像 (b)正規化

図 1 正規化の画像例

3. 物体領域の抽出

HOG 特徴量は画像中の物体の位置、サイズに不変ではない。そのため、図 2(a)のような画像で学習した場合、評価画像 A は正しく認識できても、評価画像 B は誤認識してしまう。そこで本研究では、前処理で画像中の物体領域を抽出後、その物体領域に対して特徴抽出をすることにより、物体の位置、サイズにより生じる影響を吸収し、認識精度を向上させることを検討する。画像入力から特徴抽出までの流れと各工程処理を施した画像例を図 3 に示す。

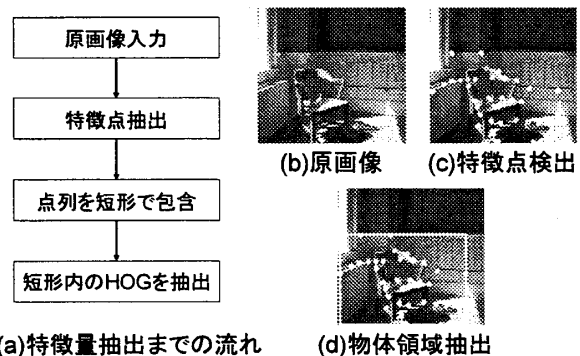
本手法では、画像の拡大・縮小や回転に対して頑強に特徴点を検出可能な Harris オペレータ^[3]を用いることで、図 3(c)のような、入力画像上の輝度エッジの交点やコーナーなどを特徴点として検出する。この Harris オペレータは回転、拡大などの画像の変形操作を行った場合に他のオペレータに比べて、同じ位置に特徴点が検出される再現度が最も高いオペレータであるとされている。次に、抽出された特徴点を包含する矩形を求める。このとき、矩形に傾きが生じると、その後の HOG 特徴量の抽出で影響が発生するため、傾きのない矩形で包含する。その矩形内部を物体領域として、図 3(d)の矩形で囲まれた領域のように抽出する。

最後に物体領域の HOG 特徴量を抽出する。矩形領域の形状は大小、縦長、横長と様々であるが本研究で用いる HOG 特徴量の抽出手法であれば、このようなサイズ、縦横比に依存せず、精度良く特徴量を抽出できると考えられる。



(a)学習画像 (b)評価画像A (c)評価画像B

図 2 位置、サイズの異なる物体例



(a)特徴量抽出までの流れ (d)物体領域抽出

図 3 特徴量抽出までの流れと各工程の画像例

4. 認識実験

4.1 認識手法

認識手法には、高い認識精度を得るために識別精度の高い学習機械で知られるサポートベクターマシン(SVM)を使用する。また、SVMは2値クラス分類器であるため、多クラス認識を行うためには複数のSVMを必要とする。本研究ではガウスクERNELを用いた1対1の非線形SVMを使用する。そのため、Nカテゴリを認識するために学習処理により $N(N-1)/2$ 個のSVMを構築する。そして認識処理では全ての組み合わせの結果、最も投票率の高いカテゴリを認識結果とする。

4.2 実験概要

認識実験にはカルフォルニア工科大学で作成された画像データセット Caltech101 から bass, cannon, ceiling_fan, chair, cup, lotus, soccerball, starfish, stop-sign, watch の10カテゴリを用いる。実験では、カテゴリ毎ランダムに30枚の学習画像を選択し、残りを評価画像として用いる。そのため、学習画像は合計で300枚、評価画像は482枚となる。この作業を3回繰り返し、平均認識率で評価する。

実験では、原画像からのみの特徴抽出、領域抽出後の画像からのみの特徴抽出、両者からの特徴抽出の3パターンで学習データを得て、学習・認識の比較を行う。原画像と領域抽出後の画像からの特徴抽出では、1枚の画像から2つの学習データを得ることになるので、600枚の画像データで学習することになる。また、認識時には原画像、領域抽出後の画像の2つの認識結果が与えられるが、SVMの出力値が大きいカテゴリ(最大出力(原画像、領域抽出後))を最終的な認識結果として判定する。

5. 実験結果・検討

10カテゴリの内、chair, cup, lotus, stop-sign, watch と全体の認識結果を図4に示す。図4より、原画像から抽出されるHOG特徴と領域抽出後の画像から抽出されるHOG特徴による認識結果を比較すると、カテゴリ別ではcupを除いて原画像の方が高い認識率となり、全体でも原画像の方が2%高い認識率を得る結果となった。これは、前処理の段階で正確に領域抽出ができずに学習画像、評価画像として使われてしまった画像があるためと考えられる。図5に領域抽出の失敗例を示す。一方で、認識結果の画像を各々見ていくと原画像では正しく認識できないが、領域抽出後の画像では正しく認識できる場合もある。そのため、学習、認識処理に原画像、領域抽出後の画像から特徴抽出を行った場合、学習データの増加、相補性の効果もあり認識率は向上した。これらの結果より、領域抽出の精度を向上させることが可能になれば、さらに認識率を向上できると考えられる。

図4の原画像と領域抽出後の画像を用いた実験結果では、認識結果を最大出力(原画像、領域抽出後)のようにしたとき、一方では正確に認識しているにもかかわらず、誤認識の出力が大きいため、誤認識の判定が出ている場合もある。そのため、2出力の内、一方が正解であれば、認識されたとみなし(OR出力(原画像、領域抽出後))実験を行った。その結果を図6に示す。図6で示されるように原画像、最大出力(原画像、領域抽出後)を用いた結果と比較してすべ

てのカテゴリで認識率が向上していることがわかる。また、全体でも原画像だけを用いた場合と比較すると約10%、最大出力(原画像、領域抽出後)の場合と比較すると約7%向上する結果となった。この結果より、2つのSVM出力の比較の際により精度の高い判定が可能であれば、認識率をさらに向上できると考えられる。

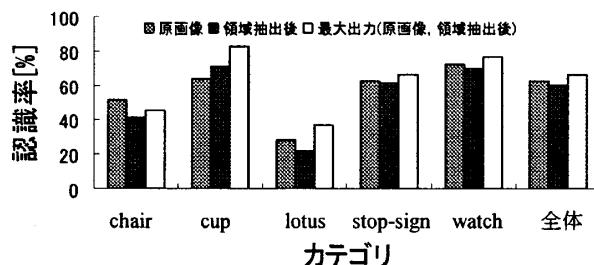


図4 領域抽出による認識率の変化

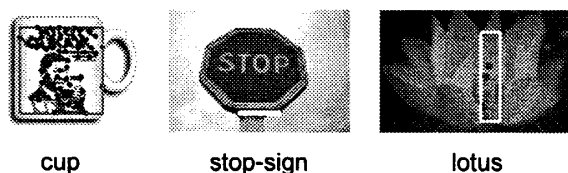


図5 領域抽出の失敗例

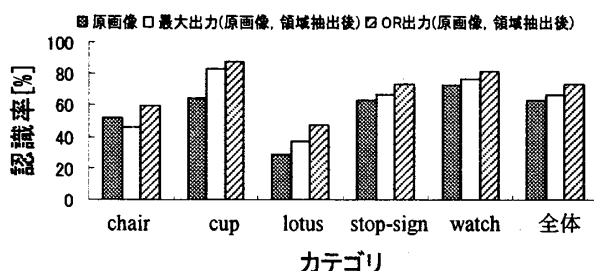


図6 OR出力との比較

6. まとめ

本研究ではHOG特徴量を抽出する前処理で物体領域を抽出することにより、物体の位置、大きさにより生じる影響を吸収し、その有効性を示した。今後は、より精度の高い物体領域の抽出方法、また原画像と領域抽出後画像を用いた場合の2つのSVM出力の比較時の判定について検討する。

参考文献

- [1] 藤吉弘亘, "Gradientベースの特徴抽出-SIFTとHOG-", 信学技報, PRMU2007-82, pp211-pp223(2007)
- [2] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection", IEEE Computer Vision and Pattern Recognition, pp. 886-893(2005)
- [3] C. Harris and M. Stephens, "A combined corner and edge detector", Proc. of Fourth Alvey Vision Conference, pp. 147-151(1988)

† 秋田県立大学大学院システム科学技術研究科
Graduate School of Systems Science and Technology, Akita
Prefectural University