

物体領域特徴の自動選定とマルチカーネル学習を用いた 特徴統合による一般物体認識

中鹿 亘[†] 須賀 晃[†] 滝口 哲也^{††} 有木 康雄^{††}

[†] 神戸大学大学院工学研究科 〒 657-8501 兵庫県神戸市灘区六甲台町 1-1

^{††} 神戸大学自然科学系先端融合研究環 〒 657-8501 兵庫県神戸市灘区六甲台町 1-1

E-mail: †{nakashika,akira1234}@me.cs.scitec.kobe-u.ac.jp, ††{takigu,ariki}@kobe-u.ac.jp

あらまし 近年、一般物体認識の研究において、画像特徴の統合による識別手法が注目を集めている。しかし特徴統合による認識を行う際、画像中の背景領域には物体を認識する上で誤認識を招く特徴や、精度を下げるノイズが多く含まれている。そこで本研究では、事前に大まかな背景領域の除去を行って、抽出される特徴を物体領域内の特徴に選定し、マルチカーネル学習によって特徴量を統合した一般物体認識手法を提案する。これにより、背景ノイズの少ない選定された特徴を用いるため、より正確な重みが学習できる。また、従来の重み付き統合手法では、特徴量ごとに重みを求めていたが、本稿では、より詳細に有効な特徴次元を見るために、各次元ごとの重み付けによる特徴統合を行っている。実験結果より、Saliency Map と Graph Cuts による大まかな背景除去を行うことで従来手法に比べて識別精度が向上し、背景除去による自動特徴選定の有効性が確認できた。更に、各次元ごとのマルチカーネル学習による重み付け処理を行うことで、識別精度を向上させることができた。

キーワード 一般物体認識, マルチカーネル学習, 特徴統合, Graph Cuts, SIFT, HOG

Automatic Selection of Object Region and Feature Integration by Multiple Kernel Learning for Generic Object Recognition

Toru NAKASHIKA[†], Akira SUGA[†], Tetsuya TAKIGUCHI^{††}, and Yasuo ARIKI^{††}

[†] Graduate School of Engineering, Kobe University 1-1, Rokkodai, Nada, Kobe, 657-8501 Japan

^{††} Organization of Advanced Science and Technology, Kobe University 1-1, Rokkodai, Nada, Kobe, 657-8501 Japan

E-mail: †{nakashika,akira1234}@me.cs.scitec.kobe-u.ac.jp, ††{takigu,ariki}@kobe-u.ac.jp

Abstract Recently, in the research of generic object recognition, classification technique based on integration of image features is drawing attention. However, in the classifying technique by feature integration, there are some features that may cause misrecognition of objects and a lot of noises that degrades the recognition accuracy in an image data. In this paper, we propose feature selection in the object area restricted by reducing background area in it, and multiple kernel learning (MKL) to weight each dimension as well as features themselves. This enables the accurate and effective weighting since the weight is computed for each dimension by using selected feature. Experimental results indicate the validity of automatic feature selection. Classification performance is improved by using background reduction technique by Saliency and Graph Cuts, and each dimensional weighting method by MKL.

Key words generic object recognition, multi kernel learning, feature integration, SIFT, HOG

1. はじめに

近年、デジタルカメラの普及や、WWW(World Wide Web) 上の画像の爆発的な増大によって、大量の画像データの分類や検索、画像に対する自動キーワード付けなどの実現の期待が高まっている。こうした背景の中で、デジタル画像の内容を計算機に自動的に認識させる一般物

体認識の研究が近年盛んに行われてきている [1]。

一般物体認識分野の研究においては、BoF(Bag-of-Features) を特徴量として用いた手法が主流となっている [2]。これは、特定物体認識分野の研究において良いパフォーマンスを得ていた SIFT 特徴 [3] を抽象化し、一般物体認識分野に応用した技術である。近年ではさらに、人間が様々な特徴情報から物の識別を行うように、

BoF 単体の特徴量だけでなく複数の特徴量を用いて、Adaboost や TF-IDF の理論により識別度の高い特徴を選択して識別を行う、特徴選択による手法が提案されている [4], [5]。また、最新の機械学習法であるマルチカーネル学習 (Multiple Kernel Learning; MKL) を用いて複数の画像特徴を統合して高精度な認識を実現する手法が提案されている [6], [7]。

このように、近年の一般物体認識の研究において、画像特徴の統合による識別手法が注目を集めている。しかし特徴統合による認識を行う際、画像中の背景領域には物体を認識する上で誤認識を招く特徴や、精度を下げるノイズが多く含まれているため、背景領域特徴を含めたまま統合すると、認識に悪影響が出ることが考えられる。特徴統合による分類手法では、識別に有効な特徴を各物体カテゴリごとに学習するため、背景を含んだ状態での特徴の重み学習は重みそのものを誤らせる可能性があるため、背景領域特徴の事前除去は重要な処理となる。しかしながら、多くの物体カテゴリと大量の学習画像データ全てに対し、人手による背景の除去を行ったり、分類する際に毎回入力画像の背景除去処理を手動で行うのは現実的に困難である。そこで本研究では、認識の前処理として Saliency Map [8] を用いて画像中の視覚注意を引く領域を物体領域として抽出し、Graph Cuts [9], [10] によりその領域を切り出しておくことで背景の除去を行う。これにより、事前情報を用いずに学習画像、入力画像共に大まかな背景の自動除去が可能となる。

また、MKL を用いた特徴統合による画像分類では、識別に有効な特徴量を学習する。クラスの分類タスク毎に、最適な特徴量の重みを用いて認識することができるが、従来の手法では各特徴量ごとに重み付けするため、各特徴量の各次元の重みは一律である。しかし、識別に有効とされた特徴量の中でも、識別に有効な次元もあれば有効でない次元も存在すると考えられる。或いは、特徴量全体では識別にあまり有効でなくても、或る次元だけみれば有効な次元も存在する可能性がある。そこで本研究では、MKL による特徴量への重み付けを一律にするのではなく、各特徴量の各次元ごとにカーネルを用意して、より詳細な重み付けを行う。本論文では、Saliency Map と Graph Cuts を用いた自動背景除去と、MKL による各次元への重み付けによる認識を組み合わせ、ノイズの少ない状態で MKL 学習を行うことで重みの精度を上げ、認識率を向上させる手法を提案する。

2. 提案手法の概要

図 1 に提案手法の流れを示す。学習段階では、まず物体領域を大まかに切り出し、背景ノイズを抑えた物体領域の特徴を抽出し、その特徴を用いて MKL による特徴統合を行う。物体領域の抽出には、Saliency Map を用いる。一般に、画像中の物体領域は背景領域に比べて顕著性が高いため、Saliency Map により物体領域ら

しい部分を算出することができる。そして、作成された Saliency Map を Graph Cuts の seeds として用いることで自動セグメンテーションを行う。セグメンテーションは多重解像度解析を用いて解像度を変えながら複数回繰り返され、これにより大まかな物体領域が切り出される。特徴抽出では Bag-of-Features(DoG), Bag-of-Features(Grid), Color, Gabor の 4 つの画像特徴を抽出する。背景が除去された状態から特徴を抽出するため、画像全体から特徴を抽出する場合と比べて、物体領域のより正確な特徴が得られる。

テスト段階においても、同様に、Saliency Map の作成、Graph Cuts によるセグメンテーションを行い、背景を除去した上で選定された物体領域上の特徴を抽出する。

このように自動で選定された物体領域特徴を用いて、特徴統合段階では、マルチカーネル学習により識別に有効な特徴を学習する。識別器は各カテゴリごとに用意し、それぞれのカテゴリで識別に有効な特徴に重み付けを行う。この際、従来手法では各特徴量ごとに重み学習を行っていたが、本手法では各特徴の各次元ごとに重み学習を行う。これにより、識別に向いていない次元による識別結果は考慮せず、複数の特徴量のうち識別に有効な次元のみで識別を行うことができるため、より詳細な特徴統合が可能となる。

3. 物体特徴領域の自動選定

一般物体認識において、各物体の識別的な特徴量を学習したり認識を行う上で、背景から得られる情報はその精度に悪影響を与えてしまう。そこで本手法では Saliency Map を用いて事前に物体領域を大まかに特定し、Graph Cuts セグメンテーションにより背景部分を除去することで認識精度を高める。本章では、Saliency Map 及び Graph Cuts のアルゴリズムについて述べ、これらの手法を組み合わせることで自動的に背景を除去する手法について説明する。

3.1 Saliency Map

Saliency Map は、画像中の顕著性の高い領域を、視覚的注意を引く領域として抽出する手法である。一般に、物体領域は背景領域に比べて顕著性が高いため、Saliency Map により物体領域を大まかに抽出することができる。Saliency Map は Itti らによって提案された [11]。この方法では、入力画像を Gaussian Pyramid によりダウンサンプリングし、各画素において Center-Surround のスケール間差分を求め、Map を作成する。そのため、作成される Saliency Map は解像度が入力画像よりも落ちてしまい、全体的にぼやけたものになってしまう。本研究では、入力画像をダウンサンプリングするのではなく、フィルタを用意してそのスケールサイズを変えながらラスタスキャン状に画素ごとのスケール差分をと

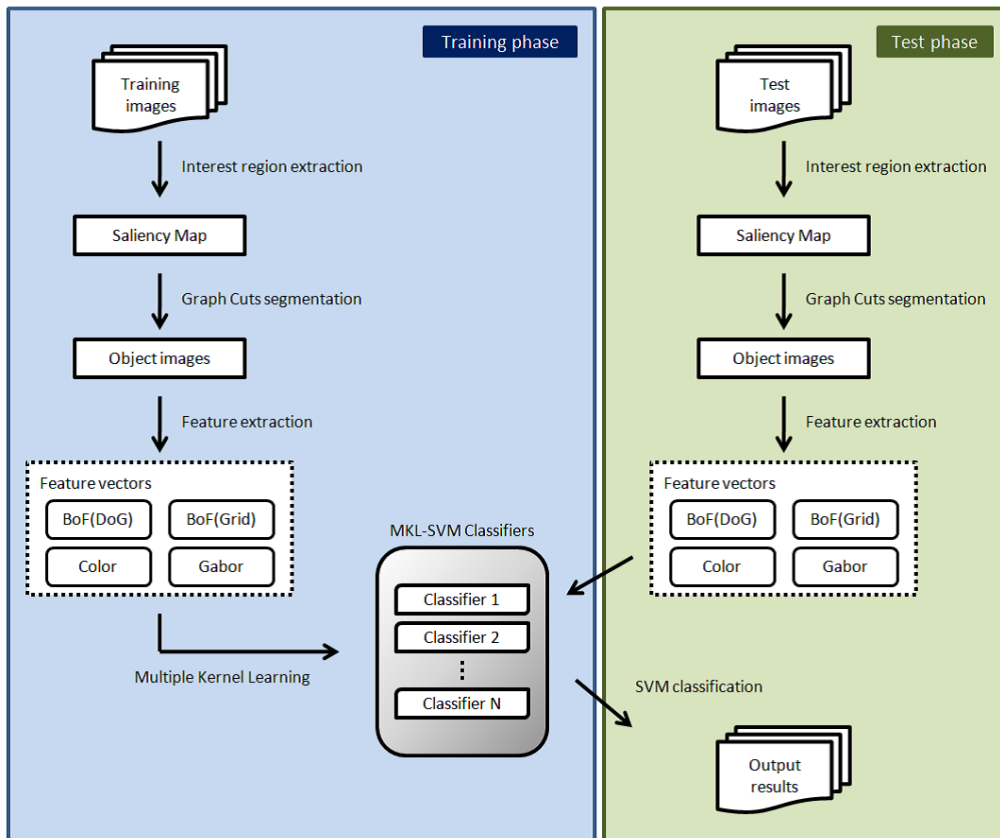


図 1 提案システム

る Achanta らの方法を用いる [12]。これにより、入力画像から解像度を落とさずに Saliency Map を作成できるため繊細な Saliency の値を算出できる。また、従来の Saliency Map の算出方法では、「画像の両端は背景である」という仮定の元で計算されていたが、本研究で採用する手法では、画像の全領域に対して同一のフィルタをかけるので、画像の中央に対象物体が存在しなくてもよいという利点がある。

3.2 Graph Cuts

画像セグメンテーションにおいて、セグメンテーション問題をエネルギー最小化問題と捉えて解を求める手法が多く提案されている。Sankes [13] や Level Set Method [14] はその例であるが、これらの手法では境界線に対するエネルギー関数から局所最小解を求める。それに対して、Graph Cuts では、各領域からエネルギー関数を定義し、大域最小解を求めることが可能で、領域と境界の両方の特性を用いたセグメンテーションが可能である。

本研究では従来の Graph Cuts [9] ではなく、多重解像度解析を用いて、平滑化を繰り返しながらセグメンテーションを行う手法 [15], [16] を用いる。図 2 にセグメンテーションの流れを示す。

はじめに、画像をレベル t まで多重解像度解析を行い、低周波成分である平滑化画像は n -link に、また高周波成分から局所テクスチャ特徴量を定義し、色特徴量ととも

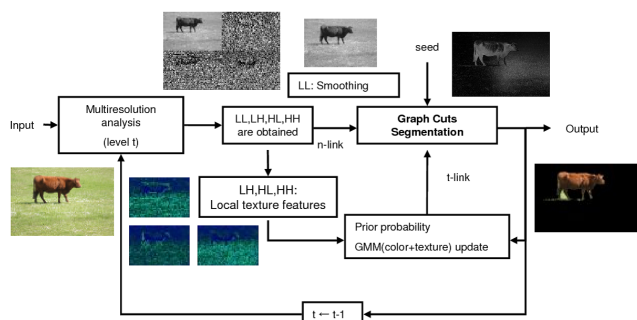


図 2 平滑化と局所テクスチャ特徴量を用いた Graph Cuts

に GMM(Gaussian Mixture Model) に適用し、物体及び背景の尤度関数を求める。同時に、1 つ上の多重解像度レベルにおける処理結果画像に距離変換を行い事前確率を推定し、事前確率と尤度関数から求めた事後確率を t -link に用いる。これを、多重解像度レベルを 1 つずつ下げながら、 $t = 0$ になるまで繰り返し行う。

前節で定義した Saliency Map を seeds として、Graph Cuts セグメンテーションを用いることで、精度良く物体領域を選定することができる。また、この手法により背景を完全に除去することは困難であるが、ほとんどの対象物体に Saliency が表れるので、背景領域を多少なりとも削ることはできる。本研究の目的は背景除去でなく一般物体認識であるため、予め背景領域を削っておき、対象物体とは関係のない余分な情報を削除することで、物

体の認識性能の向上が期待できる。

4. マルチカーネル学習による特徴統合

本手法では、背景除去によって得られる選定された画像特徴を用いて、マルチカーネル学習により識別に最適な重みを学習して画像分類を行う。従来手法では各特徴量ごとに重みを学習していたが、本手法ではより詳細な重み付けを行うため、各次元ごとにカーネルを用意して、重みの学習を行う。次節以降で統合を行う特徴量と、その重み付け学習の方法について順に述べる。

4.1 画像特徴量

本手法で画像分類に用いる特徴量と、その抽出方法について述べる。使用する特徴量は、DoGによる局所特徴量、Gridによる局所特徴量、色特徴量、大域特徴量の4種類である。

4.1.1 局所特徴量 (Bag-of-Features, DoG)

Bag-of-Featuresでは、図3に示すように、まず各学習画像からSIFT特徴を抽出する。ここではDoG(Difference-of-Gaussian)処理を使って画像中の特徴点を抽出する。これはスケールの異なるガウシアンフィルタをかけて画像を平滑化し、それぞれの差分画像から極値を検出することで特徴点を決定する手法である。この特徴点の周辺領域を4×4ブロックに分割し、各ブロックごとに8方向の勾配方向ヒストグラムを作成することでSIFT特徴が得られる。以上より回転・スケール変化にロバストな128次元の特徴量が抽出できる。次に、得られた全SIFT特徴に対してSIFT特徴空間上でk-meansクラスタリングを行う。このクラスタリングによって得られた各クラスをvisual wordとみなし、visual vocabularyを構築する。1枚の画像に対し、visual wordのヒストグラムを1000次元の特徴ベクトルで表現する。

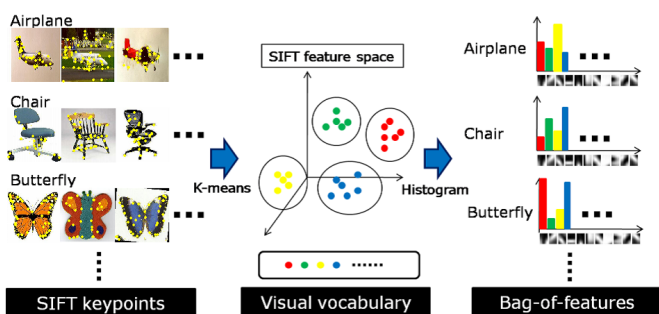


図3 Bag-of-Features

4.1.2 局所特徴量 (Bag-of-Features, Grid)

本手法では上述のDoG処理に基づく特徴検出以外にも、図4に示すように格子状に等間隔で抽出するグリッドサンプリングによる特徴点検出法を用いる。グリッドサンプリングでは、特徴点のスケールは実験的に複数の値に定められる。各円の中心が特徴点で、円の半径の大きさが特徴点のスケールの大きさを表す。このように検

出を行うことで、スケールサイズを一意に定めることはできないものの、特徴点位置に偏りのない抽出ができるため、DoGによるSIFTとは異なる表現が可能である。

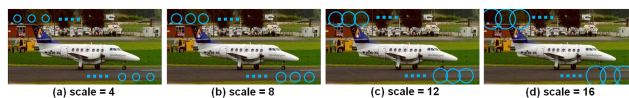


図4 グリッドサンプリングの例

4.1.3 色特徴量

SIFT特徴では、グレースケール画像からの輝度変化を用いているが、色情報は用いていない。しかしカテゴリによっては色情報が重要な場合が考えられる。そこで本研究では、図5のように、画像を3×3ブロックに分割し、位置情報を考慮した色の分布を色特徴量として用いる。分割したそれぞれの部分領域から64次元のRGBヒストグラムを作成し、それらを統合することにより画像を1つの特徴ベクトルで表現する。

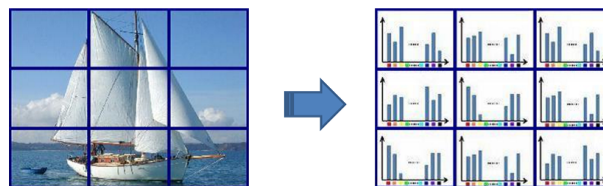


図5 色特徴量

4.1.4 テクスチャ特徴量 (Gabor Features)

ガボール特徴は、画像から局所的な濃淡情報の周期と方向を表した特徴量である。カーネルの形を固定し、周期を変えて伸び縮みさせたり、回転させて方向を変え、様々な周期や方向のフィルタを作成する。解像度 m 、方向 n のガボールフィルタは次式で表される。

$$g_{m,n}(x,y) = \frac{k_m^2}{\sigma^2} \exp\left\{-\frac{k_m^2(x^2+y^2)}{2\sigma^2}\right\} \times [\exp\{jk_m(x\cos\theta_n + y\sin\theta_n)\} - \exp(-\frac{\sigma^2}{2})] \quad (1)$$

ここで、式(1)の k_m および θ_n は、以下のように表される。

$$k_m = a^m \quad (0 \leq m \leq S-1) \quad (2)$$

$$\theta_n = \frac{n\pi}{K} \quad (0 \leq n \leq K-1) \quad (3)$$

K は方向の数、 S は解像度の数、 a は拡大率を表す。式(1)で表されるフィルタを用いて、それぞれに対応した空間周期の特徴を抽出(パターン強度を数値化)する。ガボールフィルタは、特定の向きのエッジと特定の幅のエッジを抽出する。あるサンプリング点においてガボール特徴が得られる様子を図6に示す。最後に、各フィル

夕毎に強度の平均を求め、それをヒストグラムとする。本研究では6方向、4周期の24次元のベクトルとなっている。色特徴と同様に、画像を3×3ブロックに分割し、特徴量を記述する。ガボール特徴は局所的な情報を見るので、画像の照明変動の影響を受けにくいという利点がある。ガボール特徴は局所的なパターンの特徴を表現するため SIFT と類似しているが、回転に対して不変性がないため、SIFT と異なる表現ができる。

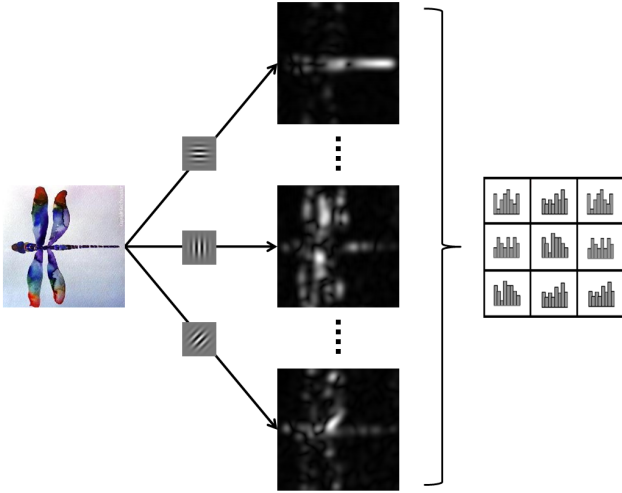


図6 ガボール特徴

4.2 マルチカーネル学習

マルチカーネル学習 (Multiple Kernel Learning; MKL) は、SVM などのカーネルを用いた識別器を複数用いた際に、それぞれのカーネルに対し最適な重みを学習する手法である。Varma らは、各分類クラスに対して最適な特徴を MKL を用いて学習する手法を提案している [6]。ここでは MKL をカーネルの選択のために用いるのではなく、複数の特徴量の最適な重みを計算するために用いており、画像特徴統合の方法として利用されている。これにより、クラスの分類タスクに対して、最適な特徴量を用いて認識することができる。本研究では、それぞれのクラスの各特徴の重みを Varma らの方法 [6] のように MKL を用いて最適な重みを求め、画像分類を行う。すなわち、特徴量を統合して画像を認識するために、マルチカーネル学習によって複数の特徴量のカーネルを線形結合した統合カーネルを作成し、それをサポートベクターマシン (SVM) に適用して特徴統合による画像認識を行う。最適なカーネル (カーネルを重みつきで線形結合したカーネル) のサブカーネルに対する重み β_j を求めることで、分類に最適な特徴量の重みを求める。

統合カーネルは以下の式のように表される。

$$K_{combined}(x, x') = \sum_{j=1}^K \beta_j k_j(x, x')$$

$$\text{with } \beta_j \geq 0, \sum_{j=1}^K \beta_j = 1 \quad (4)$$

ここで、従来の手法では各サブカーネルは各特徴量と対応しているが、本手法では、各サブカーネルは各特徴の各次元に対応している。 K は全特徴の次元数である。次節にサブカーネルの重み β_j の求め方について述べる。

4.2.1 MKL-SVM

MKL は SVM のフレームワークで解く方法が一般的である。2 クラス分類に対する MKL 問題において、 N 個のデータ点 $(x_i, y_i) (y_i \in \pm 1)$ が与えられたとする。この場合の MKL において解くべき最適化問題の主問題は以下のように示される。

$$\min \frac{1}{2} \left(\sum_{k=1}^K \|w_k\|^2 \right) + C \sum_{i=1}^N \xi_i \quad (5)$$

$$\text{w.r.t } w_k \in \mathbb{R}^{D_k}, \xi \in \mathbb{R}^N, b \in \mathbb{R},$$

$$\text{s.t. } \xi_i \geq 0 \text{ and}$$

$$y_i \left(\sum_{k=1}^K \langle w_k, \Phi_k(x_i) \rangle + b \right) \geq 1 - \xi_i,$$

$$\forall i = 1, \dots, N$$

ここで、 $w_k = \beta_k w'_k (\beta_k \geq 0, \forall k = 1, \dots, K), \sum_{k=1}^K \beta_k = 1$ である。

式 (5) に対して、双対問題は以下ようになる [17]。

$$\min \gamma \quad (6)$$

$$\text{w.r.t. } \gamma \in \mathbb{R}, \alpha \in \mathbb{R}^N$$

$$\text{s.t. } 0 \leq \alpha_i \leq C, \sum_{i=1}^N \alpha_i y_i = 0,$$

$$S_k(\alpha) = \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j y_i y_j \mathbf{k}_k(x_i, x_j) - \sum_{i=1}^N \alpha_i \leq \gamma,$$

$$\forall k = 1, \dots, K$$

ここで、 $\mathbf{k}_k(x_i, x_j) = \langle \Phi_k(x_i), \Phi_k(x_j) \rangle$, N は学習データの個数である。単一カーネルの双対問題との違いとしては、カーネルごとに $S_k(\alpha) \leq \gamma$ という拘束条件がある。また、 $\sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j y_i y_j \mathbf{k}_k(x_i, x_j)$ を最大化する代わりに、全カーネルで共通の上限値の γ を式 (6) のように最小化する。 $K = 1$ の場合は、上記の問題は標準の SVM の双対問題と等価になる。

式 (6) を解くため、次のような鞍点問題を考える。

$$L := \gamma + \sum_{k=1}^K \beta_k (S_k(\alpha) - \gamma) \quad (7)$$

これを $\gamma \in \mathbb{R}, \alpha \in \mathbb{R}^N (0 \leq \alpha_i \leq C, \sum_{i=1}^N \alpha_i y_i = 0)$ に関して最小化し、 $\beta \in \mathbb{R}^K (0 \leq \beta) \leq 1$ に関して最大化すれ

ばよい。 γ を 0 とすると、 $\sum_{k=1}^K \beta_k = 1$ の条件のもと、 $L = S(\alpha, \beta) := \sum_{k=1}^K \beta_k S_k(\alpha)$ のように式 (7) を簡略化できる。この Min-Max 問題は、次のような Semi-Infinite Linear Program として書き表せる。

$$\begin{aligned} \max \quad & \theta & (8) \\ \text{w.r.t.} \quad & \theta \in \mathbb{R}, \beta \in \mathbb{R}^K \\ \text{s.t.} \quad & 0 \leq \beta, \sum_k \beta_k = 1 \text{ and } \sum_{k=1}^K \beta_k S_k(\alpha) \geq \theta \\ \text{for all} \quad & \alpha \in \mathbb{R}^N \text{ with } 0 \leq \alpha \leq C \text{ and } \sum_i y_i \alpha_i = 0 \end{aligned}$$

式 (8) を解くために、Sonnenburg らは単一カーネルの SVM 学習の反復による解法を提案している [18]。この解法では、まず初めにカーネル重み β_j を均等重みに初期化する。次に β_j を固定して、統合カーネルを単一のカーネルとみなして SVM 学習を行い α_i を求める。最後に、求めた α_i を固定して、 $\sum_{k=1}^K \beta_k S_k(\alpha)$ が増加するように β_j を変化させていく方法である。詳細なアルゴリズムを図 7 に示す。

```

 $S^0 = 1, \theta^1 = -\infty, \beta_k^1 = \frac{1}{K}$  for  $k = 1, \dots, K$ 
for  $t = 1, 2, \dots$  do
  Compute  $\alpha^t = \arg \min_{\alpha \in C} \sum_{k=1}^K \beta_k^t S_k(\alpha)$  by single
  kernel algorithm with  $\mathbf{k} = \sum_{k=1}^K \beta_k^t \mathbf{k}_k$ 
   $S^t = \sum_{k=1}^K \beta_k^t S_k^t$ , where  $S_k^t = S_k(\alpha^t)$ 
  if  $|1 - \frac{S^t}{\theta^t}| \leq \epsilon_{MKL}$  then
    break
  end if
   $(\beta^{t+1}, \theta^{t+1}) = \arg \max \theta$ 
  w.r.t  $\beta \in \mathbb{R}^K, \theta \in \mathbb{R}$ 
  s.t.  $0 \leq \beta, \sum_{k=1}^K \beta_k = 1$  and  $\sum_{k=1}^K \beta_k S_k^r \geq \theta$  for
   $r = 1, \dots, t$ 
end for

```

図 7 MKL 双対問題の解法アルゴリズム

5. 評価実験

実験ではまず Saliency Map 及び Graph Cuts による物体領域の自動セグメンテーション精度評価を行う。次に、従来の MKL と、提案手法である各次元ごとの MKL による分類精度の比較を行う。また、セグメンテーションによる物体特徴の選定を行った場合と行わなかった場合の分類率についても検証する。

5.1 セグメンテーション精度評価

5.1.1 実験条件

セグメンテーション実験で用いるデータセットは、人・動物・車・花などが写った 50 枚の画像からなる Grab Cuts Database [19] を使用する。物体数は 1 画像につき 1 物体で、各画像には物体・背景領域の正解マスクデータが予め用意されている。

5.1.2 評価尺度

予め用意されている正解マスク画像を正解領域とし、セグメンテーションの誤検出率を検証する。誤検出率は以下の式のように、物体を背景と誤検出した割合 (Over Segmentation) と、背景を物体と誤検出した割合 (Under Segmentation) から求める。

$$\text{Err}[\%] = \left(\frac{E_O}{P} + \frac{E_B}{P} \right) \times 100 \quad (9)$$

ここで、 E_O は物体領域の誤検出ピクセル数、 E_B は背景領域の誤検出ピクセル数、 P は画像全体のピクセル総数である。

5.1.3 実験結果

初期値である多重解像度レベル t の値を変化させて、セグメンテーションを行った結果を表 1 に示す。

表に示されるように、Saliency Map を seeds とした自動セグメンテーションにおいて、解像度レベルを変えながら繰り返しセグメンテーションすることで、物体領域・背景領域ともに 7% 前後の誤検出率でセグメンテーションできていることが分かる。セグメンテーションを繰り返すことで、通常の Graph Cuts ($t = 1$) よりも誤検出率が低くなっている。これは、多重解像度解析により大域的なセグメンテーションから局所的なセグメンテーションへ順次行うことで、局所的なエラーが徐々に修正されたためと考えられる。セグメンテーション結果の例を図 8 に示す。図に示されるように、Saliency Map を seeds として Graph Cuts を行うことで、背景領域情報を大幅に削減することができていることが分かる。Saliency Map では視覚注意を引く大まかな領域しか抽出しかできないが、その領域を初期 seeds として、物体・背景領域を再学習しながら繰り返しセグメンテーションすることで各領域が修正され、検出精度が高まった。

5.2 識別精度評価

5.2.1 実験条件

識別実験に用いるデータセットは、人の顔・車・椅子・カメラなどの様々な一般物体が 101 種類含まれる Caltech-101 データベース [20] を使用する。学習枚数は 1 カテゴリあたり 10 枚、20 枚、30 枚についてそれぞれ実験を行う。使用特徴量は Bag-of-Features (DoG), Bag-of-Features (Grid), 色特徴, ガボール特徴の 4 種類を用いる。識別器は Multiple Kernel を用いた SVM で、カーネル関数は画像分類において最も性能の良かった χ^2 カーネルを用いた [21]。提案手法の有効性を検証するため、以下の 4 手法について平均分類率の比較を行った。

- 物体領域選定なし + 従来の MKL による分類
- 物体領域選定なし + 各次元の MKL による分類
- 物体領域選定あり + 従来の MKL による分類
- 物体領域選定あり + 各次元の MKL による分類

表 1 誤検出率 (%)

	t=1	t=2	t=3	t=4	t=5	t=6
Over Seg.	10.95	9.16	5.97	6.80	6.11	6.52
Under Seg.	18.03	9.56	7.25	7.26	7.26	7.27
Error (Over + Under)	28.98	18.72	13.22	14.06	13.37	13.79

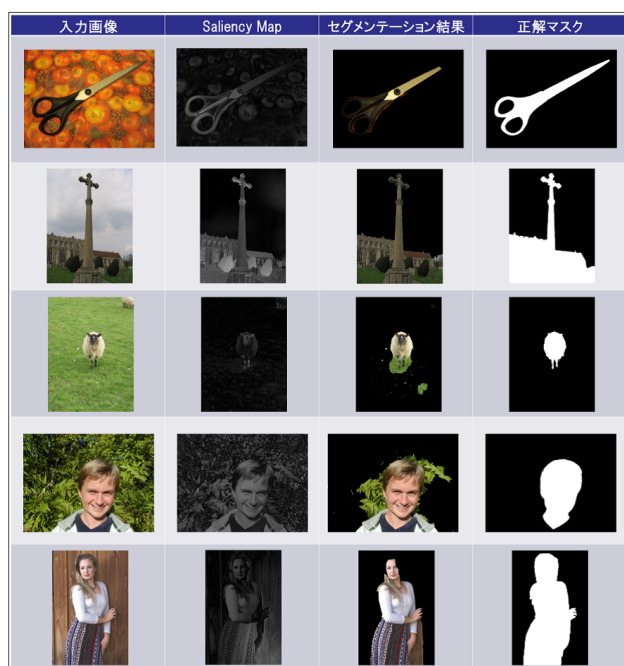


図 8 セグメンテーション結果

5.2.2 実験結果

はじめに、30 枚の学習画像を用いた際の、単独特徴量による分類と MKL 特徴統合による分類における平均分類率の比較を表 2 に示す。

表 2 単独特徴量と特徴統合による平均分類率 (%)

特徴量	平均分類率
Bag-of-Features (DoG)	47.3
Bag-of-Features (Grid)	51.2
Color Features	38.5
Gabor Features	40.5
Multiple Kernel Learning	68.7

表より、単独特徴量による分類に比べて、MKL を用いて特徴統合を行うと大幅に分類精度が向上しており、特徴統合による分類の有効性が確認できる。

次に、この MKL による特徴統合による手法に加え、Saliency Map と Graph Cuts を用いた特徴選定処理と、各次元の MKL による各次元への重み付け処理を加えた場合の平均分類率の結果を図 9 に示す。図中の a~d は、実験条件で述べた (a)~(d) の手法に対応している。

結果を見ると、従来手法 (a) に比べて、各次元の重み付けを行った場合 (b)、Saliency による特徴選定処理を行った場合 (c) のいずれにおいても、分類精度が向上し

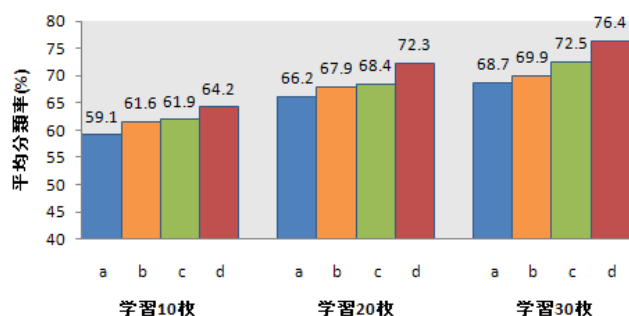


図 9 提案手法による平均分類率 (%)

ていることが確認できる。また、各次元の重み付けと、特徴選定処理を両方行った場合 (d) では、更に大きく精度が向上している。これは、特徴選定処理において、ノイズとなる背景領域の特徴が削減され、より正確な物体領域の特徴が得られたことと、選定された特徴に詳細な重み付けが行われたことにより、相乗的な効果があったためと考えられる。

6. おわりに

本論文では、Saliency Map と Graph Cuts を用いた背景領域の除去による物体特徴選定と、各特徴量の各次元ごとにカーネルを用意し、各次元ごとの詳細な特徴の重みを学習する特徴統合手法を提案した。

従来の重み付き統合手法では、各物体カテゴリごとに、各特徴量の重みをその識別度から求めていたが、特徴量の各次元に対しては一律に同じ重みが与えられていた。そのため、特徴の各次元ごとの有効性・非有効性が考慮されないという問題があった。また、物体カテゴリごとに識別的な特徴を学習する MKL などの特徴統合による手法では、背景ノイズを含めたまま重みを算出すると、重みそのものに誤りが生じてしまう可能性があり、背景領域特徴の事前除去を行うことは重要な課題であった。

提案手法では、Saliency Map と Graph Cuts を用いて事前に大まかな背景領域の除去を行い、背景ノイズを削減することで抽出される特徴を物体領域内の特徴に選定した。また、各特徴量の各次元ごとにカーネルを用意し、各次元ごとの詳細な特徴の重みを学習する手法により、詳細な特徴統合を実現した。

実験の結果、Saliency Map と Graph Cuts による大まかな背景除去を行うことで従来手法に比べて識別精度が向上し、背景除去による自動特徴選定の有効性が確認できた。更に、各次元ごとの MKL による重み付け処理を行うことで、細かい重み付けが可能となり、相乗的に識

別精度が向上することを確認した。

今後の課題として、複数物体の同時識別への拡張が挙げられる。現在のシステムは、基本的に1画像1物体を想定したものとなっている。Saliency Mapは、画像中の視覚注意を引く領域を抽出できるが、物体個別に切り分けることができないため、同一画像中の複数の物体を同時に切り出してきて認識することが難しいためである。複数物体認識へ拡張するためには各物体固有のSaliencyを学習するなどして、各物体を抽出する手法を検討する必要がある。また、特徴統合の際に各特徴次元間の関連性を考慮し、関連性の強い次元を結びつけてカーネル学習をすることで、精度の向上を検討する予定である。

文 献

- [1] 柳井啓司, “一般物体認識の現状と今後,” 情報処理学会論文誌: コンピュータビジョン・イメージメディア, vol. 48, no. SIG16 (CVIM19), pp. 1–24, 2007.
- [2] G. Csurka, “Visual categorization with bags of keypoints,” Proc. of ECCV Workshop on Statistical Learning in Computer Vision, 1–22, 2004.
- [3] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” Journal of Computer Vision, pp.91–110, 2004.
- [4] 土屋成光, 藤吉弘巨, “Boost 学習に基づく特徴量の貢献度を用いた特徴選択手法,” 画像の認識・理解シンポジウム, MIRU2008, IS2-1, pp. 599–604, 2008.
- [5] 須賀晃, 滝口哲也, 有木康雄, “複数特徴量の重み付け統合による一般物体認識,” 情報処理学会平成 21 年度関西支部大会, C-03, 2009.
- [6] M. Varma, D. Ray, “Learning the discriminative power-invariance trade-off,” In Proceedings of the IEEE International Conference on Computer Vision, Rio de Janeiro, Brazil, October. 2007.
- [7] A. Vedaldi, V. Gulshan, M. Varma, A. Zisserman, “Multiple Kernels for Object Detection,” In Proceedings of the International Conference on Computer Vision, Kyoto, Japan, September. 2009.
- [8] L. Itti, C. Koch, E. Niebur, “A Model of Saliency-based Visual Attention for Rapid Scene Analysis,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 20, no. 11, pp. 1254–1259, 1998.
- [9] Y. Boykov, G. Funka-Lea, “Graph Cuts and Efficient N-D Image Segmentation,” International Journal of Computer Vision, 70(2), pp. 109–131, 2006.
- [10] 石川博, “グラフカット,” 情報処理学会研究報告, 2007-CVIM-158-26, pp. 193–204, 2007.
- [11] L. Itti, C. Koch, E. Niebur, “A Model of Saliency-based Visual Attention for Rapid Scene Analysis,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 20, no. 11, pp. 1254–1259, 1998.
- [12] R. Achanta, “Salient Region Detection and Segmentation,” 6th International Conference on Computer Vision Systems, pp.66–75, 2008.
- [13] M. Kass, A. Witkin, and D. Terzopoulos, “Snakes: Active contour models,” International Journal of Computer Vision, vol. 1, no. 4, pp. 321–331, 1988.
- [14] James A. Sethian, “Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry Fluid Mechanics,” Computer Vision, and Materials Science. Cambridge University Press, 1999.
- [15] T.Nagahashi, H.Fujiyoshi, T.Kanade, “Image Segmentation Using Iterated Graph Cuts Based on Multi-scale Smoothing,” ACCV 2007, Part II, LNCS 4844, pp. 806–816, 2007.
- [16] Keita Fukuda, Tetsuya Takiguchi, Yasuo Ariki, “GRAPH CUTS BY USING LOCAL TEXTURE FEATURES OF WAVELET COEFFICIENT FOR IMAGE SEGMENTATION,” ICME 2008 (International Conference on Multimedia and Expo), WPM2-L3.2, pp. 881–884, 2008.
- [17] F. R. Bach, G. R. G. Lanckriet, M. Jordan, “Multiple kernel learning, conic duality, and the SMO algorithm,” Proc. of International Conference on Machine Learning, 2004.
- [18] S. Sonnenburg, G. Ratsch, C. Schölkopf, B. Schölkopf, “Large Scale Multiple Kernel Learning,” Journal of Machine Learning Research, vol. 7, pp. 1531–1565, 2006.
- [19] GrabCut Image Database, <http://research.microsoft.com/en-us/um/cambridge/projects/visionimagevideoediting/segmentation/grabcut>
- [20] Caltech 101 Database, http://www.vision.caltech.edu/Image_Datasets/Caltech101/
- [21] J. Zhang, M. Marszalek, S. Lazebnik, C. Schmid, “Local Features and Kernels for Classification of Texture and Object Categories,” A Comprehensive Study, International Journal of Computer Vision, vol. 73, no. 2, pp. 213–238, 2007.