修　　士　　論　　文

一般物体を対象とした
複数候補提示下での分類性能の向上

Improving precisions in multi-candidate generic object image recognition

指導教員　　近山 隆 教授

東京大学大学院
工学系研究科
電子工学専攻

氏 名　　076443 栗田 哲平

提 出 日　　平成 21 年 2 月 3 日

概 要

従来，一般物体認識問題に対する手法への評価は，主にテストデータに対応するクラスを一意に推定する際の精度で性能を決定してきた．しかし一般には複数候補を提示するようなシステムを必要とする場合も多く，その場合の性能は正解を上位候補と出来るか否かの問題になる．そこで本稿では，一般物体認識のデータセットにおいて複数候補提示下での精度について着目し，その分類性能の向上を試みる．まず予備実験として，様々な条件下で一般物体認識問題についての精度傾向の解析を行い，複数候補提示下の分類に適した構成・パラメータの設定等を得る．その結果としての知見を利用し，複数クラス分類において 1 対 1 で構成される Support Vector Machine(SVM) の各識別平面でクラス毎の学習データに対して動的な最適化を行う．結果として，複数候補提示下での分類性能を向上させたことを示す．

目 次

図 目 次

表 目 次

第1章　序論

1.1　背景と目的

1.1.1　背景

近年，画像情報はインターネット上のコンテンツなどを通してその利用が急激に増加している．その画像情報の検索技術についても，web 上でのテキストと関連付けた検索手法のみならず，画像情報を頼りにした検索手法も多くなり，画像認識の研究が活躍している．そのように画像認識技術の需要が高まっている中で，画像認識でも一般の物体をその対象名で認識する事を generic object recognition（一般物体認識）と言い，数十年の間，研究がなされている．この研究が長期間続けられ，今再度盛んになっている理由として，一般の物体を計算機に認識させる事は非常に困難であるが，計算機の進化と共に大規模なデータを扱えるようになり様々な手法を試す事が可能になってきたことが挙げられる．また，画像は人々の生活に深く密接しているということも大きな原動力となっていると考えられる．

計算機による認識とは，結局は様々なセンサから入力される信号を抽象的な概念に写像をする処理（図 1.1）になる．パターンと概念の間には，大きなセマンティックギャップが存在し，計算機による処理は困難を極める．これが一般物体認識を困難にさせている大きな原因である．

現在，様々な一般物体認識の手法が考案され，データベースセットに対しての精度を高めようとしているが，その認識率を限りなく 100%に近づけるのは現状では困難である．そもそも実用性を考慮すると，たとえ計算機がパターンを概念に写像できたとしても，その認識結果をどのようにして使用するのか，といういわゆるアクションの部分が存在しなければ人間の役には立たず，そのような有効なアクションが取れる分野は限られてくる．画像認識の各段階において，どのような手法を採用するのかは分野や処理の目的により変化してくる．実時間性が必要になってくる場合には，処理の精度が多少悪くても高速なアルゴリズムを採用しなければならなく，逆に間違った例を提示する事が致命傷になる場合は，提示を無理にしないという判断も必要になってくる．人間の直感に近い認識結果が必要な時は，認識率の精度が多少悪くなっても，認知科学的な分野の手法を用いたアルゴリズムを選択する場合も出てくる．

また，システムによっては必ずしも一つの候補に絞る必要もない場合が考えられる．現在の一般物体認識問題に対する手法への評価は，主にテストデータに対応するクラスを一意に推定する際の精度によって性能を比べ，手法の優劣を決定している事が多い．しかし一般には複数候補を提示するようなシステムを必要とする場合も多く，その場合の性能は正解を上位候補と出来るか否かの問題になってくる．

具体例として，図 1.2 のように大量の画像データから物品検索を行う際に，候補を複数提示しても良いが漏れなく検索をするという要件を持ったシステム構築を想定する．この場合には基本的に，検索されるデータとして取り過ぎ (false Positive) をしても多少良いが，取り残し (false Negative) の部分が出てしまうと，システムとして致命的になってしまうので，なるべくその部分をなくすように再現率 (recall) を維持しながら，適合率 (precision) を高める工夫が必要となる．

近年の一般物体認識の研究では，精度としてテストデータに対応するクラスを一意に推定する際の精度 (precision) によって性能を比較しているものが殆どである．これはあまりパラメータなどの知見が無いことや，また様々な研究者が一つの目標に対して競い合っていることに起因している．もちろん高速な手法で精度が限りなく 100%に近くなればそれで良いのであるが，現状では困難である．

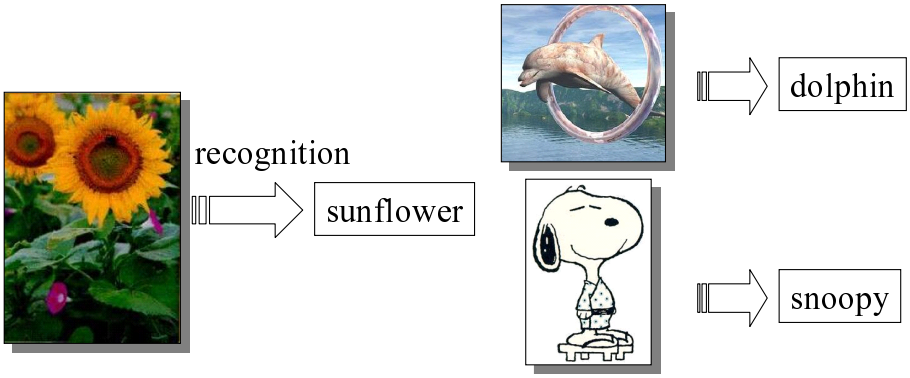


図 1.1: 一般物体の認識

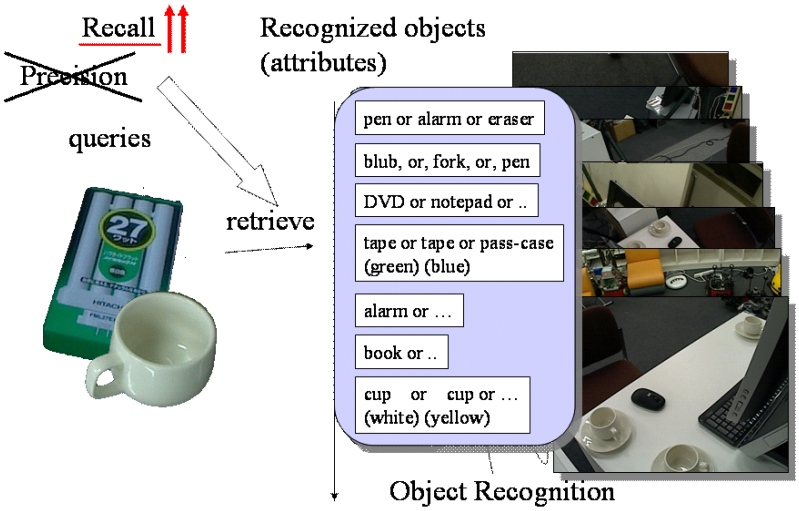


図 1.2: 大量の画像データからの検索システムの例

1.1.2 本研究の目的

本研究では実用性を見据えて，一般物体認識について複数候補提示下での精度傾向について着目し，その分類性能を向上させる事を目的とする．ここで最適化をするべき事項は，精度が 100%近くに収束する候補提示数を小さくすることであり，従来のようなテストデータに対応するクラスを一意に推定する際の精度をただ高めることとは異なる事に注意して欲しい．

1.2 画像認識の中での本研究の位置づけ

本研究の位置づけを述べる前に，一般的な画像の認識とその過程について簡単に解説する．

1.2.1 画像の認識

画像の認識とその過程は，対象・目的によって非常に多種多様である．その中で基本と考えられる概念を図 1.3 に示す．この処理全体の中で，入力はそのときの認識・分類の対象となる画像とし，出力はあらかじめ定められた数種類のクラス名とする．

以下に処理過程の各ステップについて説明する．

a) 変換および前処理

入力画像に処理を施して b) のセグメンテーション以下を行いやすくする．本研究内では深く関わらない部分であるが，これを前処理と呼び，画質が十分良く事前に処理行わなくてもオリジナルデータのまま b) のセグメンテーションを行うことが可能であるようならこの段階は通さなくても良い．b) のセグメンテーション自身を前処理と呼ぶ場合もある．具体的な前処理の例としては，ノイズの除去やその緩和，レンズ歪みの補正のためのカメラキャリブレーション，色の濃度値の変動範囲の調整やピンボケの修正，などが挙げられる．

b) セグメンテーション

例えば，物体の大きさを計りたいときには，事前にその物体を画像中の背景から切り離す処理をしておく必要がある．多くの物体が一枚の入力画像中に存在するような問題設定では，それらを事前に別々に分離する事が可能なようにしてなければならない．このように，処理の対象となる図形，あるいは画像中で次に処理すべき対象となる部分を切り出す処理を，図形の切り出し (セグメンテーション) と呼ぶ．本研究で大きく関わる画像中からの物体領域の自動取得もこの部分である．

c) 特徴量の抽出

画像中の（物体の）特徴量を得る処理である．画像のパターン認識では，最終的には決定の結果，クラス名を導くことになる．抽出された特徴の形として主に用いられるのは数値の組（ベクトル）である，これを特徴ベクトルという．近年，一般物体認識においては画像全体の特徴を求めるのではなく，画像中に局所的に存在する特徴量（局所特徴量）を扱ってクラス分類を行うことが主流になっている．ただし局所特徴量をそのままの形を用いて判別器等で分類を行うのは困難であるので，d) の決定（クラス名，ラベルの選択) の前に，局所特徴量を線形なベクトル表現に落とす処理を行うのが普通である．本研究では一般物体認識として２次元画像を扱うが，３次元形状を復元する処理などにおいては，入力装置の３次元的な位置・画像を撮影したカメラパラメータなどの情報の他に，複数

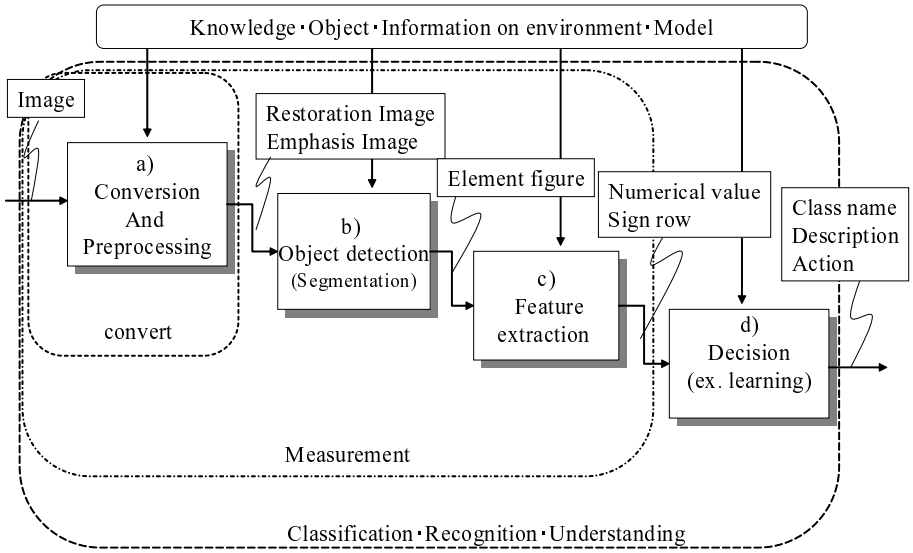


図 1.3: 画像認識の流れ

毎の画像における特徴点の対応付けが非常に重要な問題となってくる．対応付けが上手くいくように抽出する特徴の選択をする処理や，２つ以上の特徴点の関係を利用し対応を考えるなどの処理が必要になってくる場合もある．このような画像の分類や物体同定に必要となる一連の処理を特徴抽出と呼ぶ.

d) 最終的な決定

画像のパターン認識とは，与えられた画像に関して何らかの判断・決定を行うことを示している．これを計算機上で処理する場合には，分類を行うべき結果はクラス名であり，その中から最も適当なものを選択する，という形式になる．c) 特徴抽出で得られた特徴量を用いて，パターンマッチングや機械学習の手法などによって最終的な同定や分類を行う．ただし，認識として最終的な決定を行う事項は目的や画像上での状態などによって変わってくる．以下に主な３つ例を挙げる．

1) 画像全体の分類 画像情報からそれが分類されるべきクラスを判別し，決定を行う.

2) 部分画像の分類 基本的に 1) と同じであるが，こちらは１つの画像中に判定対象が複数存在する.

3) 対象物の決定 画像中に様々なものが存在している場合に，それらがどこにあり，なんであるかを判断する.

1.2.2 本研究の位置づけ

本研究で認識の対象とするのは，基本的に１画像中に物体が１つ存在するような画像を扱う一般物体認識であるので，決定の種類は 1) となる．だが更に今回我々はその中で複数候補提示下での性

能について着目する，つまりクラスとしての候補を複数個提示した場合毎での精度での検証を行う．

我々の研究では，一般物体認識について複数候補提示下での分類性能を向上させる事を最終的な目的としている．その枠組みを構成するために特に焦点を当てたのが上記の b) と d) であり，その 2 つの過程を要素として含んだ構成で，手法を提案している．

本論文では，他の研究で行われているようにデータに対してクラスを一意に推定する際の精度のみで性能を判断せず，候補数に従った精度傾向全般を見て，その性能向上を図るような方法について提案し，その手法の効果について実験による検証を行う．

1.3 本論文の構成

以降、本論文の構成は次のようになっている。第 2 章では，関連研究として一般物体認識手法について近年成果が出ているものを取り上げる，第 3 章では，今回データの分類をするのに用いる学習器である Support Vector Machine の複数クラスの分類への拡張，順位付けについて述べる．第 4 章では，用いる特徴量の詳細，そして事前準備として Caltech101 データセットに対して，様々な条件下で分類を行い，その結果について示す．また，この予備実験で得られた知見を提案手法の構成に役立てる．第 5 章では，提案する手法についてその意図および処理手順の概要を述べる．第 6 章では，提案手法を適用した実験結果を示しその考察を行う．最後に第 7 章にて，本論文のまとめと今後の課題を述べる．

第2章　関連研究

一般物体認識の研究として，近年特に成果を上げているものを挙げる．

2.1　一般物体認識に対する研究動向

一般物体認識へのアプローチとして，2000 年代前半は Region-based approach（領域ベースのアプローチ）が広まった．Region-based approach とは各画像に対して 1 つずつクラス分類をせず，データベースの中にある大量の画像に対して，画像毎に複数個の関連するキーワードを関連付けるようにする領域ベースのアプローチである．Barnard らは訓練画像に対して複数のキーワードが関連付けられている Corel 画像データベースセットを用い，テストデータを領域分割し，各領域に対しての自動的なアノテーションを行っている [25]．具体的には，テキスト翻訳の統計的機械翻訳の手法を応用し，領域分割で画像から抽出した全領域を片方の言語で書かれた文，画像に関連付けられた複数のキーワードをもう一方の言語で書かれた文とみなして，キーワードが関連付けられた画像を大量に用意する事により，画像間の対応の確率モデル（image-word translation model）を学習させ，部分領域へのアノテーションを成している．以上で述べた Translation model に代表される Region-based approach は，2003 年に Barnard らの論文 [25] が，ECCV（European Conference on Computer Vision）でベストペーパーを受賞するなど，当初は注目を集めていたが，近年の Generic Object Recognition の分野では主流な手法では無くなってきており，このアプローチをベースとした研究もあまり発表されていない．大きな理由として，認識の精度というものが最初に行う画像のセグメンテーションに因ってきてしまう事が挙げられる．領域の分割が上手くいかない画像には，そもそも効果を成さないとも言われており，汎用性の面からも避けられていると考えられる．

代わりになるものとして，近年一般物体認識に対するアプローチとして盛んになされているのが Part-based Approach（局所特徴量に基づくアプローチ）であり，Bag-of-keypoint [26] [30] や Constellation model [27] [28] など様々な手法での分類が試みられている．そのような局所特徴に基づいて画像データのクラスを推定するに当たって頻繁に使用されるのが Caltech101 データベースセットである．

今回，各研究の精度評価データベースとして，Caltech-101 での精度結果に従い調査を行った．Caltech101 はカルフォルニア工科大学の Fei-Fei によって収集され，作られた画像データベースである [11] [14]．表 2.1 で示されるような 101 の物体カテゴリと追加背景からなる 102 のクラスで構成され，31 から 800 の画像をカテゴリ毎に含んでいる．今日利用できるデータベースの中で最も多種多様なオブジェクトを含んでいるとされている．殆どの画像は 300 × 300 程度の中解像度のものであ

る．色，姿勢そして照明が画像ごとに変化しているので，分類は困難とされ，研究者達のなかでは非常にチャレンジングなデータベースとして，現在，評価画像データの標準となっている．

Caltech101 などのようなデータベースセットを対象とした一般物体認識関係の既存研究のアプローチにおける主な着眼点は，ほぼ以下に示す５つの例に分類される．

特徴量の提案・組み合わせ 研究例は多く，画像全体の特徴量や局所特徴量について様々な提案がなされている．近年では局所特徴量だけでなく，その関係性も用いる事が多くなっている．計算機の処理能力の向上と共に複雑な特徴量が処理可能になっているためである．

特徴量の表現手法 Bag-of-keypoints 等やその量子化手法についての改良など．

学習手法の工夫 一般物体認識に適したカーネル関数や確率・判別モデルそのものの提案など．

より妥当な学習データに最適化 学習領域の絞込みや認識領域の分離など．

大量の特徴量を上手く用いる 各特徴量に対するカーネル重みの自動調整など．

その Caltech-101 データベースセットを使用した画像分類技術での認識率の上位５位を示したものが表 2.2 である．ここにおける認識率とは訓練画像を 30 枚とし，テスト画像を残りとした結果の平均値となっている．2006 年の時点で最も高い認識率を出しているのは Zhang らの手法 [1] で，66.23%となっている．

本研究では，Lazebnik らの Spatial Pylamid Matching [2] の手法，そして Lazebnik らが提案した表現およびカーネルを類似度計算や学習に使用して訓練データ中の Region of Interest の学習を行っている 2007 年に提案された Bosch ら [6] の手法を提案アプローチ中に盛り込んでいる．よってこれらの手法の具体的な中身について次節より解説を行う．他の近年成果を出している手法については [24] に詳細がある．

表 2.1: Caltech-101 が持つクラス

BACKGROUNDGoogle	cougarface	ibis	rooster
Faces	crab	inlineSkate	saxophone
FacesEasy	crayfish	joshuaTree	schooner
Leopards	crocodile	kangaroo	scissors
Motorbikes	crocodileHead	ketch	scorpion
accordion	cup	lamp	seaHorse
airplanes	dalmatian	laptop	snoopy
anchor	dollarBill	llama	soccerBall
ant	dolphin	lobster	stapler
barrel	dragonfly	lotus	starfish
bass	electricGuitar	mandolin	stegosaurus
beaver	elephant	mayfly	stopSign
binocular	emu	menorah	strawberry
bonsai	euphonium	metronome	sunflower
brain	ewer	minaret	tick
brontosaurus	ferry	nautilus	trilobite
buddha	flamingo	octopus	umbrella
butterfly	flamingoHead	okapi	watch
camera	garfield	pagoda	waterLilly
cannon	gerenuk	panda	wheelchair
carSide	gramophone	pigeon	wildCat
ceilingFan	grandPiano	pizza	windsorChair
cellphone	hawksbill	platypus	wrench
chair	headphone	pyramid	yinYang
chandelier	hedgehog	revolver	cougarBody
helicopter	rhino		

表 2.2: Caltech101 に対する既存研究の分類性能

Rank	Paper	Method	Result(%)
1	[1]	SVM-KNN	66.23
2	[2]	Spatial Matching	64.60
3	[3]	Constellation model	63.00
4	[4]	Pyramid Matching	58.00
5	[5]	SVM	56.00

2.2 Spatial Pyramid Matching

Spatial Pyramid Matching [2] は Grauman らが提案した Pyramid Matching [4] を空間的な要素も考慮するように改良したものである．局所特徴量の集合によって物体を識別する Part-based アプローチを SVM に導入するには，画像間の類似度を計算するカーネル関数を定義する必要がある．Pyramid Match Kernel は 2 つのヒストグラムの bag 間の部分マッチングに基づいて類似度を計算するカーネル関数である．

2.2.1 Pyramid Matching

Pyramid Matching は画像のヒストグラムでの各解像度レベルでのマッチングに重み付けを行う事で得られる (図 2.1)，ヒストグラムの解像度レベルが同じ時に，2 つの点がグリッドの同一セルに落とされたら 2 つの点はマッチしたとみなされる．強い解像度レベルの時でのマッチングは，粗い解像度レベル時でのマッチングに比べより重要，つまり大きい重みをもつように設定される．ヒストグラムのレベルが l の時の各セルでのマッチング数が $I(H_X^l, H_Y^l) = I^l$ で表される時，カーネル関数は式 (2.2) となる．

$$I(H_X^l, H_Y^l) = \sum_{i=1}^{D} \min(H_X^l(i), H_Y^l(i)) = I^l \tag{2.1}$$

$$\kappa^L(X, Y) = I^L + \sum_{l=0}^{L-1} \frac{1}{2^{L-l}} (I^l - I^{l+1}) \tag{2.2}$$

$$= \frac{1}{2^L} I^0 + \sum_{l=1}^{L} \frac{1}{2^{L-l+1}} I^l \tag{2.3}$$

H_X^l と H_Y^l は解像度 l での画像 X と画像 Y のヒストグラムを示している．$H_X^l(i)$ と $H_Y^l(i)$ はグリッドの i 番目のセルに落ちた，各ヒストグラムにおける点（局所特徴）の数である．

2.2.2 Spatial Matching Scheme

前述した Pyramid Matching は高次元の空間特徴の集合の正確なマッチングを成しているが，同時に全ての空間情報を捨ててしまっている．Spatial Pyramid Matching では，空間情報を保持したまま，Pyramid Matching を成すように考えられた手法である．全ての特徴ベクトルを M の離散型へと量子ベクトル化し，同じ型の特徴だけマッチングを行う．各特徴のチャネルはそれぞれ特徴座標を持つ．ヒストグラムの解像度の各レベルにて，空間は 4 等分にされていき，その各グリッドで同一セルに落ちた点でのマッチング数の重み合計を各チャネルで取る（図 2.2）．カーネル関数 $K^L(X, Y)$ は，各特徴チャネルカーネルの合計となる（式 (2.4)）．

$$K^L(X, Y) = \sum_{m=1}^{M} \kappa^L(X_m, Y_m) \tag{2.4}$$

表 2.3: Performance on Spatial Pyramid Matching

	Weak Feature		Strong Feature	
L	Single-level	Pyramid	Single-level	Pyramid
0	15.5(± 0.9)		41.2(± 1.2)	
1	31.4(± 1.2)	32.8(± 1.3)	55.9(± 0.9)	57.0(± 0.8)
2	47.2 (± 1.1)	49.3(±.4)	63.6(± 0.9)	64.6(± 0.8)
3	31.4(± 1.2)	54.0(± 1.1)	60.3(± 0.9)	64.6(± 0.7)

2.2.3 評価結果

Lazebnik らは，今回提案したカーネル関数を SVM に適用する際に,「弱い特徴」,「強い特徴」という 2 つの種類の特徴量を用いて実験を行っている．弱い特徴は，エッジ点を二つの異なった解像度における 8 方向からの出力である 16 次元の特徴ベクトルを示している．強い特徴は，8 ピクセル空間において，グリッド上で計算された 16 × 16 のピクセル区画での SIFT 特徴量である．SIFT 特徴量は多重解像度分析とガウシアン微分を用い，更に特徴量をベクトルで表している．SIFT は後に詳しく解説するが，特定物体認識に特に有効な特徴量として知られており，近年，一般物体に対する認識においても，特徴表現を上手く線形など扱いやすい形に落とすことによって，よく機能する事が知られてきている．この実験では最終的に k-means を用いた Bag-of-keypoints（第 4 章に詳細）として局所特徴を量子ベクトル化しているが，最終的な特徴ベクトルの次元数は 200 としている．

Spatial Pyramid Matching について Caltech-101 を対象に実験を行った結果を表 2.3 に示す．学習データ数は 30 としている．Single-level は単純な Pyramid Matching を示しており，位置情報を考慮せずにヒストグラムの各レベルでのみマッチングを取るようなカーネル関数で実験を行った結果である．実験結果を見ると，Single-level よりも Pyramid Matching を行った方が，弱い特徴よりも強い特徴を取った方が良い結果が出る事がわかる．どのレベルまでマッチングを取ると有効であるかの検証であるが，結果では Single-level ではレベル 3 よりも 2 の方が良い精度が出ているが，Pyramid では，レベルが上がっても認識率が落ちなくなっている．これはカーネル関数の特性上，位置情報も考慮して各レベルでの重みをつけているために，誤ったマッチングで大きい重みがついてしまう事が避けられるようになっているためであると推測される．

図 2.3 に Caltech-101 に対して，この手法を用いた時の認識率が最も良かった対象物群と，最も悪かった対象物群について示した．人工の建造物や椅子・木などの認識率が良かったのは，特に形状の特徴がはっきりしているためであると考えられる．逆に蟻やワニ・ビーバーなどの動物の認識率が悪い理由として，それぞれ背景に擬態してしまっていて，特徴点が取りづらくなっているためである．

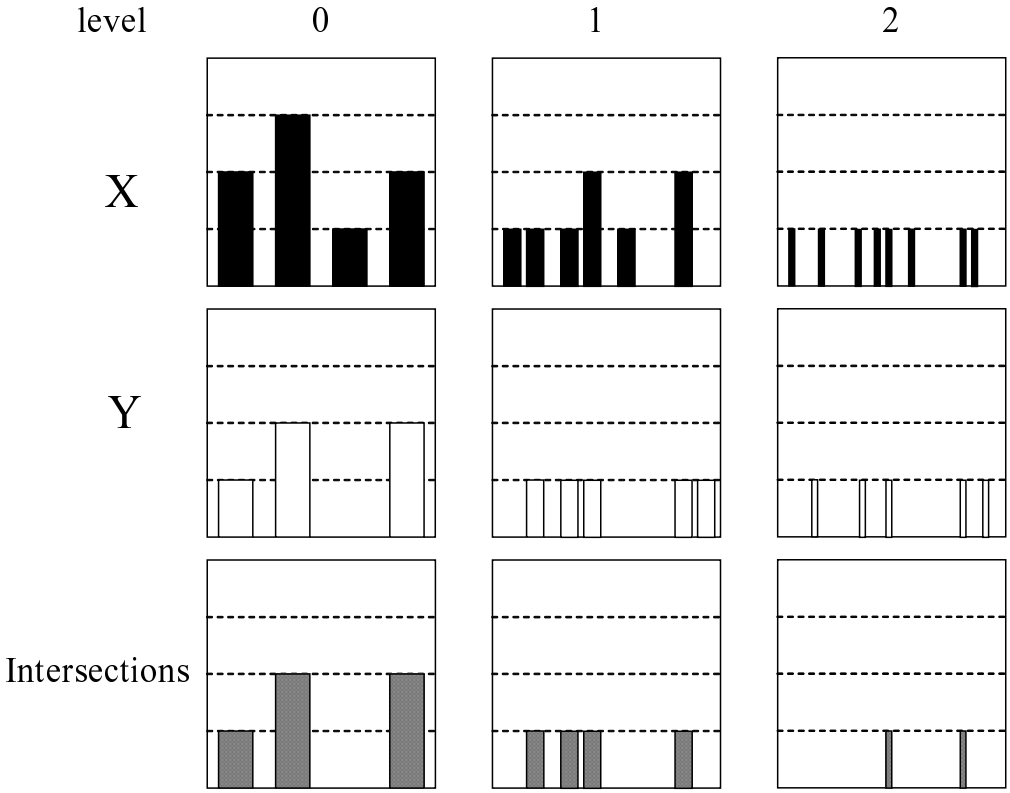


図 2.1: Pyramid Matching

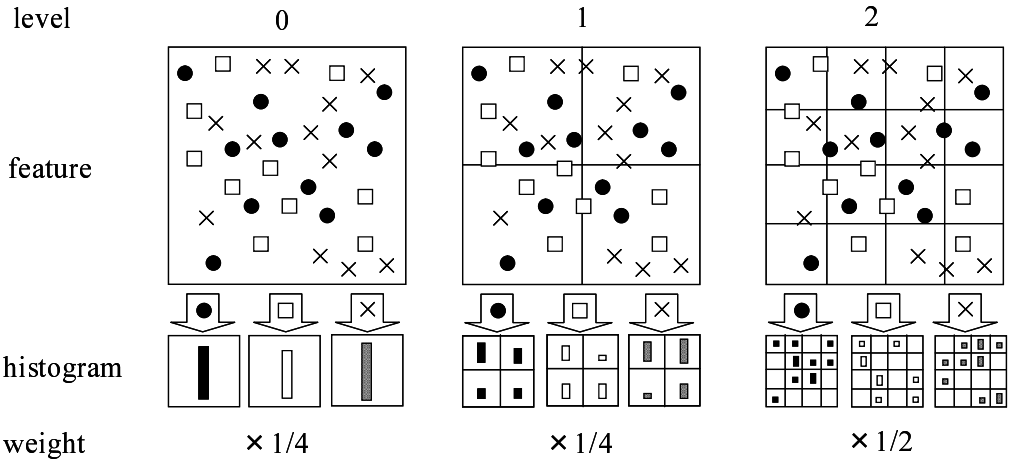


図 2.2: Spatial Pyramid Matching

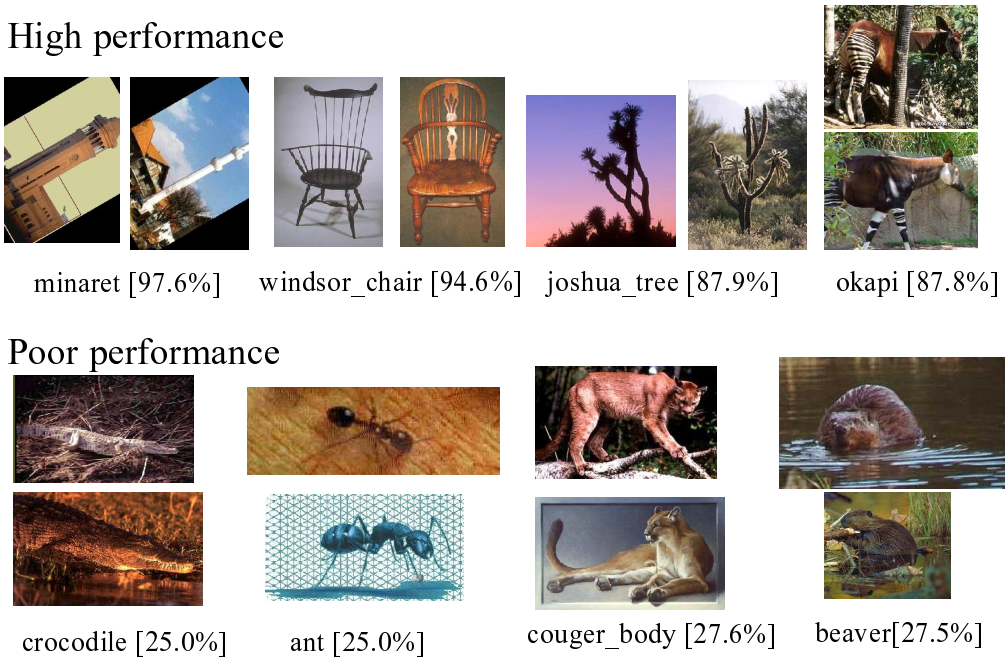


図 2.3: 分類精度の高いクラス・低いクラス（Spatial Matching）

2.3 学習領域の絞込み

一般物体認識をする上で用いる画像データは物体と一緒に背景などのそのクラスとは関係のないものが写っている事が多く，そういった所から局所特徴量を検出してしまうと実際分類をする際に度々邪魔になってしまうことがなる．

Bosch らは，あらかじめ学習画像群からその物体がある推定される関心領域 (Region of interest) を自動的に得て，その中で Spatial pyramid 表現 [2] を用いて SVM および random forests で判別モデルを構成し分類を行い，高い精度の認識に成功している [6].

探索対象の画像 i 中の Region of Interest を r_i に対し，そのクラスの訓練データ中で類似するような特徴量を持つ他の画像 j 群のサブセットがあると仮定する．その時，画像 j において，類似すると考えられる Region of Interestr_j を決め，以下のコスト関数を最適化することによって，画像 i 中の Region of Interestr_i を決定する．

$$L_i = \max_{r_j} \sum_{j=1}^{s} K(D(r_i), D(r_j)) \tag{2.5}$$

Region of Interest の探索を行った実験結果の例を図 2.4 に示す.

また Bosch らが Caltech-256 を対象として Region of Interest の探索を行った際のサブセット s(0〜4) の値に対する精度傾向を調査した結果を以下の表に示す.

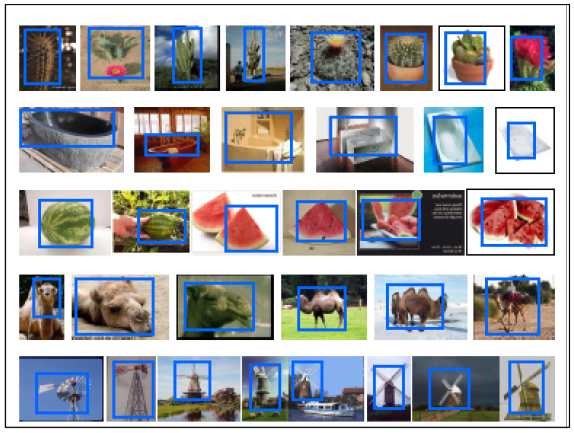


図 2.4: Region of Interest の探索 (Bosch)

表 2.4: サブセット s に対する精度傾向

s=0(探索なし)	s=1	s=2	s=3	s=4
38.7(± 1.3)	42.5(± 1.0)	42.9(± 1.0)	43.5(± 1.1)	42.9(± 1.)

2.4 その他のアプローチ

本研究の提案手法等とは直接関連しないが，一般物体認識分野において近年成果を上げている関連研究について簡潔に述べる．

2.4.1 SVM-KNN

Zhang らは，特徴量を判別器に入れ分類させる際，大きなデータセットについて SVM の識別性能を保持したまま効率よく学習させる手法として SVM-KNN[1] を提案している．SVM-KNN では，2 つの機械学習の手法，K-最近傍法と Support Vector Machine(SVM) を利用している．K-最近傍法は，他の分類器（例えば線形判別や決定木）などでは特徴空間の明確な構造を要求するのに対して，視覚物体認識における巨大なマルチクラスの性質を簡単に扱える．しかし有限のサンプリング中に存在するノイズによって高い偏りを度々発生させてしまうことがある．SVM は，少数のクラスとデータの中では他の分類器よりも汎化性能が高く，精度の良い分類が可能だと言われているが，大きなデータセット全体で SVM を学習させるのは時間がかかり，マルチクラスへの拡張が最近傍法ほど自然ではない．この 2 手法の長所を持ち越したものが，SVM-KNN であり，その特徴として「マルチクラスのデータセットを簡単に扱える」,「最近傍法や SVM よりも性能が良い」,「SVM では処理しにくいような問題でも有効性を維持出来る」,「様々な距離関数が使用できる」,「他の手法よりもデータセットでの実験でよい結果を出している」などが挙げられる．通常の SVM のように最初に全体で識別器を生成するのではなく，図 2.5 のように，まず一つのクエリが入ったら近傍の学習データを取り出し，その中で SVM を構成するというのが基本的な考え方である．Zhang らはこのような識別器を用い，一般物体認識において従来よりも高い精度を出している．用いている特徴量や距離基準については [32][33] で詳しく述べられている．

2.4.2 各特徴量におけるカーネルの重み学習

Varma らは，特徴としての記述子が複数種類あった場合，それを分類器に組み込む時の各記述子の識別力と汎化性のトレードオフを学習し，それを使用して多くの特徴量を学習して最終的な一般物体認識を行っている．そして様々な物体認識のデータベースセットにおいて近年で最高（条件はあるが Caltech101 で 80%程度）の精度を出している [7].

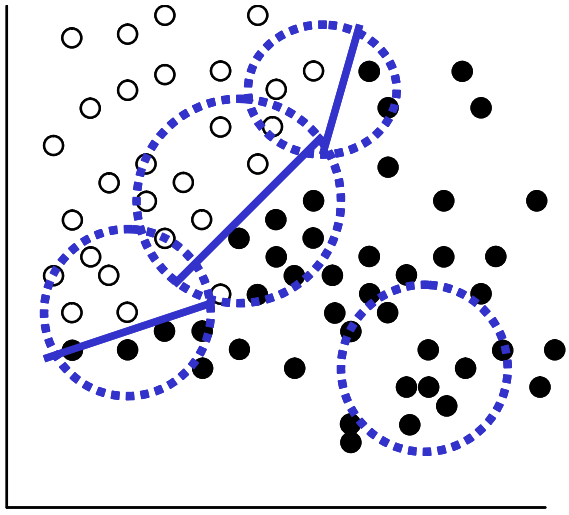


図 2.5: SVM-KNN

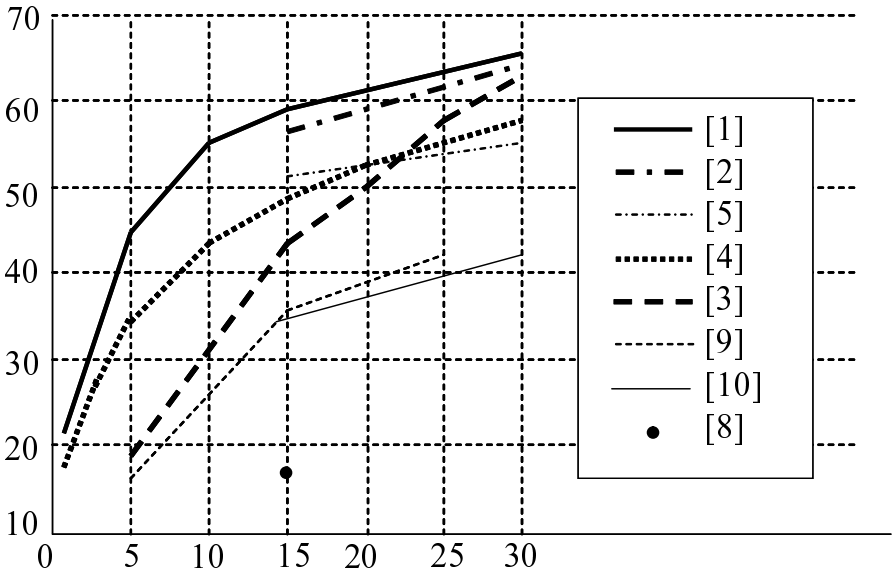


図 2.6: 学習データに対する精度

2.5 Caltech-101 に対する既存研究の分類性能

図 2.6 は，Caltech-101 データベースを対象とした，学習データ数に対する認識率をグラフにしたものである．今回，手法の内容についても記した SVM-KNN と Spatial Pyramid Matching が双方とも他の手法と比較して，どの学習データ数でも認識率が優れている事が分かる．また，元祖の constellation model である Fei-Fei らの手法 [8] での認識率は，学習データ数 15 に対して 17.7%であり，ここ数年での一般物体認識分野における研究の大きな進捗状況がわかる．

2.6 本章のまとめ

本章では，一般物体認識について，本研究と関連する手法および近年特に成果を出している研究の手法を中心に紹介してきた．

だがどの研究でも複数候補提示下での精度については着目しておらず，数種類の特徴量を組み合わせて使用していることが多い．本研究では複数候補提示下での分類性能の向上を最終的な目的としている．そのためにはまず複数候補提示下での精度が様々な条件かでどのような傾向を示すのかを調査する必要がある．第４章で行う予備実験では，シンプルな特徴量でどの程度までデータセットの２値問題を分類でき，多値的な分類を成功させられるかを示す．またそこで，根本的に分離できない２値問題を探索・発見する．更にパラメータの設定の方法での有効性の違いを示す．

その予備実験での分類については Support Vector Machine を用いる．次章からその SVM について，複数クラスの分類に用いる理由を含め，その多クラスへの拡張と順位付け問題まで詳細を述べる．

第3章　複数クラス分類問題のためのSVM

3.1　2クラス分類問題におけるSVM

本章では画像認識を行うために用いる分類器としての２値クラスSVMについて述べる．SVMは最適分離超平面の考え方を元にしてVapnikらにより考案され，近年になって非線形への問題への拡張としてカーネル学習法などと組み合わされてきた[12][13]．入力ベクトルを非線形に写像した高次元の特徴空間上で，１つの線形識別関数を構成する．そもそもSVMでのクラス分類の目的は，高次元特徴空間において，２値問題について,「正しく」分類するような，つまり汎化限界を最適にするような超平面を，計算量的に「効率良く」（１０万事例程のオーダーの問題も扱えるように）学習するような方法を提供する事である．高次元特徴空間において写像された分離超平面は，元の入力空間では局面になって，最終的に非線形な識別関数を構成する．

以上で述べたSVMは数多くある現在のパターン認識手法の中でも，最も性能が優秀な学習モデルの一つとして知られている．だが，SVMは２クラス識別問題において基本的には定式化されているので，画像認識など，多クラスの問題を扱うような識別器を構成するには更に工夫が必要となる．具体的な手法については後に述べるが，多数の２クラスSVMを組み合わせる事で，多クラスの識別を可能としている．そのアプローチも多数あるが，それらを統一化した手法も提案されている．

3.1.1　線形SVM

本節では，線形のSVMについて解説する．

SVMの最も単純なモデルとして，最大マージンクラス分類器がある．これは，線形分離可能なデータにのみ適用化であり，実世界のデータに用いる事は現実的ではない．しかし，構造が非常にシンプルで理解もしやすいアルゴリズムであり，これを基本として現在の実用的なSVMは構成されている．

ラベル付けがされたN個の特徴ベクトルとしてのデータがあり，それを$\{\mathbf{x}_i, y_i\}_{i=1}^{N}$とする．ここで，$\mathbf{x}_i$は$i$番目のデータの特徴ベクトル，$y_i$はそのデータについているラベルである．

学習データの集合Rが線形分離可能であるならば，関数のマージンが1である以下の式を満たす識別関数のパラメータ$\{\mathbf{w}, b\}$が存在する事になる．

$$\mathbf{w} \cdot \mathbf{x}_i + b \quad \geq \quad 1 \ \ (if\ y_i = +1) \tag{3.1}$$

$$\mathbf{w} \cdot \mathbf{x}_i + b \quad \leq \quad 1 \ \ (if\ y_i = -1) \tag{3.2}$$

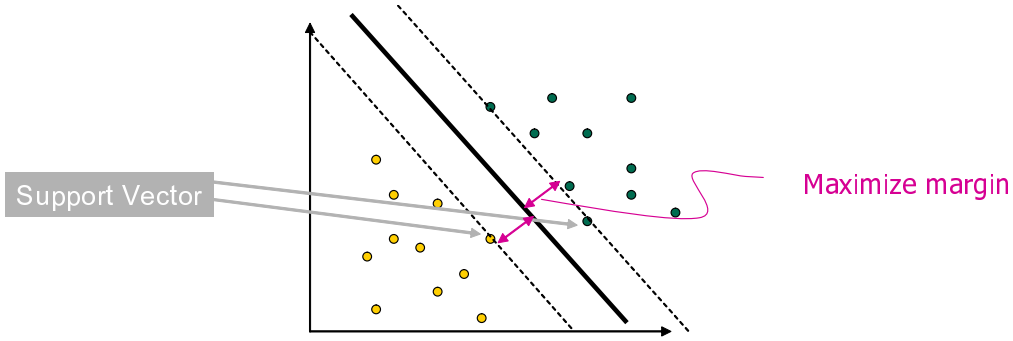


図 3.1: マージン最大化

上記の境界を関数とした 2 枚の超平面で学習データは完全に分離されており，この間にはデータが存在しないという事を示している．

そのような超平面の場合，識別する平面とこれらの超平面との距離をマージンと呼ぶ．この時のマージンが大きい方が，汎化性能が高い事が知られている．実際に上記の平面間のマージン距離 γ を求めると以下のようになる．

$$\gamma = \frac{|\mathbf{w}^T\mathbf{x}' + b - 1|}{\| \mathbf{w} \|} + \frac{|\mathbf{w}^T\mathbf{x}' + b + 1|}{\| \mathbf{w} \|} \tag{3.3}$$

$$= \frac{|-1|}{\| \mathbf{w} \|} + \frac{|1|}{\| \mathbf{w} \|} = \frac{2}{\| \mathbf{w} \|} \tag{3.4}$$

SVM の最大の特徴はこのマージン γ が最大となる分離平面を求める「マージン最大化」にあり，学習データの近くをぎりぎり通るのではなく，出来るだけ余裕があるように分離するような識別平面が求められる (図 3.1)．これによって，SVM は高い汎化能力（未学習データに対しても高い識別性能）を発揮できる．

このマージンを最大化するためには $\| \mathbf{w} \|$ を最大化すればいい事がわかる．しかし一般には，訓練サンプル集合を誤りなく分けるパラメータは一意には決まらない．

ここから，マージン γ を最大化するようなパラメータ $\mathbf{w}$ と b を求める事は，結局以下の制約が付いた最適化問題を解く事と等価になる．

$$\text{minimise}_{\mathbf{w},b} \qquad L(\mathbf{w}) = \frac{1}{2}\| \mathbf{w} \|^2 \tag{3.5}$$

$$\text{subject to} \qquad y_i(\mathbf{w}^\mathbf{T}\mathbf{x}_i + b) \geq 1 \quad i = 1, ...l \tag{3.6}$$

超平面 $(\mathbf{w}, b)$ はこの最適化問題の最適解となる．

この最適化問題は二次計画問題であり，対応する双対問題へと変換する方法などによって解かれている．この主問題のラグランジアンを考えると，目的関数は以下のようになる．

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2}\| \mathbf{w} \|^2 - \sum_{i=1}^{l} \alpha[y_i(\mathbf{w}^T\mathbf{x}_i + b) - 1] \tag{3.7}$$

$\alpha_i \geq 0$ はラグランジュ乗数である．ここで対応する双対問題を得るために $\mathbf{w}$ と b に関し偏微分を行い定常性を仮定する．その時，停留店では以下の関係が成り立つ．

$$\frac{\partial L(\mathbf{w}, b, \alpha)}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^{l} y_i \alpha_i \mathbf{x}_i = \mathbf{0} \tag{3.8}$$

$$\frac{\partial L(\mathbf{w}, b, \alpha)}{\partial b} = \sum_{i=1}^{l} y_i \alpha_i = 0 \tag{3.9}$$

この関係を上の主問題としての目的関数，式 (3.7) に代入すると，以下のような関係がわかる．

$$\begin{aligned} L(\mathbf{w}, b, \alpha) &= \frac{1}{2} \sum_{i,j=1}^{l} y_i y_j \alpha_i \alpha_j \mathbf{x}_i^T \mathbf{x}_j \\ &- \sum_{i=1}^{l} y_i y_j \alpha_i \alpha_j \mathbf{x}_i^T \mathbf{x}_j + \sum_{i=1}^{l} \alpha_i \qquad (3.10) \\ &= \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} y_i y_j \alpha_i \alpha_j \mathbf{x}_i^T \mathbf{x}_j \qquad (3.11) \end{aligned}$$

ここから以下のような命題が得られる．

$$\text{maximise } \mathrm{W}(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} y_i y_j \alpha_i \alpha_j \mathbf{x}_i^T \mathbf{x}_j \tag{3.12}$$

$$\text{subject to} \quad y_i(\mathbf{w}^{\mathbf{T}} \mathbf{x}_i + b) \geq 1 \;\; i = 1, ...l \tag{3.13}$$

パラメータ α^* がこの二次最適化問題の最適解の時，重みベクトル $\mathbf{w}^* = \sum_{i=1}^{l} y_i \alpha_i^* \mathbf{x}_i$ は，マージンが $\gamma = \frac{2}{\|\mathbf{w}\|}$ である最大マージンをもつ超平面を実現する．

ここで b の値は双対問題には存在せず，解 b^* は制約から以下のように解かれる．

$$b^* = -\frac{\max\limits_{y_i=-1}(\mathbf{w}^T \mathrm{x}_i) + \min\limits_{y_i=1}(\mathbf{w}^* \mathbf{x}_i)}{2} \tag{3.14}$$

3.1.2 ソフトマージン最適化

前節で解説した最大マージンクラス分類器としての線形 SVM は，SVM の概念として非常に重要であるが，前述したように実世界の問題に対して適用をするのは，線形分離の可能性から言って困難である．最大マージンクラス分類器は完全に無矛盾な仮説，つまり訓練誤差が無いような識別関数を常に生成しようとする．これはマージンに依存した汎化誤差の限界を考慮しているためである．

マージンに依存するような設計にすると，全体での少数のノイズの影響が大きくなり，ノイズが散在する実世界のデータにその分類器をそのまま適用することは，信頼性の面から言って利用価値が低い．そして線形分離不可能の問題に対して，主問題の実行領域が空であり双対問題は限界の無い目的関数を持つことになり，最適化問題を主問題として解く事が出来ない．

そこで多少の誤識別について外れ値を設けることで許容するような尺度を用いて分類器を設計し直す方法がある．これはソフトマージンの手法と呼ばれている．

その中の代表的な手法の一つ (C-SVC) について示す．この手法では最適化をするべき SVM の問題は次式のようになる。

$$\text{Minimize} \qquad L(\mathbf{w}) = \frac{1}{2}||\mathbf{w}||^2 + C\sum_{i=1}^{l}\xi_i \tag{3.15}$$

$$\text{subject to} \qquad y_i(\mathbf{w}\cdot\mathbf{x}_i + b) \geq 1 - \xi_i,\ \xi_i \geq 0,\ C > 0 \tag{3.16}$$

C が識別誤差に対して与えらるペナルティであり，ξ は緩めの度合いを示している．この式は今までの線形 SVM と同じように解くことが出来る．

本研究の予備実験でもソフトマージンの手法を組み込んでいる．そのパラメータの設定による精度傾向についての調査も行っている．

3.1.3 カーネルトリック

線形分類器を用いて，データの形が非線形であるような事象を学習するには．非線形な特徴集合を選択して，データを異なった表現に書き換える必要がある．

線形分類器の特性の一つとして，双対問題として表現できるという事が挙げられる．つまり仮説が学習データの線形結合で表現が可能であるということである．ここから，決定規則がテストデータと学習データの内積を用いて表現できることがわかる．ここでもし，特徴空間内の内積 $(\Phi^T(\mathbf{x}_i)\Phi(\mathbf{x})$ が元の入力データでの関数として直接計算が可能ならば，非線形学習マシンを構築する固定した非線形写像によりデータを特徴空間に変換し，特徴空間内での識別に線形マシンを使用するという段階を併合することが出来るようになる．このような直接計算を行うために用いる以下のようなカーネル関数を用いて非線形データを計算量を削減しつつように扱えるようにする手法はカーネルトリックと呼ばれている．

$$K(\mathbf{x}, \mathbf{z}) \quad = \quad \Phi(\mathbf{x}) \cdot \Phi(\mathbf{z}) \tag{3.17}$$

このような性質を持ったカーネル関数を用いれば，非線形のデータを SVM に適用するための特徴ベクトルの写像および内積によって起こる膨大な演算が 1 回カーネルを計算するだけですんでしまう．

3.2 複数のクラスを対象としたSVM

前章で扱ったSVMは基本的には2値の識別問題を対象として定式化されている．しかし実世界の問題では2クラスより多いようなクラスを識別するような分類問題に直面する場合も多い．そのため2値問題の複数クラス問題へ拡張する手法が必要になる．

多クラス分類とは，各データが予め用意されたクラスのどれかに分類をすることを言う．データ $\mathbf{x}$ が条件としてあったときのクラス分類の事後確率 $P(i|\mathbf{x})$ が最大となるようなクラスを選択するように通常設計する．事後確率を近似するために，正規混合分布・カーネル k-nearest neighbor やカーネル密度推定などのパラメトリック・ノンパラメトリックな手法が用いられる．

しかし2値分類問題に限ると，SVMのようなマージン制御に従った高い汎化性能を持つような分類器を用いる事が出来る．前述したように，基本的に多クラスの識別問題を2値分類器で扱うためには，2クラスの判別モデルを組み合わせる事になる．これは複数クラス問題を2値問題のセットへと縮小させる考えを前提として提案されている．SVMの基本的な構成のまま，多値の判別関数を直接構成する手法（Multiclass-SVM [21] [23]）も提案されているが，今回は組み合わせで構成する手法を対象とする．これは，各クラス間の分類問題が同じ仮説の上に成り立っているとは限らないこと，個々の問題においての詳細なデータの解析が難しい，ことが主な理由である．

3.2.1 One-versus-All

全クラス数が k で，$\mathbf{x}$ を入力としたとき，k 個の各SVMの識別関数 $g_j(\mathbf{x})$ は，

$$g_j(\mathbf{x}) \quad = \quad \mathbf{w}_j\mathbf{x} + b_j \tag{3.18}$$

と定義されるとする．ここでは単純化のため対象のクラス j とその他のクラスの間に分離が可能であるとすると仮定している．その時，もし入力 $\mathbf{x}$ がクラス j のラベルを保持していれば，式 (3.19) が成立する事がわかる．

$$\mathbf{w}_j\mathbf{x} + b_j \quad \geq \quad 0 \tag{3.19}$$

各識別境界は上式の境界で与えられる（図 3.2 左）.

以上の事を利用し，この識別関数において各データが正しくその保持クラスに分離される状況においては，マルチクラスへの適用も出来ている．この手法は従来のSVMをいわばそのまま用いている最も単純な手法であり，計算量も少なく実装も簡単である．しかし，クラス数が増加すれば確実に分離出来る状況はなくなり，上で述べた仮定が成り立たなくなる．すると，誤って他のクラスに属してしまう場合だけでなく，どこかのクラスに確実に属している前提があっても，そのデータがどこのクラスにも属していないという結果が出てしまうこともある．これは全体で分離性を仮定しているため，データ量が多くなるとどうしても無理が出てきてしまうからであると考えられる．また，各クラスで別々に識別問題を解いているので，そこで出力された結果全体に対する最適解の保証は無い．

そこで，同じ1対他の分類器を構成する考え方であるが，部分的な分離可能性を仮定する手法について述べる．ある任意のクラス j とそれ以外のクラスの間に分離可能性を仮定したとき，上で述

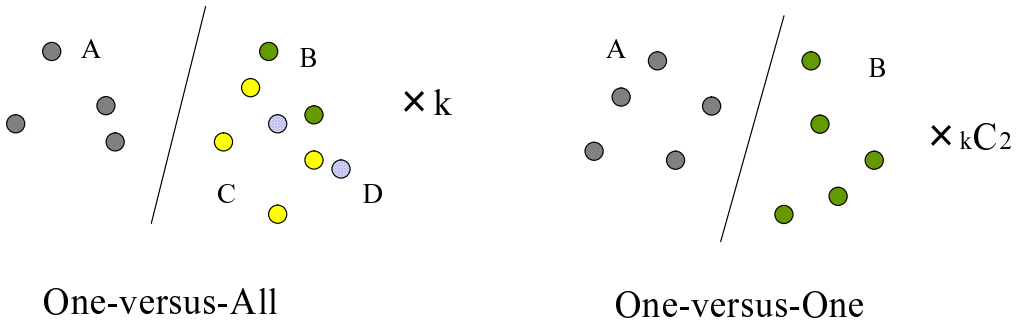


図 3.2: One-versus-All と One-versus-One

べた通り識別関数は式 (3.18) となる．もし入力 $\mathbf{x}$ がクラス j のラベルを保持していれば，他の全てのクラス $j'(j' = 1...k, j' \neq j)$ の識別関数と比較して，式 (3.20) が成立する事がわかる．

$$\mathbf{w}_j\mathbf{x} + b_j \quad \geq \quad \mathbf{w}_{j'}\mathbf{x} + b_{j'} \tag{3.20}$$

このとき，クラス j と j' の間の識別境界は $(\mathbf{w}_j - \mathbf{w}_{j'})\mathbf{x} + (b_j - b_{j'}) = 0$ となる．以上のような識別境界を構成すれば，部分的な分離性を仮定している事になる．これは，各識別関数における入力データに対する出力マージンの値を比較し，クラスを決定している事となる．結果的に誤っていても，全てのデータはどこかのラベルに属することになる．k 個の識別関数を構成し，その各々に全体のデータ数 N を入力して学習する事になるので，計算量的には，クラスが多い問題にも対処できるようになっている．各 2 値 SVM の識別関数をデータ数 n で構築するオーダーは，大体どのような実装においても $O(n^3) \sim O(n^4)$ 程度であると知られている．

このような任意のクラスとその他全てのクラスで構成された識別関数を利用し，2 値クラス SVM を多クラス問題対応させるような手法を One-versus-All(または One-versus-Rest) のアプローチと呼ぶ．

3.2.2 One-versus-One

前節で述べた One-versus-All の他に，Hastie らによって提案された，One-versus-One（または All-Pair）という 2 値問題の多クラスへの拡張手法がある [16]．これは任意のクラスのペアどうしでの学習データを用いて識別関数を構成し，それを多クラス分類問題に利用している．

任意の 2 つのクラス $j_1, j_2(j_1, j_2 = 1...k, j_1 \neq j_2)$ の間に分離可能性があると仮定する．その時，全ての 2 クラスの組み合わせ $\frac{k(k-1)}{2}$ 個の識別関数が以下の式で構成される．

$$g_{j_1,j_2}(\mathbf{x}) \quad = \quad \mathbf{w}_{j_1,j_2}\mathbf{x} + b_{j_1,j_2} \tag{3.21}$$

この時，もし入力 $\mathbf{x}$ がクラス j のラベルを保持していれば以下が成り立つ．

$$\mathbf{w}_{j_1,j_2}\mathbf{x} + b_{j_1,j_2} \quad \geq \quad 0 \tag{3.22}$$

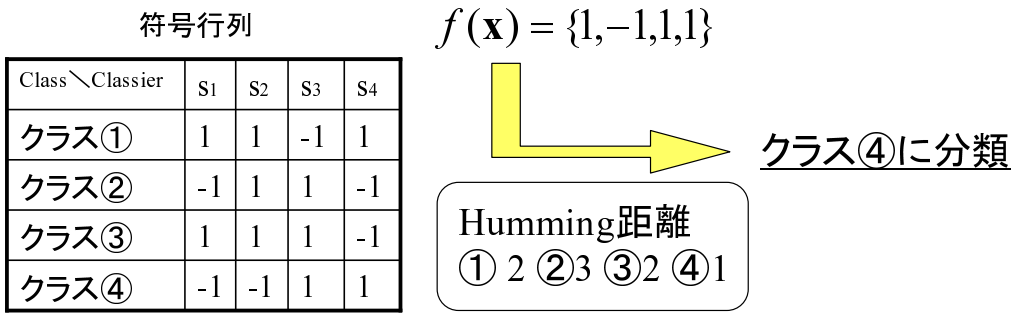


図 3.3: Error Correcting Output Codes

各識別境界は上式の境界 $\mathbf{w}_{j_1,j_2}\mathbf{x} + b_{j_1,j_2} = 0$ で与えられる．この手法では，ペアというそれぞれの仮説において分離性があるとして，学習器を構築，その結果を統合（多数決など）して扱う，といったものである．各 2 クラスの学習データ同士で部分的に識別関数を構築するため，非常にデータ量が小さくなり，完全分離可能性が高まる．また，2 値分類器の元々の意味として持つ，2 つの仮説を正しく分離するというものとも一致していると考えられる．$\frac{k(k-1)}{2}$ の識別関数を構成し（図 3.2 右），その各々にその学習器に用いるクラスのデータ $s(s = \frac{N}{k})$ を 2 つ入力して学習するので，計算量の観点でみると，多くのクラスがある場合には大きくなってしまうが，学習データの数が少ない場合には，各分類器を高速に作成できる．

しかし，最終的なそのクラスのラベルを結果として出力するために，その識別関数を構築する 2 ペア以外のクラスを持つデータもそこに当てはめるので，その他の識別関数においてどのような分離が行われるかの予測が困難である．また，各ペアで仮説を立て，そこで識別関数を構築するので，最終的な結果が決定不能な領域に属するデータというものがしばしば出てしまう．複数の学習器での出力を多数決で取るという処置は単純で実装も容易であるが，最適性の観点からは裏づけがない．

3.2.3 Error Correcting Output Codes

複数クラスの問題を扱うためのより一般的な手法が，Dietterich らによって与えられている [17]．それは，各クラスのラベルと分類器によっての出力が関連付けされた符号行列 $M \in \{-1, +1\}^{k \times l}$ を元に考えられている．これはラベルが y であるクラスが，s 番目の分類器によって出力される値 f_s を $M(y, s)$ にマッピングしている事を示している．ここで入力 $\mathbf{x}$ が与えられたとき，行列 M における y の仮説の各値が $f_1(\mathbf{x})...f_l(\mathbf{x})$ に最も近いような（例えば Hamming 距離などを取って）ラベル y を求める．この処理をデコードと呼ぶ．その y が入力 $\mathbf{x}$ のラベルとして予測される．

具体例を図 3.3 で示す．ここでは距離として Humming 距離を用いている．最終的に距離が最も近いクラスに決定されていることがわかる。この例では 4 つのクラスへの分類をする際に 4 つの分類器を使用しているが，更に少なくすることが出来る．

以上のような考え方を誤り訂正符号行列，Error Correcting Output Codes(ECOC) と呼ぶ．この

手法の利点として，構築する識別平面，つまり分類器の数に制約がないことである．もちろん One-versus-All や One-versus-One の手法のように分類器の数を多くすれば頑強性は高まる．しかし同時に計算量も増える．そのような精度と計算量のトレードオフを図るための手法といって良い．

3.2.4 多クラス SVM の一般化

以上のようなアルゴリズムを一般化統合する [18]．行列の各分類器によって得られる値 f_s を拡張し，符号行列 $M \in \{-1, 0, +1\}^{k \times l}$ を定める．ここで $M(y, s)$ が 0 の場合は，分類器によってそのデータがどのように分類されているかを考慮しなくて良いという事である．分類器 $s = 1..l$ について，その中の分類アルゴリズムを使用して，全学習データデータ $\mathbf{x}_i$ によって，$M(y_i, s)$ にラベル付けを行う．

例えば，One-versus-All での多クラス SVM は，M は全対角要素が+1 であり，他の要素が 0 であるような $k \times k$ 行列となる．また，One-versus-One の手法では，M は各列が別個のペアクラス (j_1, j_2) の分類器である $k \times \frac{k(k-1)}{2}$ 行列となる．ペア (j_1, j_2) での分類器を $s(j_1, j_2)$ とすると，$M(j_1, s(j_1, j_2)) = +1, M(j_2, s(j_1, j_2)) = -1$ となり，また j_1, j_2 以外の全てのクラス j_{other} に対して，$M(j_{other}, s(j_1, j_2)) = 0$ となる．

3.2.5 多クラス SVM の順位出力への拡張

前節まで，SVM の多クラス化について議論してきた．今回，我々は一般物体認識において，複数候補を提示した場合の精度を確認する．その為に，多クラス SVM での出力を最もそれらしいラベルのみならず，もっともらしさの基準を決定し，それを元に順位推定をしなければならない．先ほどの符号行列と各々の分類器によって生成された個々のデータに対する出力値が小さい程，そのデータのクラスであるというように順位付けを行う．

さて，One-versus-All の元の形を符号行列とデコードで記述するには，前述したように全対角要素が $+1$ であり，他の要素が 0 であるような $k \times k$ 行列である M を用い，個々のデータに対する各分類器での出力値 ± 1 の以下に定められる Humming 距離 d_H を求める．

$$d_H(M(r), f(\mathbf{x})) \quad = \quad \sum_{s=1}^{l} \frac{1 - \mathrm{sign}(M(r, s) f_s(\mathbf{x}))}{2} \tag{3.23}$$

通常のクラス分類では各データについて以下に示される y にクラスが決定される．

$$y \quad = \quad \arg\min_r d_H(M(r), f(\mathbf{x})) \tag{3.24}$$

また順位出力の形では，i 番目にそれらしいクラスは，$i-1$ 番目にて出力されたクラスを除いた集合を用いて上述した式で決定される．しかし前述したように One-versus-All は全体で分離性を仮定しているため，どこかのクラスに確実に属している前提があっても，そのデータがどこのクラスにも属していない結果が出てしまう．また，符号行列が対角成分にしか値を持たないので，分類器ラベルの出力が ± 1 しか持たないと順位付けがほぼ不可能になる．

部分的な分離可能性を仮定した One-versus-All を符号行列とデコードで記述するには，先ほどと同じ符号行列と Humming 距離を用いた場合，個々のデータに対する各分類器の識別関数値の中で最も大きいものを+1 とし，その他を 0 とした出力を使用する．この操作を行うとどこかのクラスには必ず属することになるが，そのクラス以外の優位性が全く出なくなるので順位出力は不可能になる．そこで識別関数の値をそのまま，個々のデータに対する各分類器の出力として用いる．識別関数値を原形で用いると前述した符号行列との Humming 距離を取ることが出来なくなるので，Humming 距離の代わりに以下に示す損失関数に基づいた距離 d_L を用い，その値を元にクラスを決定する．関数は以下の式のような指数ベースの損失関数などが代表として挙げられる．

$$d_L(M(r), f(\mathbf{x})) \quad = \quad \sum_{s=1}^{l} \frac{1}{1 + e^{2M(r,s)f_s(\mathbf{x})}} \tag{3.25}$$

同じように，各データについて以下に示される y にクラスが決定される．

$$y \quad = \quad \arg\min_{r} d_L(M(r), f(\mathbf{x})) \tag{3.26}$$

このようなデコーディング手法を loss-based decoding と呼ぶ．

One-versus-One を符号行列で一般化するには，前述したように，M は各列が別個のペアクラス (j_1, j_2) の分類器である $k \times \frac{k(k-1)}{2}$ 行列となる．その符号行列と各データの出力の間で式 (3.23) で表される Humming 距離を取る．One-versus-One では，部分的な分離性を仮定しているために，このアプローチでも順位出力が可能である．また，One-versus-All で行ったように各データでの出力に識別関数値をそのまま用いて，符号行列との間で損失ベースの距離を取って順位出力に変換する事も出来る．

3.3 本章のまとめ

本章では，複数クラスの分類を行う Support Vector Machine について述べた．基本的な線形 SVM の概念から，それを組み合わせた多クラス分類，また順位付けと一般化について解説を行った．

次章では本研究の目的のための，一般物体の認識について，最終的な提案手法に必要となる条件・パラメータの設定等を得るための予備実験を行った結果を示す．そこで，本章中で述べた各 SVM の構成手法での実験結果についてそれぞれ述べる．その他にも様々な特徴量やパラメータの条件下で実験を行った結果を示し，最終的に提案するアプローチへの材料として使用する．

第4章　Caltechデータセットに対する予備実験

今回，我々は予備実験として Caltech101 のデータベースセット [11] [14] に対して，多クラス SVM を適用し，候補数にしたがった精度がどの程度になるかを検証する．Caltech101 データベースセットについては第２章で述べた通りである．このデータセットにおける認識率とは訓練画像を 30 枚とし，テスト画像を残りとした結果の平均値となっている．2006 年の時点で最も高い認識率を出しているのは Zhang らの手法 [1] で，66.23%となっていた．また 2007 年には，関連研究で触れたようなカーネルの重みを推定して組み合わせる手法や，学習領域を絞る手法などを用いて，80%以上の精度を出す研究も発表されてきている [7][6].

4.1　予備実験の目的

序章でも書いたように，本研究の目的は複数候補を提示した場合での分類性能の向上である．その場合最適化したいのは，精度が 100%近くに収束する候補提示数を最小化することであり，テストデータに対して一意にクラスを推定した際の精度が高い手法を用いることが必ずしも良いとは限らない．更に今回予備実験として検証を行うのは，特徴量や学習領域を上手く選択した場合での結果を得るというよりは，シンプルな特徴でどの程度まで分類可能なのかということを確かめ，その特徴量ではどのようにしても分離が出来ないクラス間の問題を発見し，根本的な問題を提示する事にある．また各 SVM 構成手法での候補提示数に従った優位性や，パラメータの設定について適切な方法を発見するという事も目的としてある．従って今回の予備実験では用いる特徴量やその組み合わせについてはあまり詳細な議論をしないが，２つの単純な Bag-of-keypoints の特徴表現については多少実験を行い，優位性について示す.

4.2 局所特徴量 (SIFT)

認識を行うための局所特徴量として，今回はSIFT を用いる．SIFT(Scale-Invariant Feature Transform) とは，1999 年に Lowe によって提案された，特徴点の検出・局所特徴量の記述を行うアルゴリズムである [15]．この局所特徴量は，画像の回転・スケール変化・照明変化に対してロバストに働く事が知られており，近年の物体認識において，最も用いられている特徴記述の一つとなっている．SIFT は以下に示す 4 段階の処理からなる．

1. **Scale-Space extrema detection**　平滑化画像の差 (Dfference of Gaussian) を計算し，注目画素の近傍での極値点から Keypoint を検出し，その中で Keypoint の値が最も高くなるようなスケールパラメータを求める．

   Difference of Gaussian（以下 DoG) 画像は以下の式で求まる．

$$D(u,v,\sigma) \quad = \quad (G(x,y,k\sigma) - G(x,y,\sigma)) * I(u,v) \tag{4.1}$$

   これはスケールの異なったガウス関数と入力画像 $I(u,v)$ を畳み込んで平滑化処理を行った画像の差分である．$G(x,y,\sigma)$ はガウス関数であり，以下の式で定義される．

$$G(x,y,\sigma) \quad = \quad \frac{1}{2\pi\sigma^2}\exp(-\frac{x^2+y^2}{2\sigma^2}) \tag{4.2}$$

   σ_0 から k 倍ずつ拡大したスケールのガウス関数で行うのであるが，スケールが増加し続けると最終的に処理しきれない領域が発生してしまい，更に計算量も大きくなっていく．

   そこで画像のダウンサンプリングを利用する．スケール σ で平滑化された画像を $L(\sigma)$ とする．ある画像をスケールの初期値 σ_0 で平滑化処理を行い，画像 $L_1(\sigma_0)$ を得て，スケールを k 倍して DOG 処理を行っていく．その時，得られた画像 $L_1(2\sigma_0)$ 画像とその画像を 1/2 にダウンサンプリングした画像 $L_2(\sigma_0)$ の間には $L_1(2\sigma_0) = L_2(\sigma_0)$ の関係が成り立つとされる．

   このような関係を利用し，SIFT のアルゴリズムでは画像をダウンサンプリングすることでスケール σ の最大値を制限して計算量の増加や処理できない領域の拡大を防いでいる．また，σ の連続性も保証している．

   この DOG 画像群から極値を検出して最終的なキーポイントのスケールを求める．基本的に，DOG の値が大きく異なるような結果となるスケール σ はその間で（主にエッジ情報の）大きな変化があったと考えられるので，DOG 画像から σ に従った極値を検出してキーポイントとスケールを決定する．

   まずキーポイントの検出であるが，注目画素があったときに，両端のスケール σ での隣接 9 近傍 (図 4.1) と同スケールでの 8 近傍の合計 26 近傍の DOG 値の比較を行う．その時，極値であったらその画素をキーポイント候補点として検出する．スケール σ が小さい値の DOG 画像

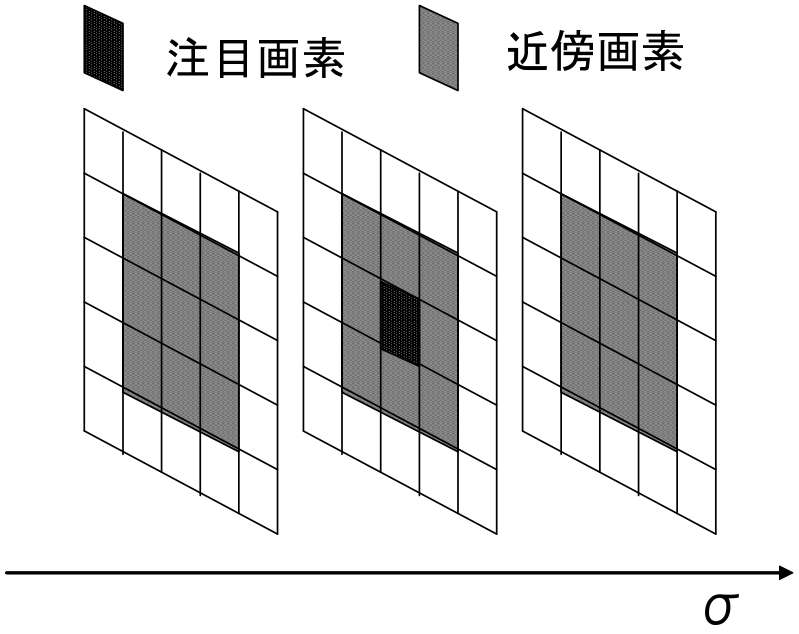


図 4.1: DOG 画像からの極値検出

から検査を行い，一度検出された画素が再度異なる DOG 画像で極値となってもキーポイント候補点として検出しない．

次にキーポイント候補点各々におけるスケール σ の決定であるが，その候補点としての画素それぞれで σ を変化させた際の極値を採用する．ダウンサンプリングした結果 1/2 のサイズになった画像では σ は比例して 1/2 倍となる．このようにして SIFT は拡大・縮小への不変性を得ている．

2. Keypoint localization 1 で行ったキーポイント候補点の検出では単純に極値の画素だけを得ているので，周りの傾向しだいでは，DOG 出力がさほど高くなくてもキーポイントと判断されることがある．またエッジ上の点が非常に検出されやすくなってしまう．そのようなものをなるべく排除するために，キーポイント候補点の中から，主曲率やサブピクセル位置推定を行って求めたコントラスト値によって更に安定した候補だけに絞り込む．

２次元ヘッセ行列 **H** を以下のようにして得る．

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \tag{4.3}$$

ここでの D は導関数であり，DOG 値の 2 値微分を行い得られる．この行列の対角成分の和

Tr(**H**) の二乗と Det(**H**) の比について，以下のような閾値処理の式を通し，条件を満たさないキーポイント候補点は削除する．

$$\frac{\mathrm{Tr}(\mathbf{H})^2}{\mathrm{Det}(\mathbf{H})} < \frac{(\gamma_{th}+1)^2}{\gamma_{th}} \tag{4.4}$$

この閾値 γ はヘッセ行列の各固有値の比率としての意味を持つ．つまり，γ をヘッセ行列の第一の固有値と第二の固有値の比率 ($\alpha = \gamma\beta$) とすると，$Tr(H)$ の二乗と $Det(H)$ の比は以下のように計算できる．

$$\mathrm{Tr}(\mathbf{H}) = D_{xx} + D_{yy} = \alpha\beta \tag{4.5}$$

$$\mathrm{Det}(\mathbf{H}) = D_{xx}Dyy - (D_{xy})^2 = \alpha\beta \tag{4.6}$$

$$\frac{\mathrm{Tr}(\mathbf{H})^2}{\mathrm{Det}(\mathbf{H})} = \frac{(\alpha+\beta)^2}{\alpha\beta} = \frac{(\gamma+1)^2}{\gamma} \tag{4.7}$$

つまり，$Tr(H)$ の二乗と $Det(H)$ の比は固有値の値に関わらずその比で決定される．コーナーは元ベクトル (x,y) において全方向で大きな変動があり，つまり固有値はどちらも大きくなることになる，よってこのような特徴を持つ候補点はあまり削除されない．削除したいエッジなどの特徴点は固有値の比率が大きくなるという特性を利用して，このような閾値処理を行い候補点の削除を行う．

3. Orientation 特徴量の記述を行うために，キーポイントの領域内の輝度勾配ヒストグラムを作成し，そのピークから各 Keypoint のオリエンテーションの算出をする．

SIFT は回転・スケール変化に対して柔軟であるという特徴を持つ．これは処理 1 でスケールに対して不変性を持たせている他に，最終的に特徴量を得る際に向きの正規化を行っているからである．各キーポイントのオリエンテーションを求めるために，まず以下の式で各画素の勾配情報を得る．

$$m(u,v) = \sqrt{f_u(u,v)^2 + f_v(u,v)^2} \tag{4.8}$$

$$\theta(u,v) = \tan^{-1}\frac{f_v(u,v)}{f_u(u,v)} \tag{4.9}$$

$$f_u(u,v) = L(u+1,v) - L(u-1,v) \tag{4.10}$$

$$f_v(u,v) = L(u,v+1) - L(u,v-1) \tag{4.11}$$

$m(u,v)$ と $\theta(u,v)$ は平滑化画像 $L(u,v)$ における勾配強度と勾配方向を示している．また，各々の局所領域における画素の座標を (x_{local}, y_{local}) とすると，その重み付き方向ヒストグラム h

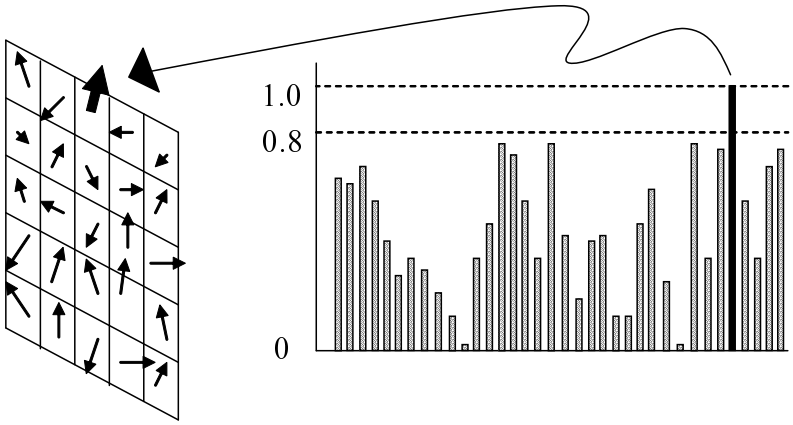


図 4.2: オリエンテーションの算出

は局所領域内の勾配情報を用いて構成される．h_θ を 36 方向に量子化したヒストグラムの各ビンとすると，その値は式 (4.12) で求まる．

$$h_{\theta'} = \sum_x \sum_y w(x_{local}, y_{local}) \cdot \delta[\theta', \theta(x_{local}, y_{local})] \tag{4.12}$$

$$w(x_{local}, y_{local}) = G(x_{local}, y_{local}, \sigma) \cdot m(x_{local}, y_{local}) \tag{4.13}$$

δ はデルタ関数であり，勾配方向が θ' に含まれている場合のみ 1 を返す．$w(x_{local}, y_{local})$ は局所領域内の各々の画素に対する重みを表しており，ガウス窓と勾配強度から得られる．つまりキーポイントにより近い特徴点がより重視されるようになる．ここで用いるガウス関数のスケール σ は，処理 1 により得られたキーポイントのスケールである．

ヒストグラムの最大値を得て，その値の 80%以上となるような勾配方向をキーポイントのオリエンテーションとして割り当て使用する (図 4.2).

4. Keypoint descriptor オリエンテーションを正規化した後，同じように輝度勾配ヒストグラムから，特徴量の記述を行う．

キーポイントとしての各局所領域を処理 3 で得られたオリエンテーション方向へと回転させ正規化を行う．最終的な特徴量の記述は周辺領域（キーポイントを中心として，保持するスケールを半径とした円内）の勾配情報から行う．領域内における 4×4 の 16 ブロックにおいて，処理 3 のようにそれぞれ 8 方向の勾配方向ヒストグラムを構成する．それら全ての情報（各ブロック 16 それぞれで 8 方向）の 128 の値を 1 つの局所特徴として記述する (図 4.3)．光源の変化によって勾配強度等が変化してしまう影響を少なくするため，局所特徴記述は各ベクトルの値の総和を用いて正規化される．

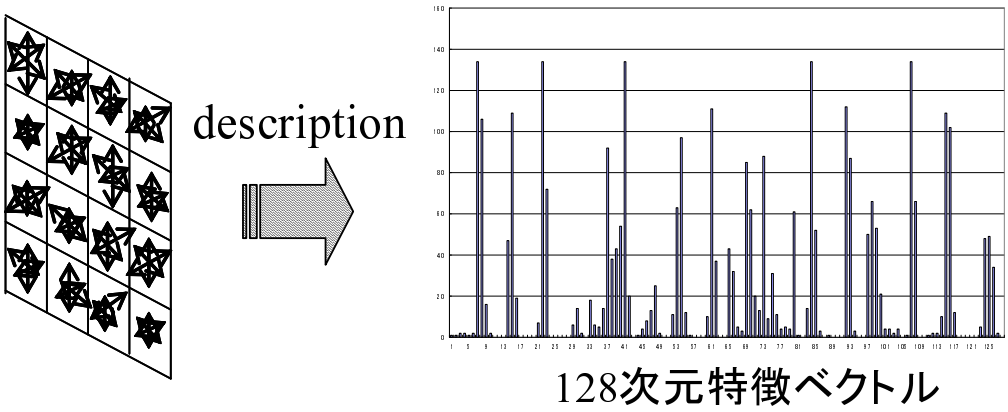


図 4.3: SIFT 記述子

更に詳細な導出方法については文献を参考して欲しい [20]．この SIFT アルゴリズム自体への研究も盛んであり，近年では SIFT の安定性を向上させるために主成分分析を適用した PCA-SIFT [31] や SIFT の欠点である色相情報が無くなるという事を考慮して改良した手法 [29] などが提案されている．

今回，SIFT で抽出された局所特徴量は 128 次元で表現されるものを用いる．SIFT は前述の通り，画像のスケール・回転変化に対してロバストに働くため，特定物体の同定には非常に有効な特徴量と言っていい．しかし，一般物体認識問題などに関するクラス分類に対して，SIFT 特徴量をそのまま利用する事は困難である．そこで Bag-of-keypoints のアプローチを用いる．

4.3 Bag-of-keypoints Approach

Bag-of-keypoints [26] は局所領域の特徴量（keypoints）のみで認識を行う手法であり，Constellation model（局所領域の相対的位置の情報も確率モデル化する手法）と同等の認識結果を出している．統計的な言語処理での Bag-of-word model [22](図 4.4) と類似しており，対象画像を，位置を無視して局所特徴の集合として考える (図 4.5)．具体的な処理としては，クラスタリング手法（今回は以下で説明する k-means を利用した）を用い，局所特徴で構成される特徴ベクトルをベクトル量子化させ，局所特徴を word として扱えるようにする．最終的な分類を行う際には画像中におけるその word の数をカウントして，ヒストグラム化する．そして各クラスの学習画像を用いて判別モデルを構成して，未知のテストデータに対して分類を行う (図 4.6)．ここでの分類には，前章で説明した多クラスの SVM を使用する．

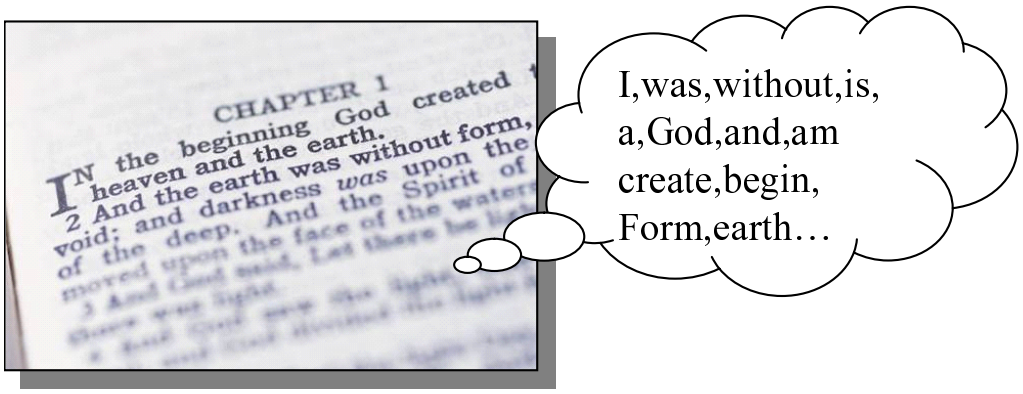


図 4.4: Bag-of-words

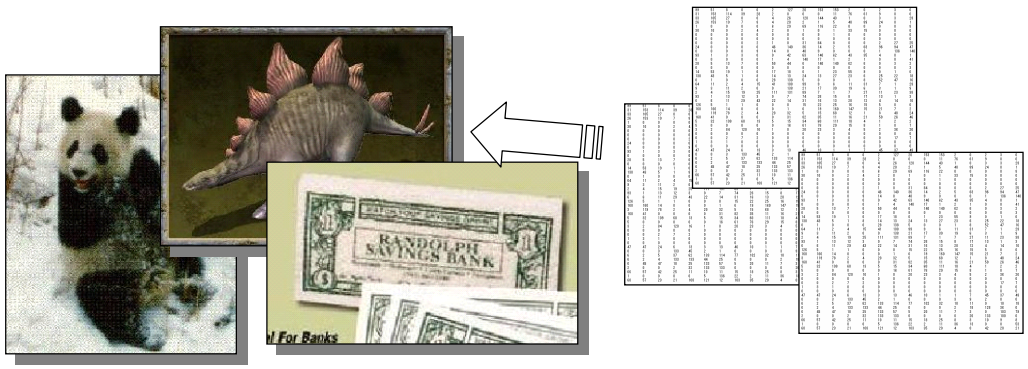


図 4.5: Bag-of-keypoints

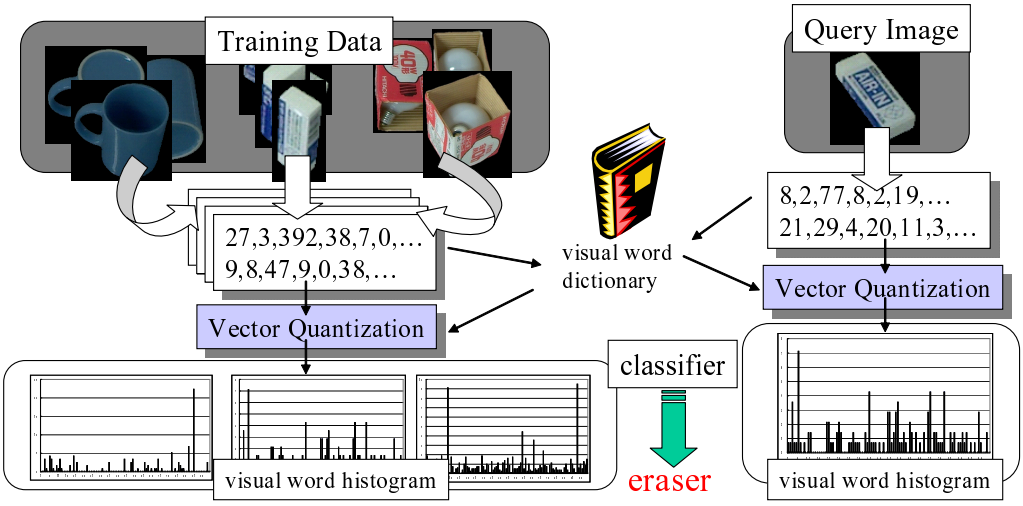


図 4.6: Bag-of-keypoints Overview

ここで各ヒストグラムの各ビンの値が最終的に使用される特徴量となるが，その値を 2 通りの形にする．一方は単純に各 word クラスタ中心に近いものをカウントして，その合計で正規化した値を利用する．もう一方は，Feature Weight という考え方を導入する．

4.3.1 Feature Weight

Feature Weight は，言語処理で導入されている特徴量であるが，各 word がどれだけ重要なものであるかを考慮し，言わば重み付けを行っている．自然言語処理における tf-idf の考え方と同じであり，背景などで頻繁に登場するような特徴クラスタの影響を少なくするような手法であり，画像においても意味が出てくると考えられる．画像の総数を N とし，特徴クラスタ x が出現する画像の個数を n_x とした時，特徴クラスタ x の逆画像頻度 idf，は式 (4.14) となる．

$$\mathrm{idf}_x = \log \frac{N}{n_x} \tag{4.14}$$

また，特徴クラスタ x が画像 d 中に出現する数を oc_{xd} とし，出現特徴の集合を W としたとき，特徴クラスタ x の画像 d 内での出現頻度 tf は式 (5.6) となる．

$$\mathrm{tf}_{xd} = \frac{\mathrm{ptf}_{xd}}{\sqrt{\sum_{i \in W} \mathrm{ptf}_{id}^2}} \tag{4.15}$$

$$\left(\mathrm{ptf}_{xd} = 0.5 + 0.5 \times \frac{oc_{xd}}{\max_{i \in W} oc_{id}} \right) \tag{4.16}$$

上の tf 値，idf 値を用いて，重要度 Feature Weight は以下のようにして求まる．

$$\text{Feature Weight}_{xd} = \mathrm{tf}_{xd} \times \mathrm{idf}_x \tag{4.17}$$

図 4.7: SIFT の抽出と，属するクラスタの Feature Weight

図 4.7 では Feature Weight 値で重要であると判断されたクラスタに量子化されるような SIFT の局所特徴を濃い円で，重要でないと判断されたクラスタに入るような SIFT を薄い円で描写している．背景などでも多く出てくるような場所では SIFT が抽出されていてもその Feature Weight は小さくなる．また，ある画像例 (cougar face・ibis) についての最終的な特徴量としての単純な正規化と Feature Weight のヒストグラムを示したものが図 4.8 である．Feature Weight は単純な正規化での特徴量に完全には依存していないことがわかる．

以上の Feature Weight の値と通常の正規化の 2 通りの値をヒストグラムの特徴量として用いる．

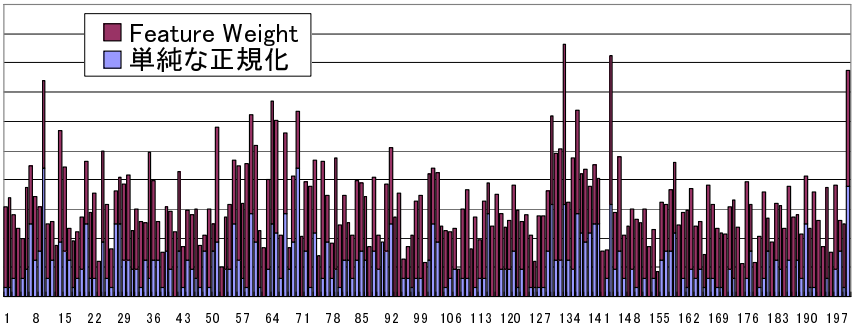
4.3.2 Clustering

クラスタリング（clustering）は，学習パターンの空間内での分布状態を見て，クラスタ（パターンが密に存在する塊状の部分）に分割をする処理のことである．ただし，高次元（三次元以上の）パターンでは人が視覚的にクラスタを発見することは難しいため，クラスタリングを自動的に行う必要がある．そのため多くの方法が工夫されている．以下に一例（k-means）を示す．

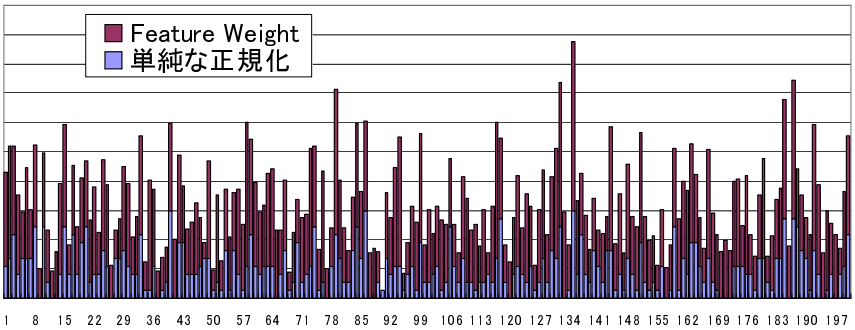
クラスタリングアルゴリズム k-means

(a) 学習パターンから任意に K 個選択し，それらを各クラスタの中心．$c_1(0), ..., c_k(0)$ とする．i$\leftarrow 0$

(b) 各学習パターンをそれと最も近いクラスタ中心 $c_k(i)1$ ($k \in 1, ...K$) に属させる．



cougar faceの1例



ibisの1例

図 4.8: 単純正規化と Feature Weight

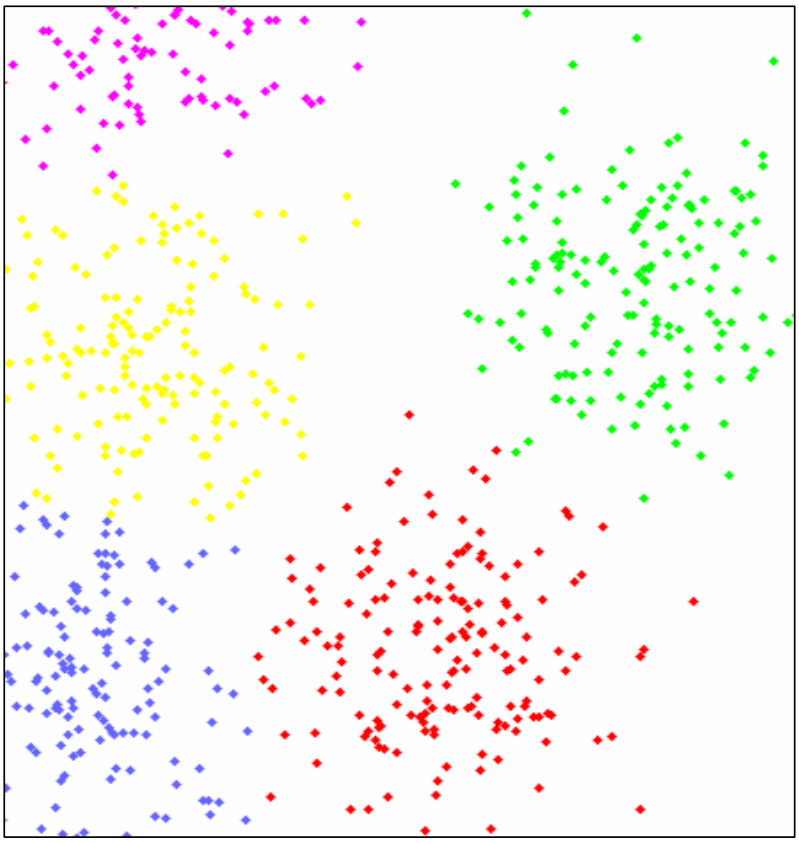


図 4.9: k-means

(c) 各クラスタごとに (b) で求めたクラスタに属するパターンの平均を求め，それを新しいクラスタ中心 $c_k(i+1)$ ($k = 1, ..., K$) とする．

(d) (c) の処理でクラスタ中心が全く変化しない場合は終了．それ以外は (b) へ．

この手法は図 4.9 のように直感的でわかりやすいが，最初に選ぶクラスタ中心によってクラスタリングの結果が変化してくること，はじめにクラスタの総数 K を与えて以後固定すること，に注意しなければならない．

4.4 SVM に用いるカーネル関数

今回サポートベクターマシーンはソフトマージンのパラメータが組み込まれているものを使用する．

また，線形学習器の識別能力は限られており，実際に実験を行っても満足のいく精度の出ない結果が出るので，前章で説明をしたカーネル表現を用いる．前章でも述べたがカーネルは線形学習器の識別能力を向上するために，データを高次元特徴空間に写像する．線形の学習器で双対表現を使用するので，写像を陰な処理として最終的な結果を得ることが出来る．つまり，線形学習器の内積を妥当なカーネル関数に置換して，カーネルを通して特徴ベクトルの内積を計算することにより，高次元特徴空間に対する非線形写像を陰なものとして，パラメータを増加することなく求められる．

今回の実験では，カーネル関数は以下に示される多項式カーネル K_{poly}，シグモイドカーネル K_{sig}，RBF(Radial Basis Function，動径基底関数) カーネル K_{RBF} を使用する．

$$K_{poly}(\mathbf{x},\mathbf{z}) = (\mathbf{x}^T\mathbf{z}+c)^d \tag{4.18}$$

$$K_{sig}(\mathbf{x},\mathbf{z}) = \tanh(\mathbf{x}^T\mathbf{z}+c) \tag{4.19}$$

$$K_{RBF}(\mathbf{x},\mathbf{z}) = \exp(-\frac{1}{\gamma}^2|\mathbf{z}-\mathbf{x}|^2) \tag{4.20}$$

4.5 各識別関数でのパラメータの調整

前述した Caltech のデータベースに対して，Bag-of-keypoints で得られたヒストグラムを特徴として，判別モデルを構成した SVM で順位出力を行った結果について考察を行う．また，学習データ及びテストデータを用いて，SVM におけるソフトマージン制御パラメータ c，カーネル及びそのハイパーパラメータを調整をして，今回用いる特徴を用いた場合にどの程度分類できるかの可能性を知る．

多クラス分類では，複数の識別平面を構成し，それを用いて最終的なクラスの決定・順位付けを行っている．ここで各分類器が同じ仮説の上に成り立っているとは限らないので，各々において最適なカーネルおよびパラメータを調整する必要がある．調整用のデータを用いてソフトマージンのパラメータ，カーネルのハイパーパラメータを調整する場合，グリッド的な探索が現在最も妥当だと言われている．つまり指数関数的にパラメータを変化させて，良い結果のパラメータをその中で採用し，徐々に最適なパラメータに近づけるようなヒューリスティックな手法である．今回は各 2 値分類器上でそのグリッド的な探索を用いて最も良い精度が出るようにパラメータの調整を行った．また，カーネルの選択もその探索に含めた．

4.6 予備実験の結果

4.6.1 予備実験の条件

以下に示す One-versus-All，One-versus-One 各々の多クラス拡張において，理想的なパラメータを用いると記している場合はテストデータをチューニングデータとしてパラメータを調整する．以下の結果は全て学習データを 30，テストデータは 50 を限度とし，１０回ランダムに各データを各クラスで収集した際の結果の平均である（各精度の標準偏差は予備実験中，提示数に因らず 1.5 を越えなかった）．使用した計算機は CPU が Xeon 3.06GHz dual・メモリは 2GB のものを一台使用した．実装したプログラムは C++言語で記述している．

学習時間の目安として当計算機を用いて，クラスタ数 200・Feature Weight で RBF カーネルを使用した場合，One-versus-One で 621 秒，One-versus-All で 246 秒であった．

4.6.2 妥当な Bag-of-keypoints のクラスタ数と最終的な特徴量の調査

一般に Caltech101 に対する Bag-of-keypoints のクラスタ数は 200〜1000 程度が望ましい結果が得られるという事が既存研究によって分かっている．今回の事前実験として簡易に利用できるデータマイニングツール Weka [34] に実装されているアルゴリズム SMO を使用して単純な正規化と Feature Weight におけるクラスタ数に従った精度傾向を調べたところ，計算量と精度のバランスは 200〜400 程度が良い（図 4.10）という結果が出た．今回は，その妥当なクラスタ数での候補提示数による傾向の違いを観察した．その結果が図 4.11 である．クラスタ数 200 での実験結果よりも 300・400 での結果の方が提示数全般に渡ってよい精度を出しているが，提示数が 30 を越えるとほぼ 100%に近くなり，あまり変わらなくなる．

特徴量間（単純な正規化と Feature Weight）の精度比較を行う．One-versus-One にて各分類器でカーネル及びパラメータを調整した状態での単純な正規化と Feature Weight の精度を比べたものが図 4.12 である（クラスタ数は 200 と 300）．単純な正規化よりも Feature Weight の方が候補提示数によらず精度が高くなっている．これはクラスタ数を増減させても同じような結果が出た（クラスタ数 100 以下では実験を行っていない）．

以上の実験結果から以降全ての実験は特徴量として FeatureWeight を，Bag-of-keypoints のクラスタ数は 200・300 を用いて行っている．以降の結果は RBF カーネルで理想的なパラメータを用いたときの精度傾向を示している．つまり該当の特徴量を用い SVM で分類を行った際での限界の値となる。

4.6.3 One-versus-One と One-versus-All での精度傾向の調査

One-versus-One は識別関数の分類結果から，One-versus-All は識別関数値から，それぞれの符号行列との前述した別種の距離を取り，精度を求める．まず線形 SVM での各々の精度傾向を求めたもの (クラスタ数 200）を図 4.13 に示す．

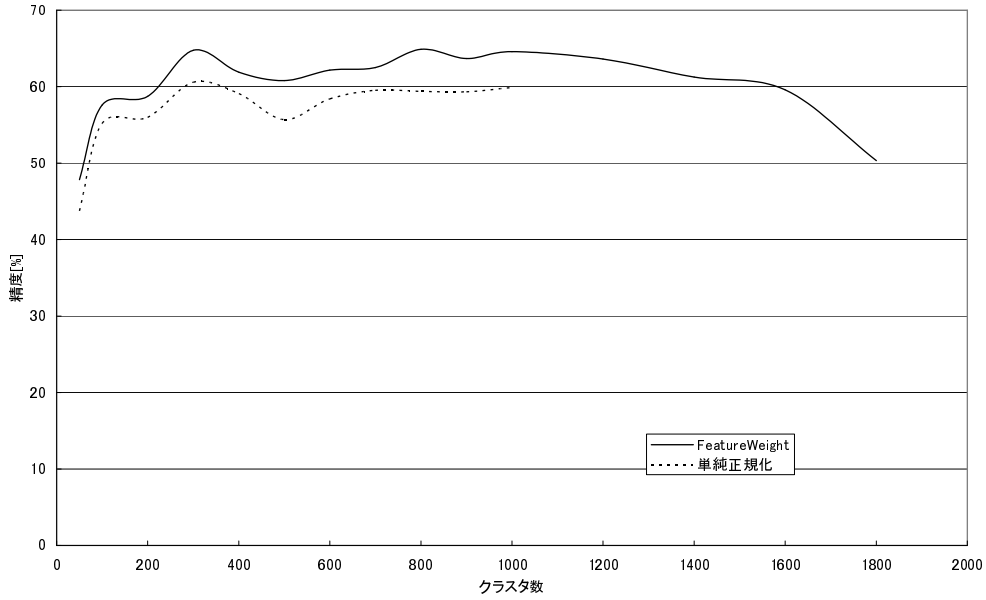


図 4.10: クラスタ数に対するクラスを一意に推定したときの精度 (weka)

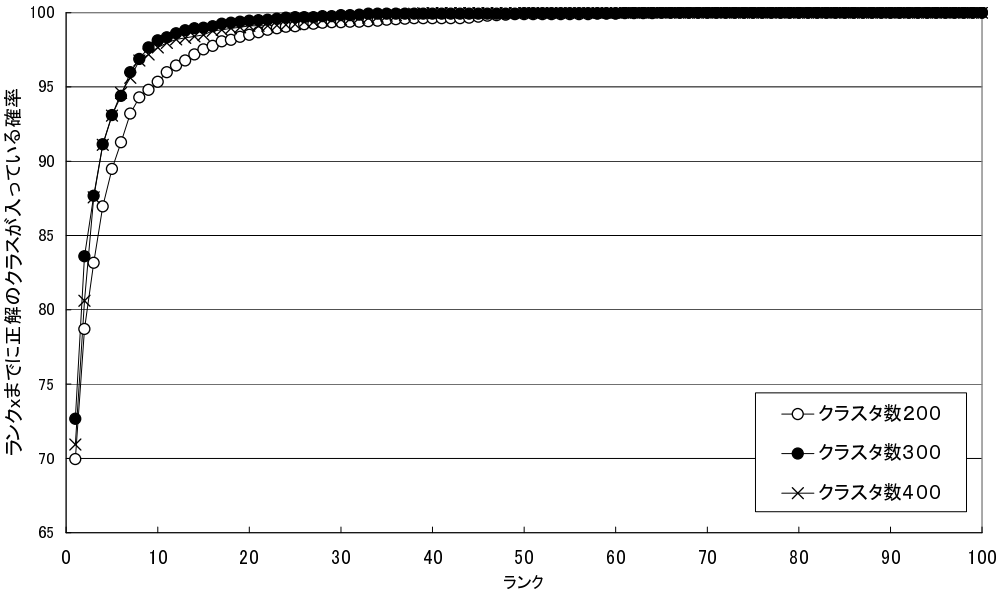


図 4.11: クラスタ数の違いによる精度傾向の比較

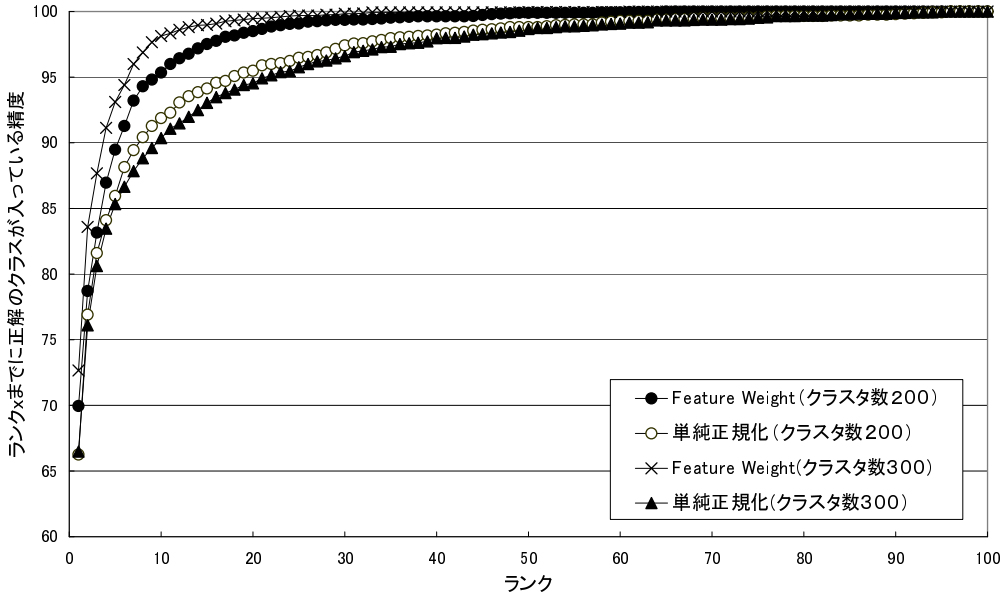


図 4.12: 特徴量での精度比較（単純正規化と Feature Weight）

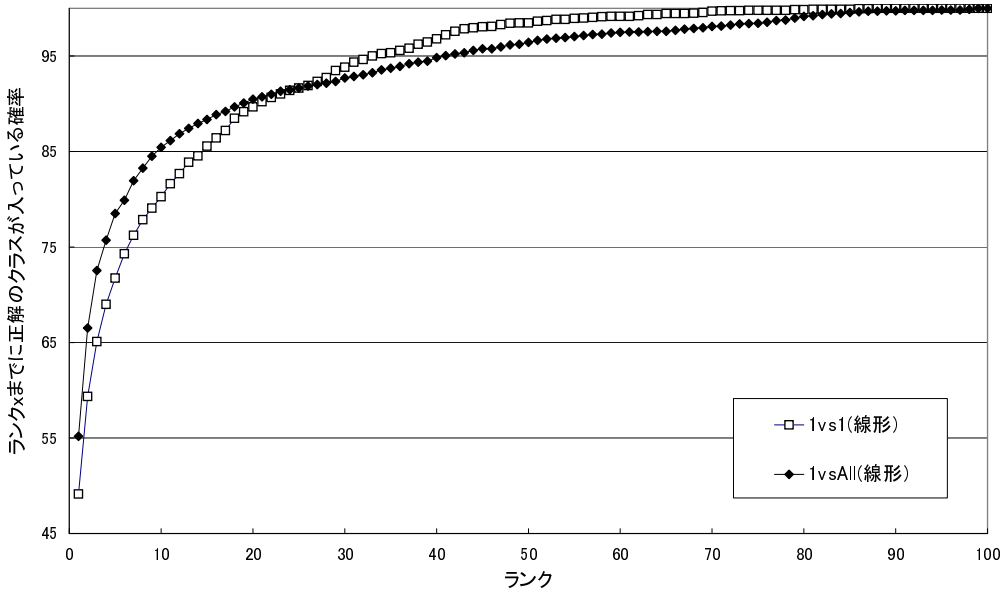


図 4.13: 線形 SVM での One-versus-All と One-versus-One の精度傾向の比較

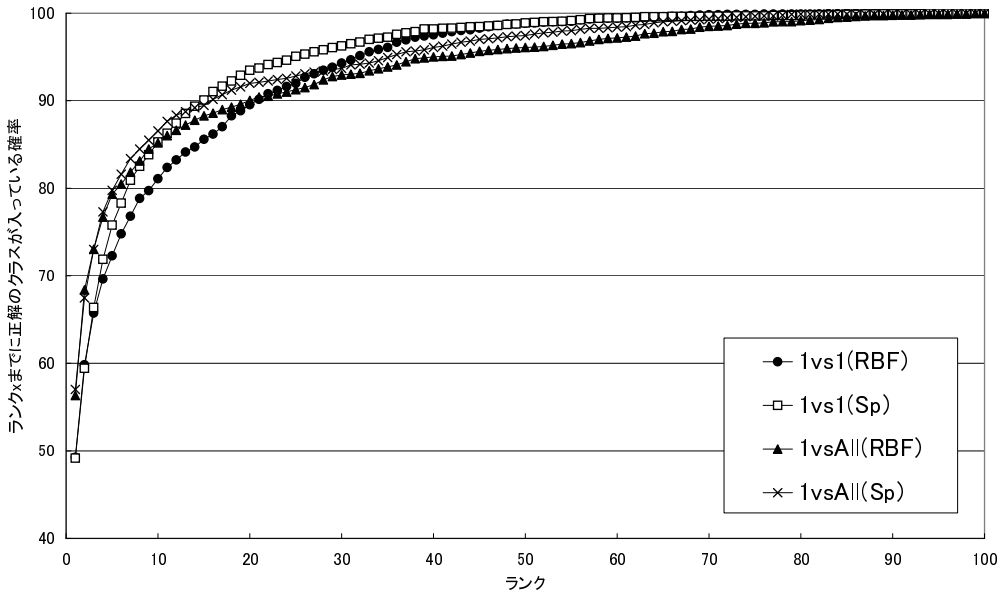


図 4.14: 各カーネルでの One-versus-All と One-versus-One の精度傾向の比較 (クラスタ数 200)

更に RBF カーネルと Spatial Pyramid カーネルの L=1 において Bag-of-keypoints のクラスタ数を 200・300 として，One-versus-One と One-versus-All で SVM を構成した場合の精度傾向を比較したものが図 4.14,4.15 である（パラメータ一定）.

パラメータ等を調整していない場合には，少数の候補提示化における精度は One-versus-All の方が One-versus-One よりもかなり高い値が出る．しかし，候補提示数を増やしていくと提示数２０辺りから One-versus-One の方が高い精度が出てくるようになる．これは少数の候補提示の場合では One-versus-One で図 4.16 のような決定不能な領域に陥り誤ったクラスに分類されてしまったデータが，候補提示数が増加したことによって正解となっていくためであると推測される．つまり候補提示数が大きくなるほど One-versus-All の識別関数値を用いて決定不能領域を強制的に作らせない優位性というものがなくなっていくことによって，逆に One-versus-One での投票的な処置によって最終的なクラスを決定する方法の妥当性が識別関数値を直接用いて順位を決める方法よりも勝るためであると考えられる.

また RBF カーネルにおいて学習データのみで各識別関数を構成して，最終的なこの特徴量でカーネル及びパラメータを調整して分類できる限界についての複数候補提示化における結果を，元のチューニングしていないもの（正確には全体で最適な結果が出る一意のパラメータに設定しているもの）と比較したものが図 4.17 である（クラスタ数は 200 である）.

RBF パラメータを調整した場合では，One-versus-All は殆ど変化が見られないのに対して，One-versus-One では，著しく精度が上昇している．これは RBF カーネルで各分類器でパラメータ類を調

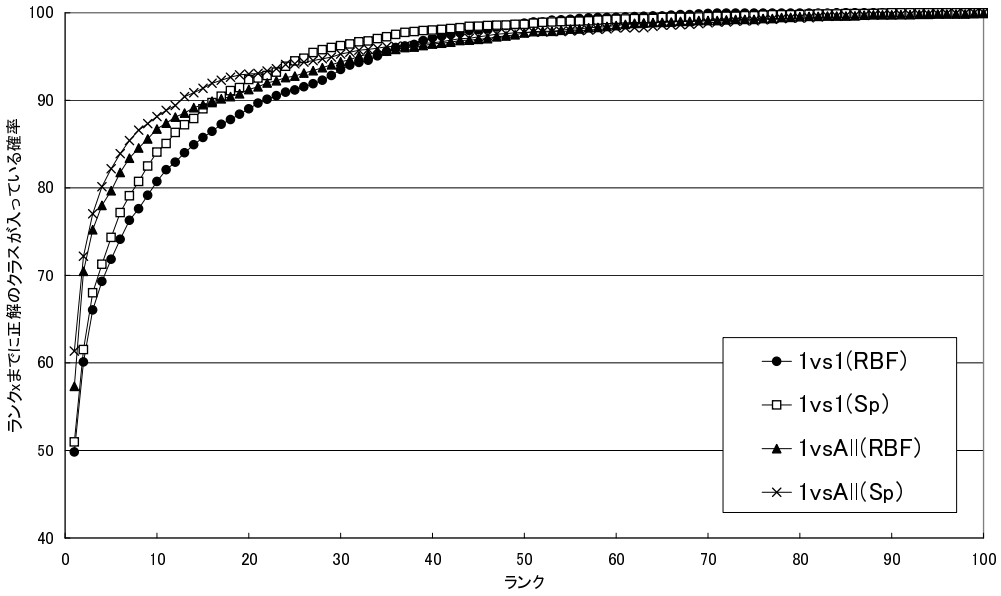


図 4.15: 各カーネルでの One-versus-All と One-versus-One の精度傾向の比較 (クラスタ数 300)

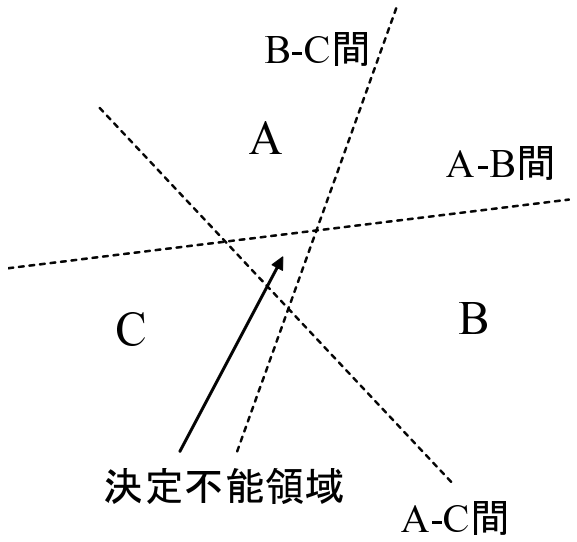


図 4.16: One-versus-One で発生する決定不能領域（3 クラス分類での例）

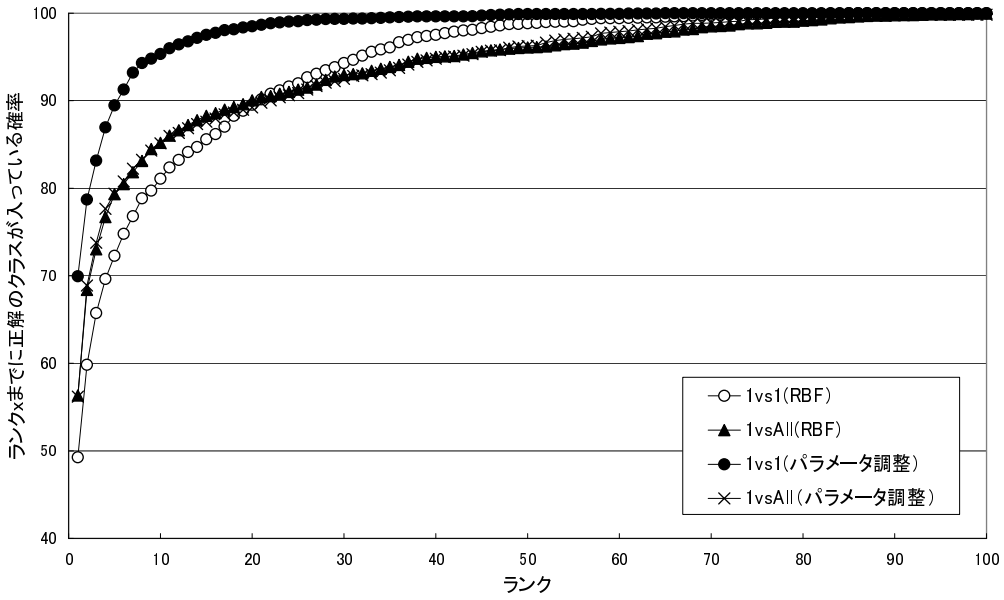


図 4.17: 各分類器でパラメータを調整した場合との比較（One-versus-All と One-versus-One）

整する事が，One-versus-All の識別関数値を順位推定に利用する事の正当性を失わせているのに対して，One-versus-One では出力は分類結果となり，その Humming 距離で候補数に従った精度を求めているため，各 2 値分類器の分類精度の上昇が直接結果に影響するためであると考えられる．

図 4.18 は，その One-versus-One に関して，前述のアプローチ（識別結果 (±1) をそのまま用いる）と識別関数値と loss decoding を用いた手法での比較を，パラメータ調整有り・無しのそれぞれについて行った結果である．識別関数値を用いると，今回の損失関数では精度が下がってしまうが，やはりパラメータ調整を行うとその評価の正当性が無くなってしまうため，精度が下がることがわかる．

4.6.4 カーネル選択をすることによる分離可能性の上昇

One-versus-One において RBF カーネルのみを用いてパラメータのみを調整したものと，カーネル関数も選択して調整したもの，および線形 SVM を用いたものでの精度の違いを示したものが図 4.19 である（Feature Weight，クラスタ数 200）．カーネルを選択していった方が多少精度の高い結果になっていることがわかる．仮説に適した分離平面を構成したのか，偶然に分離できたのかは不明であるが，3 つのカーネルのうち最も精度の高い結果の出た RBF カーネル単体を用いるよりも可能性は上がることがわかる．精度が同程度の時は RBF、シグモイド、多項式の順でカーネルを選択している．

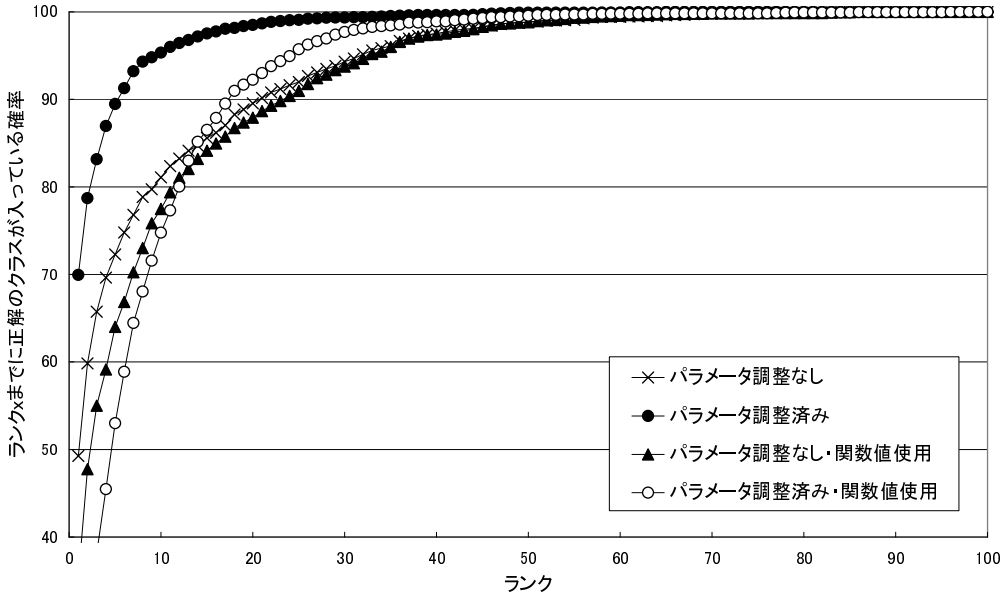


図 4.18: 識別関数値を用いた場合と識別結果を用いた場合の比較

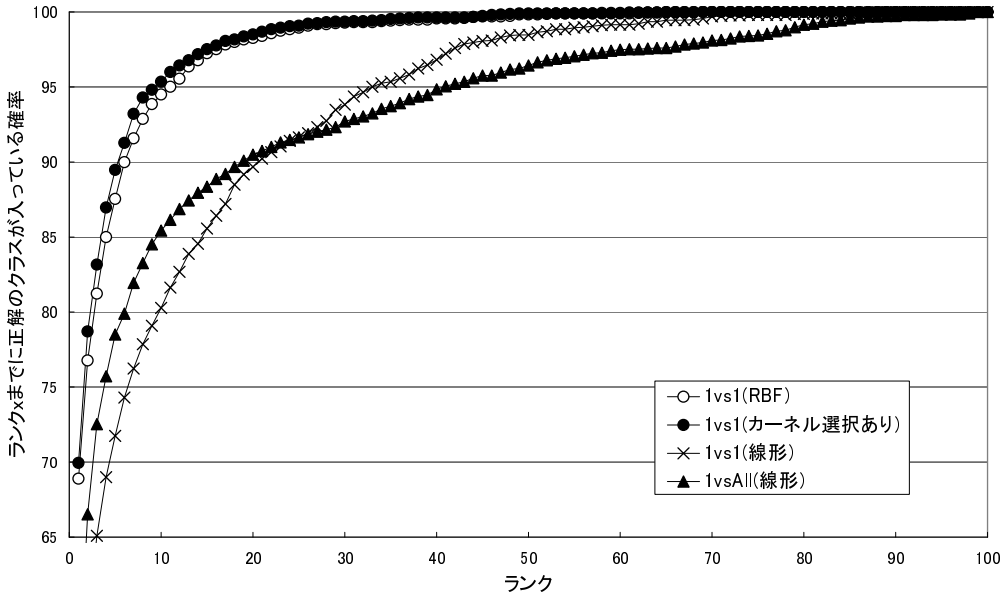


図 4.19: カーネル選択の有無による精度の違い

表 4.1: カーネルの選択割合 [%]

数	Poly	Sig	RBF	PorS	SorR	RorP	どれでも
200	2.65	4.16	17.11	0.98	4.16	2.02	68.93
300	2.35	5.94	17.56	1.21	4.46	2.39	66.08

表 4.2: Spatial Pyramid カーネルを含めたカーネルの選択割合 [%]

	Poly	Sig	RBF	Sp
唯一の最高精度を出した割合	1.57	4.66	12.75	5.73
最高精度の割合	69.66	75.14	87.41	63.36

4.6.5 カーネルの適合度の比較

識別平面全体の内のカーネル選択の割合（200・300 のクラスタ数における、多項式 (Poly)，シグモイド (Sig)，RBF，多項式かシグモイド (PorS)，シグモイドか RBF(SorR)，RBF か多項式 (RorP)，どのカーネルを選択しても最高精度が同じ (どれでも)，の選択割合）を示したものが表 4.1 である．この 3 つのカーネル中だと RBF カーネルが最も 2 値問題の仮説にフィットしていると言える．

さらに，Spatial Pyramid カーネルのレベルが 1 の場合の（つまり単純に 2 クラスの要素間の最小値をカーネル値とするような）カーネルを適用した例との比較も行った．その結果が表 4.2 である．最高精度を出す割合は RBF カーネルがやはり最も高く出て，Spatial Pyramid カーネルは最も低く出る．しかし，唯一の最高精度を出した割合を見ると Spatial Pyramid カーネルは RBF カーネルに続いて 2 番目に高くなっている．

また，このカーネルを使用した最終的な精度傾向を示したものが図 4.20 である．この結果としての精度傾向を見るとパラメータが一定のものを用いた場合には Spatial Pyramid カーネルは RBF カーネルよりも有効に働いているといってよい．各識別関数でパラメータの調整を行った場合には RBF カーネルの方が分離の可能性は高まることがわかるが，あくまでも理想的に調整が出来た場合である事に注意して欲しい．

4.6.6 各 2 値問題での分離精度

理想的にパラメータやカーネル関数の調整を行った各識別平面での，分離精度の分布を示したものが図 4.21 である．縦軸が 2 値問題中の全体精度 [%]，横軸が各 2 値問題例である．実験条件は Bag-of-keypoints のクラスタ数を 200，特徴量として Feature Weight，SVM の構成は One-versus-One となっている．

ここで，完全な (100%) の分離が出来ない 2 値問題は全体の 40%であった．また，90%の分離が出来ない 2 値問題は全体の 8%，80%では全体の 4%，70%で 2%となった．

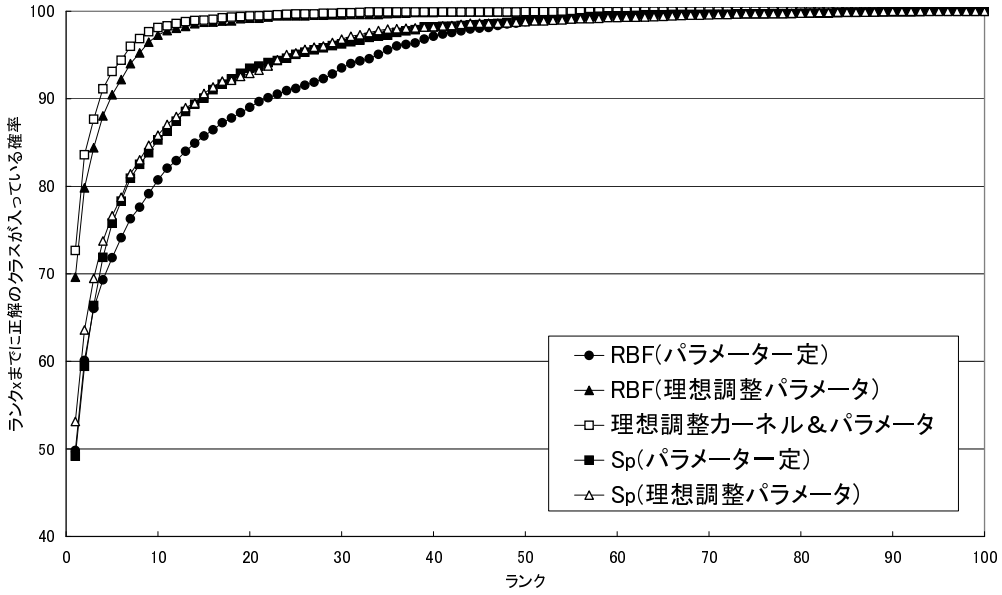


図 4.20: RBF Kernel と Spatial Pyramid Kernel 間での（理想調整パラメータ有無での）精度比較

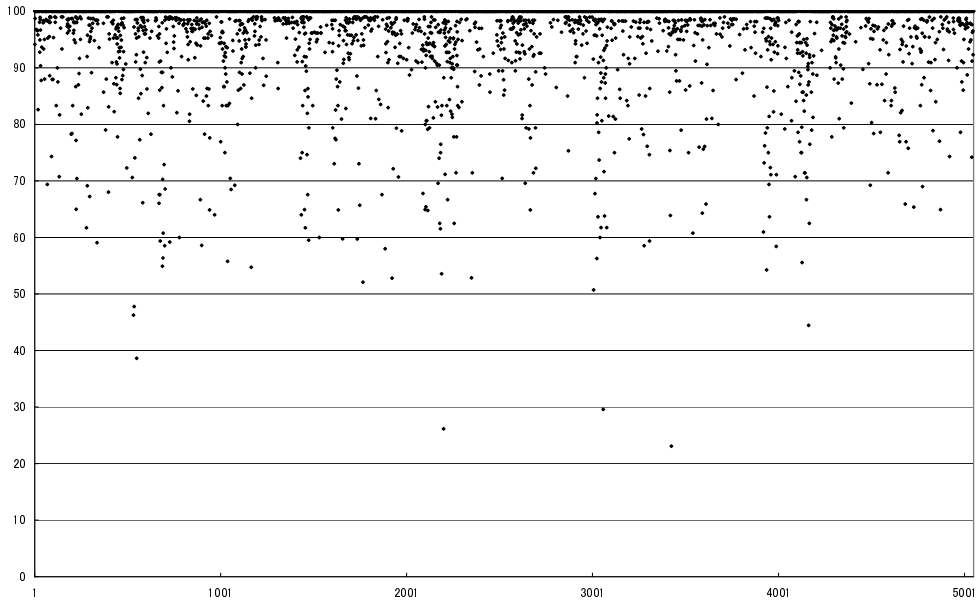


図 4.21: 各識別平面での分類精度の分布

表 4.3: 分類で多くの誤りを出した 2 値問題例

2 値問題	正例正解	正例不正解	負例正解	負例不正解
scorpion-pigeon	0	50	15	0
dolphin-flamingo	11	24	27	10
octopus-platypus	0	5	4	0
ewer-flamingohead	2	48	15	0
ewer-minaret	50	0	0	46
Faces-camera	18	32	19	1

4.6.7 分離が困難だった 2 値問題

識別においてどのような 2 値分類問題が難しく，上手く分離できなかったのかについて調査を行った．理想的にパラメータやカーネル関数の調整を行った 2 値 SVM で分離が出来ていない識別平面についてその一例を示す．テストデータに対する分類において多くの誤りを出している 2 値問題例を示したものが表 4.3・図 4.22 である．これはどれだけ識別関数のカーネルやパラメータの値を調整しても上手く分類が出来なかった問題例である．多くが片方のクラスに因ってしまい，分離が出来なくなっているものが多い．これは今回用いた特徴量では根本的に判別が無理であるという事を示している．

4.6.8 パラメータの決定方法による精度比較

一般的に未知のデータに対してパラメータ調整を行うには，学習データ・テストデータの他にチューニング用のデータがあるのが望ましい．しかし，この一般物体認識の性能の標準的な測定では学習データを 30 程度用いて残りをテストデータとしているので，合計データ数が少ないクラスもある事を考慮すると，学習データのみで行う場合は cross-validation を用いるのが妥当だと言える．10 分割交差で学習データのみでパラメータを調整した場合と，あらかじめ実験したデータのパラメータ（データをランダムにシャッフルを行い 5 通り）を知識として得ている場合での精度の平均との比較をした図が 4.23 である．各分類器のパラメータの調整は，異なった学習データおよびテストデータから予め得られている知識を用いて行ってもあまり意味を成さない事がわかる．学習データのみで交差確認で調整を行っても，提示数 50 まで見るとほぼ精度 100%となり，フィルタなどの活用においては十分実用的な精度であると言える．

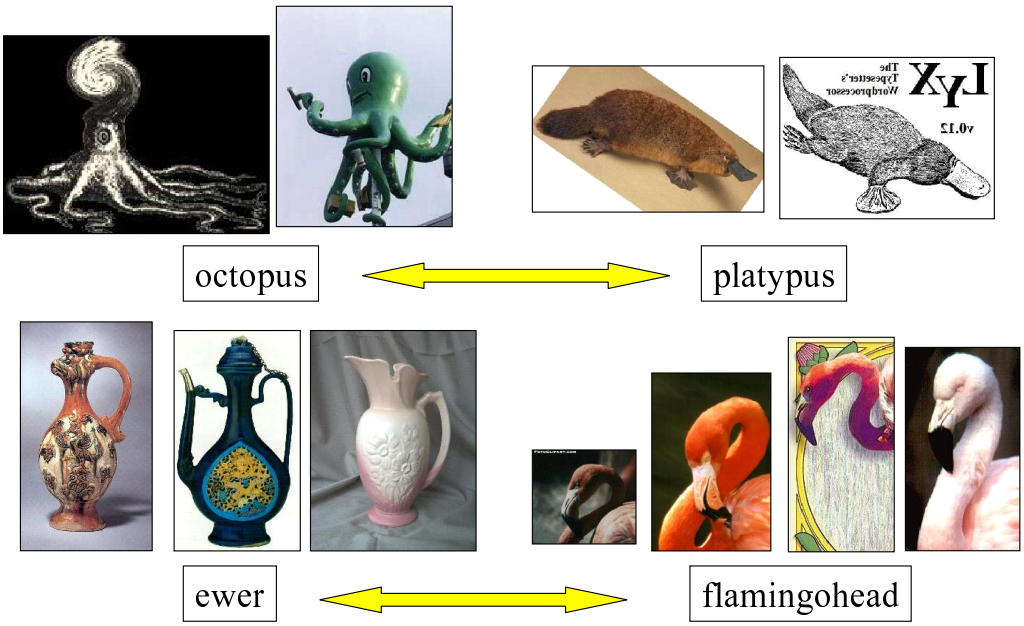


図 4.22: 分類が困難であった 2 値問題

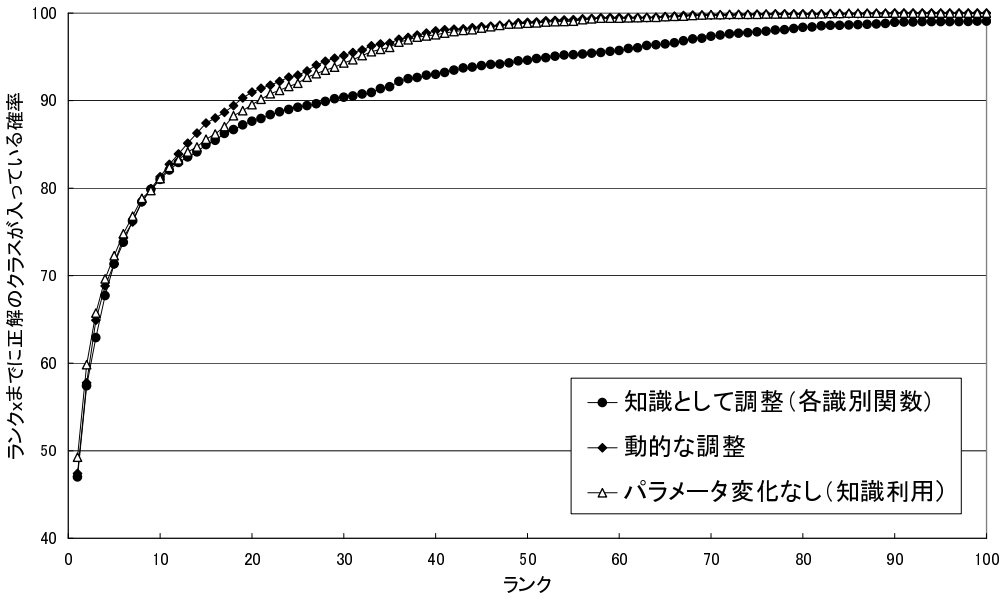


図 4.23: パラメータの決定方法での精度傾向の比較

4.7 予備実験の考察

実験結果を見ると，単純な特徴量だけを用いて Bag-of-keypoints のアプローチを取っても，パラメータが理想的に調整できれば単純な特徴量だけを用いても十分実用的な精度が出る（候補数２，３程度で既存研究の最高精度を上回る）．特に３０程度の複数候補提示した場合はその条件下でのパラメータの調整知識があればほぼ 100%の精度を出す事が出来る．

しかし，基本的に実験で得られた各分類器のパラメータを，完全に条件が同じである場合以外に適用して精度を出す事は難しい．学習データのみでも Cross Validation を用い各識別関数においてパラメータの調整を行えば，One-versus-One では候補数 50 で精度はほぼ 100%（具体低には 99%）となる（90%は 20 程度で超える）．つまりほぼ信頼できる精度で 1/2 程度にデータを削減することは可能であるが，やはり精度には不満が残る．One-versus-All でのパラメータ調整は難しく，調整無しで適用すると少ない提示数においては One-versus-One の精度を上回るが，精度が 100%近くに収束するまでの候補提示数が多い．パラメータの決定を動的に行った場合と知識として用いた場合での精度傾向の違いは，大きくは出ないので，多くの計算時間を消費して毎回動的にパラメータを決定する利点はあまり無いと考えられる．

また，根本的に分類が不可能な 2 値問題も存在し，それに対処するには適した特徴量の考案を行ったり，学習領域を上手く認識するような工夫が必要となってくる．

4.8 本章のまとめ

本章では Caltech101 のデータセットを対象として，SVM の多クラス分類への構成手法を用いて一般物体認識における複数候補提示下の分類性能傾向について様々な条件下で調査を行った．

次章では今回の実験で判明した，分離が困難な 2 値問題についての分類について議論し，複数候補提示下での精度を高めるための手法について提案する．

第5章　分類困難な２値問題に対する学習領域の最適化

本章では複数候補提示下の分類性能を高めるための手法を提案する．

予備実験により 100%近い (95%以上の) 精度に収束する提示数が少ない SVM の構成方法は，One-versus-One であるということがわかった.（ただし 90%程度を信頼できる精度とするなら One-versus-All の方が良い）また，仮に理想的にソフトマージンやカーネルのハイパーパラメータを設定したとしても，現在用いている特徴量およびその Bag-of-keypoints での表現では２値分類としてみても根本的に分離が困難な問題が存在している事がわかった．

このようなそもそも単純な特徴量を用いても分離が困難な問題に対しては，異なった性質の特徴量，もしくは大量の特徴量を学習器に投入したり，学習する領域の絞込みをおこなってノイズを少なくするなどの工夫を取らなければどうしようもない．実際近年一般物体認識に対して成果を出している研究はそのような手法を用いている．

ただし，部分的に異なった特徴量や大量の特徴量を用いるためには，学習データはその対象を絞れるが，同じ特徴次元で分類を行うためにテストデータもそれに応じた特徴量を算出しなければならなくなり，結局全テストデータについて全特徴量を検出しなければならなくなるので，非常に計算コストが高くなってしまう．

一方，学習データからの Region of Interest の学習は，学習データとして信頼できるようなモデルを構成したいという事をそもそもの意図として持っている．すなわち，テストデータも同じような処理を行わなくて良いので，部分的な学習器の構築に向いているといえる．

以上のような背景から，今回の提案手法では Bosch の訓練データでの Region of Interest の学習を大きな枠組みとして，その問題について絞る事を主眼に置いた．

5.1　訓練データからの ROI の学習

Region of Interest の学習手法の基本的な考え方については２章で述べている．

図 5.1 は，特徴量を Feature Weight，Bag-of-keypoints のクラスタ数を 200 とし，RBF カーネル ($c = 20, \gamma = 0.04$)，Spatial Pyramid カーネル ($c = 20, L = 1$) で Region of Interest を学習させたものとそのままオリジナルの画像で学習させたものでの精度の比較である．図を見るとわかるようにかならずしも精度傾向が高くなってはいなく，むしろこの条件下だと低くなっていることが多い．

図 5.2 は今回学習データ中で探索を行った得られた Region of Interest の例である．左例のように上手く抽出出来ているものもあれば右例のように，邪魔なオブジェクトが入ってしまっている場合も

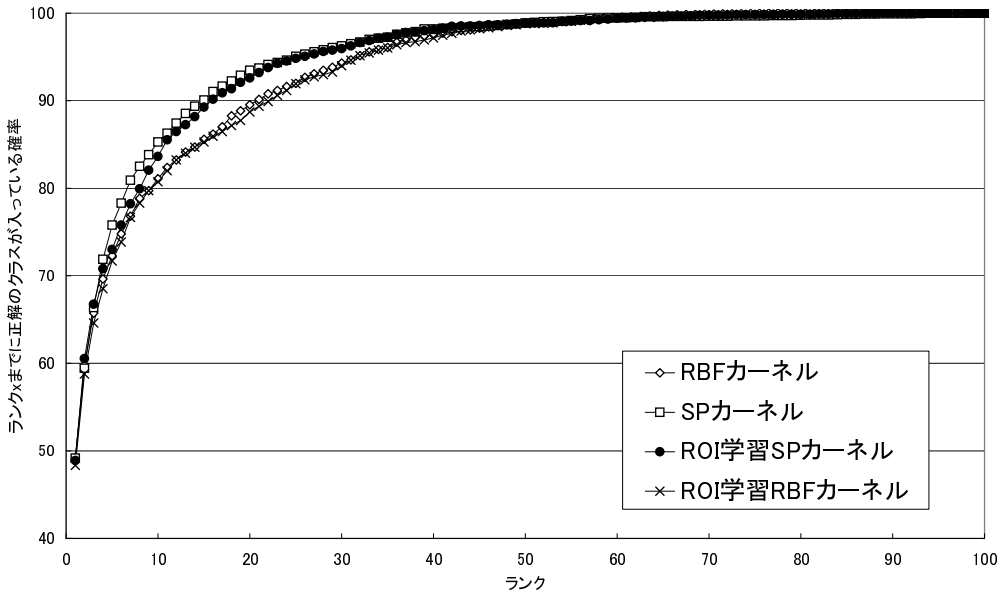


図 5.1: Region of Interest の学習の有無による精度傾向の比較

ある．

学習データ群からの Region of Interest の学習は，その物体の領域の局所特徴量だけをなるべく得ようとするために，背景に出現する局所特徴量は除去される可能性が高い，つまり結果としてノイズを除去するような形となる．

だがここで One-versus-One として分類を見たときには各問題は 2 値分類問題となるが，その時背景に出てくる局所特徴量は分類に邪魔になる場合だけではないことがわかる．もちろん物体と関係の無い背景の局所特徴量が邪魔となり，Bag-of-keypoints のヒストグラムが 2 値問題として分離しづらくなる状況も多い．しかし，(特に動物などの例においては）時に背景がいわばコンテキストとして，むしろその物体を分離しやすくするための特徴量となっている場合も出てくるのではないかと考えられる．そのような場合に Region of Interest の学習を行うような操作をすると，逆に情報が少なくなることによって分離が今まで可能だった 2 値問題が途端に困難になってしまうことになる．また，全ての学習データ群からの Region of Interest の学習を行うのは非常に計算量的なコストがかかるという問題もある．

今回我々は，Support Vector Machine の One-versus-One の構成をする中で，そのままの特徴抽出では 2 値分類が困難であるようなクラス間の問題でのみ Region of Interest の学習，つまり学習領域の絞込みを行うことを提案する．

図 5.3 のように，学習セットの中で分類が困難であるような 2 値問題をあらかじめ得て，そのような問題にだけ学習領域の絞込みを行うような操作を行う．

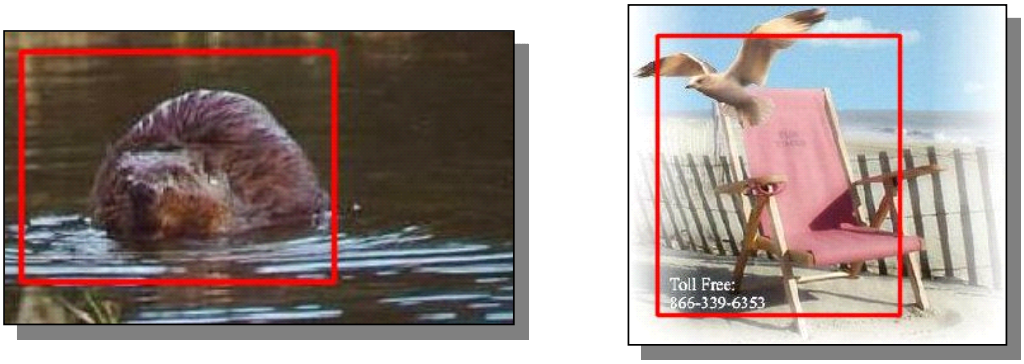


図 5.2: Region of Interest 検出結果の例

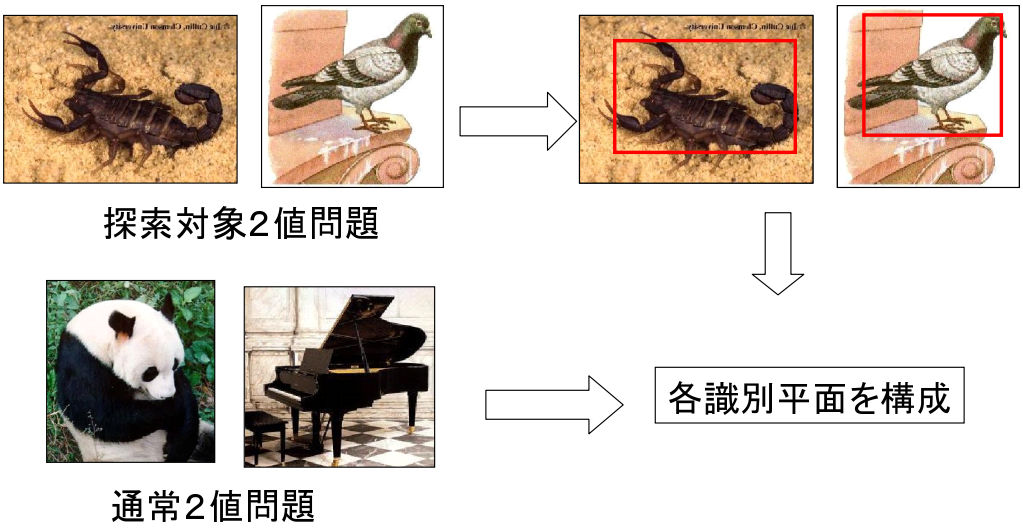


図 5.3: 提案手法

5.2 提案手法の流れ

まず領域探索対象の 2 値問題を発見しなければならないので，事前に学習データの中でさらに準学習データ，およびテストデータに分離して,（正確にはテストデータの集合を得て）その中での精度傾向を調べる操作，つまり Cross Validation（交差検定）を行う．予備実験でも同操作を行ったが，k-Cross Validation では，標本群を k 個の部分サブセットに分割し，そのうちの 1 集合を検査例，残りの $k-1$ 集合を学習例とし，全組み合わせ k 回の検定を行う．その操作の中で精度が低く出るような 2 値問題を探索対象の候補とする．

Region of Interest には Pyramid Match Kernel に基づいて類似度を計算し，その値が最適値になるものを最適化する．Pyramid Match Kernel は 2 章にも述べたように，2 つの bag 同士の部分マッチングに基づいて類似度を計算するカーネル関数である．領域探索が必要と判断された 2 値問題に属するクラスについて探索を行うのであるが，以下のコスト関数 L_i を最適化する ROIr_i を求めるのは非常にコストがかかる．

$$L_i = \max_{r_j} \sum_{j=1}^{s} K(D(r_i), D(r_j)) \tag{5.1}$$

$$K^L(X,Y) = \sum_{m=1}^{M} \kappa^L(X_m, Y_m) \tag{5.2}$$

$$\kappa^L(X,Y) = I^L + \sum_{l=0}^{L-1} \frac{1}{2^{L-l}}(I^l - I^{l+1}) \tag{5.3}$$

$$= \frac{1}{2^L} I^0 + \sum_{l=1}^{L} \frac{1}{2^{L-1+1}} I^l \tag{5.4}$$

$$I(H_X^l, H_Y^l) = \sum_{i=1}^{D} \min(H_X^l(i), H_Y^l(i)) = I^l \tag{5.5}$$

今回の実験では，近似最適としてサブセット s と比較する画像の ROIr_j の固定を行い，r_i を探索する．

更により重要なクラスタにマッチングをするように類似度計算の際に必要となる Spatial Pyramid

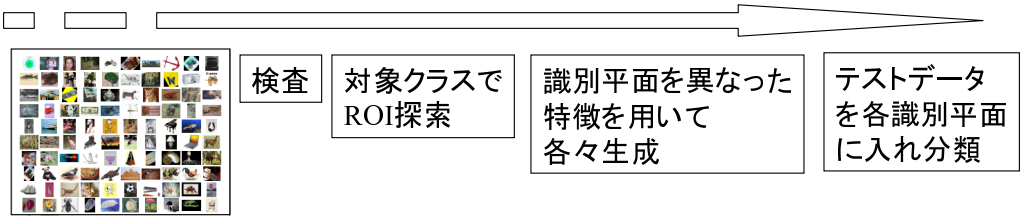


図 5.4: 簡単な流れ

表現に Feature Weight を利用する．この時，特徴クラスタ x の画像 d 内の出現頻度 tf は下式となる事は前章で述べた．

$$\mathrm{tf}_{xd} = \frac{\mathrm{ptf}_{xd}}{\sqrt{\sum_{i \in W} \mathrm{ptf}_{id}^2}} \tag{5.6}$$

$$\left(\mathrm{ptf}_{xd} = 0.5 + 0.5 \times \frac{oc_{xd}}{\max\limits_{i \in W} oc_{id}} \right) \tag{5.7}$$

特徴クラスタ x が画像 d 中に出現する数を oc_{xd} とし，出現特徴の集合を W としている．しかし Spatial Pyramid 表現の場合，各レベルで局所特徴が位置情報に従って 4 分割され，その各関心領域中でヒストグラムを作成する事になる．つまり特徴クラスタ x がレベル l での画像 d^l 内での m 番目の ROI$d^l r_m$ に出現する数を $oc_{xd^l r_m}$ とすると，レベル l の画像中の範囲 r_m における出現頻度 $\mathrm{tf}_{xd^l r_m}$ は以下のようになる．

$$\mathrm{tf}_{xd^l r_m} = \frac{\mathrm{ptf}_{xd^l r_m}}{\sqrt{\sum_{i \in W} \mathrm{ptf}_{id^l r_m}^2}} \tag{5.8}$$

$$\left(\mathrm{ptf}_{xd^l r_m} = 0.5 + 0.5 \times \frac{oc_{xd^l r_m}}{\max\limits_{i \in W} oc_{id^l r_m}} \right) \tag{5.9}$$

前述したように特徴クラスタ x の逆画像頻度 idf，は下式で示される．

$$\mathrm{idf}_x = \log \frac{N}{n_x} \tag{5.10}$$

以上で求めた出現頻度 tf と逆画像頻度 idf を用いて画像 d のレベル l における m 番目の関心領域 r_m の最終的な特徴量 $\text{Feature Weight}_{xd^l r_m}$ は式 5.11 から得られる．

$$\text{Feature Weight}_{xd^l r_m} = \mathrm{tf}_{xd^l r_m} \times \mathrm{idf}_x \tag{5.11}$$

5.3 Feature Weight の更新

Feature Weight を利用する際に，ROI として新たに学習データ群が出来上がるので，その特徴量のデータ変化に従って idf 値の変化をさせる処理を行うことも出来る．つまり重要度の更新を考慮に入れるということである．この場合は，各クラスに応じた特徴クラスタの出現数を領域探索対象に含まれる 2 値問題と含まれない問題を考慮し正規化することになる．

全クラス数が K の学習用データのあるクラス k が生成する 2 値分類識別平面のうち，領域探索対象に入っている数を e_k とする．あるクラス k においてオリジナルの学習画像の中に存在する特徴クラスタ x が出現する個数を $n^o_{k,x}$ とし，Region of Interest 内に特徴クラスタ x が出現する個数を $n^r_{k,x}$ とする．その時，更新後のクラスタ x の出現数 n_x は式 5.12 で表される．

$$n_x = \frac{1}{K}\sum_{i=1}^{K}(K - e_i - 1)n^o_{i,x} + e_i n^r_{i,x} \tag{5.12}$$

よって最終的な特徴クラスタ x の idf は以下のようにして求まる．

$$\mathrm{idf}_x = \log\frac{N}{n_x} = \log\frac{NK}{\sum_{i=1}^{K}(K - e_i - 1)n^o_{i,x} + e_i n^r_{i,x}} \tag{5.13}$$

以上のような更新を行うには各特徴クラスタの idf 値を再度求めなければならないので，もし更新を行うことの効果が少ないなら，計算コスト的にもしないほうが良い．

ここで，元の領域を絞らずにそのまま学習させた場合の各識別平面による識別率に対する，Region of Interest を学習させた場合での Feature Weight の更新を行ったものと行わなかったものでの識別率の比較を行った結果を 5.1 に示す．表中の数値は，もとの認識率よりも上回った（同値のものは除外する）識別平面数の平均 (標準偏差) を示す，Caltech101 を対象としているので全識別平面は 5050 である．

表の結果からわかるように，Feature Weight の更新をすることの有無での性能の違いはあまり出ていない．むしろ更新を行うことで精度が非常に落ちる可能性も出てくる．よって今回の一連の実験において，ROI 検出後の Feature Weight の更新は行わないものとする．

表 5.1: Feature Weight の更新の有無による違い

クラスタ数	更新なし (SP)	更新なし (RBF)	更新あり (SP)	更新あり (SP)
200	763(± 31)	748(± 27)	621(± 23)	795(± 35)
300	692(± 22)	753(± 35)	562(± 31)	178(± 94)

5.4 本章のまとめ

本章では，複数候補提示下での分類精度の上昇を図るために，部分的に Region of Interest の探索を行い，その中の局所特徴量での Bag-of-keypoints 表現を用いた各 2 値問題に対して識別関数を独立に生成する手法を提案した．

次章では本提案手法を適用した結果について，元の領域探索を行わない場合の精度と全データセットに対して Region of Interest を探索した場合の精度と比較し，その結果について示す．また，探索対象問題を決定する閾値についての検証も同時に行う．

第6章　評価

前章で提案した手法を予備実験と同じように実際にデータセット Caltech101 を対象として適用した結果を示す.

6.1　実験条件

使用計算機は予備実験と同じく CPU が Xeon 3.06GHz dual，メモリ 2GB，プログラムの記述言語は C++である．実験結果導出の手順も予備実験と同じように全ての精度はクラス毎のデータをシャッフルして 10 回実験を行った平均とし（標準偏差は精度を検査しているものについては全実験で提示数に因らず 1.2 未満であったのでグラフの優劣に影響は出ない），学習データ数は 30，テストデータ数は最大 50 としている．

領域探索のためのパラメータとして探索のサブセット s は 3，r_j は画像のオリジナルサイズ，カーネルは Spatial Pyramid Kernel（$L = 3$）を用いた．

多クラス分類のための SVM は予備実験での複数候補提示下での 100%近くへの収束時間の速さから，また手法がそもそも One-versus-One を前提として構成されているので，One-versus-One を用いている．SVM の識別平面を構成する際のカーネル関数は，予備実験から RBF カーネルと Spatial Pyramid Kernel の $L = 1$（つまり位置情報や Bag of Keypoints のレベルを考慮しない）を用いた．

予備実験の結果から，Bag-of-keypoints のクラスタ数は 200，最終的に用いる特徴量として Feature Weight を用いている．

パラメータの設定方法であるが，予備実験から動的に生成を行うのと，定置として知識として与えたもので極端な優劣は示されなかった．計算コストの事も考慮し，今回の実験ではパラメータを一意に，過去の実験から良い傾向の出ているパラメータを与えた．ここでは双方ともソフトマージンの制御パラメータ c は 20 とし，RBF カーネルのハイパーパラメータは $\gamma = 0.04$ とした．

6.2 実験結果

6.2.1 妥当な探索２値問題決定閾値

探索対象クラスを絞るために，Cross Validation を利用する．３分割交差検定において２値分類において精度が高くでない問題を探索の対象とし，その２値問題に含まれるクラスのみに Region of Interest の探索を行う．その閾値であるが，事前実験で様々な値に変化させて本手法を適用してみたところ，大体 100〜200 程度が良いという結果が出た．図 6.1 は訓練データについて Cross Validation をした結果，精度が下位の２値問題からソートしたものと，それに対応する２値問題について Region of Interest を得てから学習を行ったものの精度傾向を比較したグラフである．正の値であるほど，Region of Interest での特徴量で学習する方が分類精度の良くなる例であり，負の値であるほど何も適用しなかった方が分類精度の良くなる例である．図を見ると，正の値が出ている例が，元々の問題例の精度が高くなるにつれて少なくなっている事がわかる．図 6.1 だけでは一見その傾向が分かりづらいので，その問題例に対する精度比較数値を下位から加えていったものを示したものが図 6.2 である．100〜250 あたりの値では大きく正の値を保っているが，270 を越えたあたりから急落し，負の値になっている，これはこれ以上の問題について領域探索をしても誤分類をしてしまう可能性が高まることを示している．事前実験の結果でよい結果がでた閾値は，図 6.2 で正の値が高く出ていることがわかる．これは 100〜250 程度の候補に決定することの妥当性を示している．また，候補数の値を 50〜300 に設定した際の候補提示数 7〜9 における分類精度を示しているのが図 6.3 である．ここでも 100〜150 程度が良い結果が出ていることがわかる．

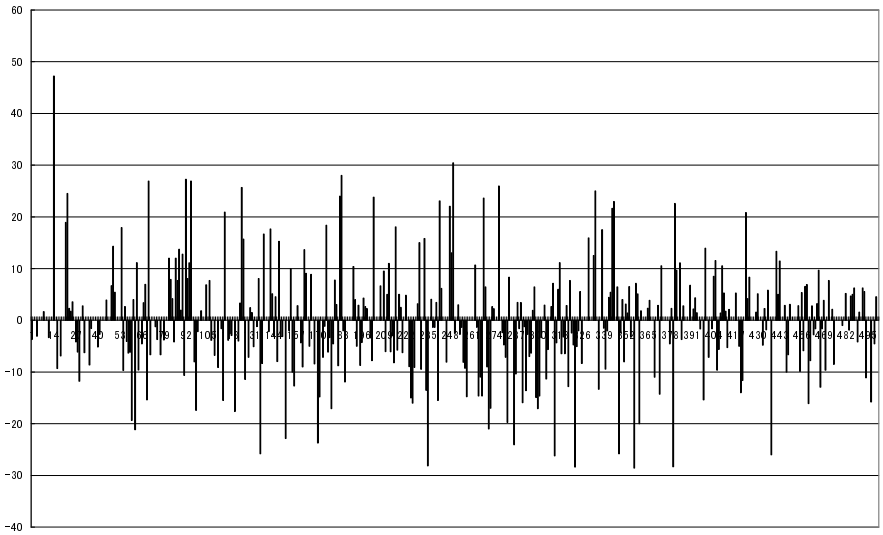


図 6.1: 精度下位からソートした元問題例と対応する ROI 後の精度比較

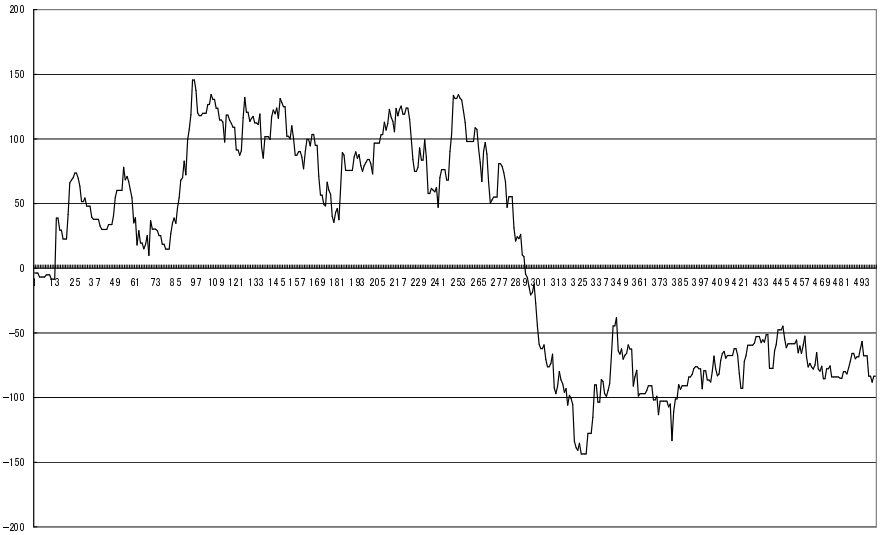


図 6.2: 精度比較した値の下位からの総和

また，領域探索を決定する閾値に従った計算時間の推移を調査したものが図 6.4 である．このグラフ上での 100[%] はデータセット内の全てのクラスについて領域探索を行った時に要する計算時間を示している．図を見るとわかるように，基本的には探索対象 2 値問題を増加するたびに単調に増加している傾向がある．

以上のような実験結果から，今回の提案手法での探索対象を決定する閾値の設定は，精度と計算量のバランスを考え 125 としている．また閾値パラメータを様々に変えて実験を行い，この値付近が妥当であるという結果を得た．

領域探索を行わず通常の手法で分類を行った中での分類下位の 2 値問題に属するクラスについて図 6.5 に示す．ここでの縦軸は分類精度が下位 300 の 2 値問題に対象のクラスが属していたらカウントしていったものである．またその実際のクラスの画像の一例を示したものが図 6.6 である．

実験過程において，2 値分類問題において精度が低いものに属しづらいクラスを示したものが図 6.7 である．

領域探索を行わない通常の特徴量を用いた結果，分類精度下位の 2 値問題に属しやすいクラスとして考えられるのは，そもそも SIFT 特徴量の段階からあまり抽出しにくいクラスや，やはり背景がノイズになっているようなクラスが挙げられる．単純な特徴だけでも分類しにくい問題に属しづらいクラスは，SIFT 特徴量が抽出しやすいもの，そもそも学習データ背景などが無いもの，そして背景がコンテキストとして働いているものが多い．

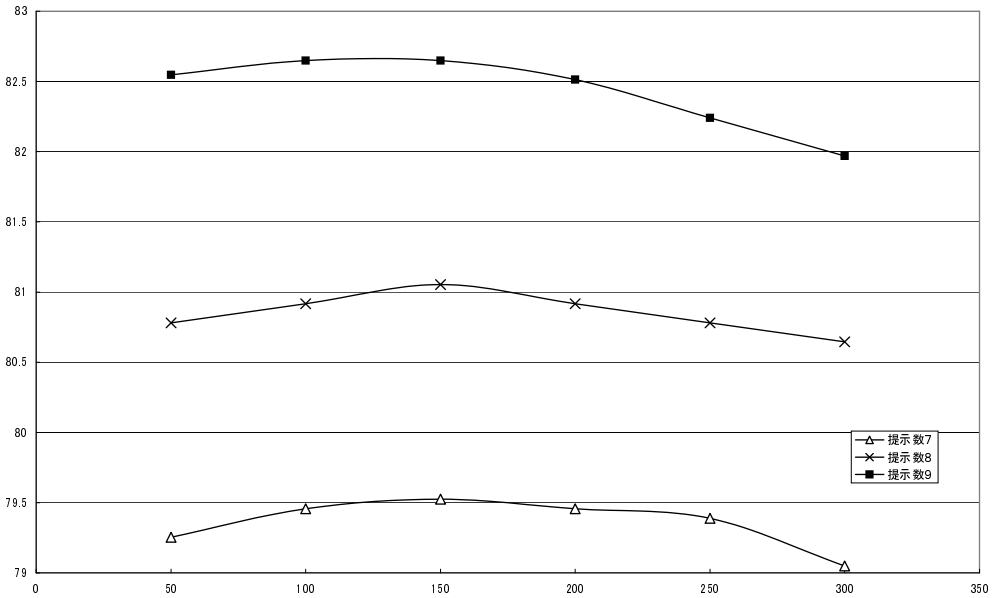


図 6.3: 候補提示数 7〜9 における，探索問題例数に対する精度

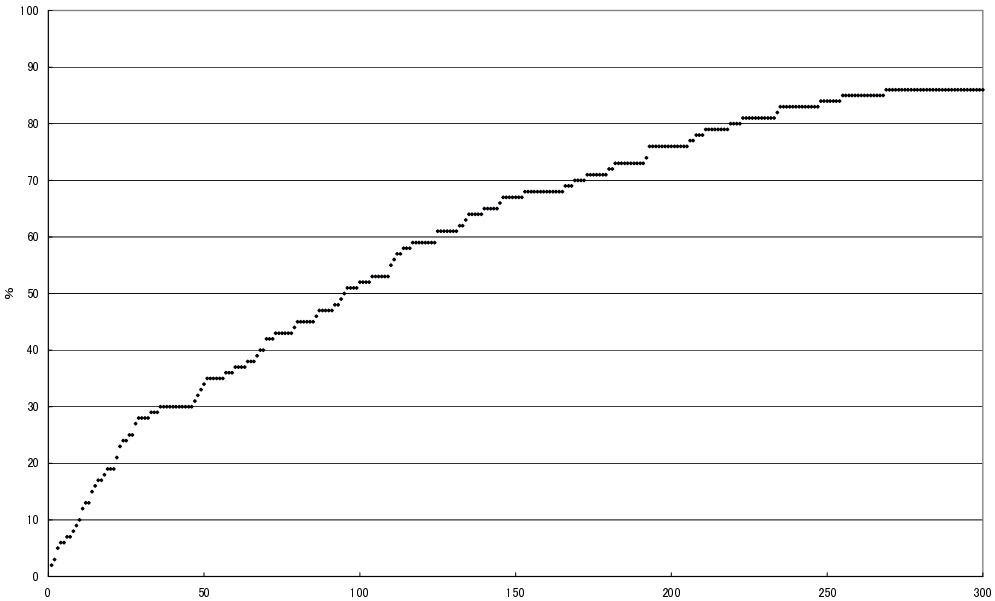


図 6.4: 探索対象決定閾値に従った計算時間の推移

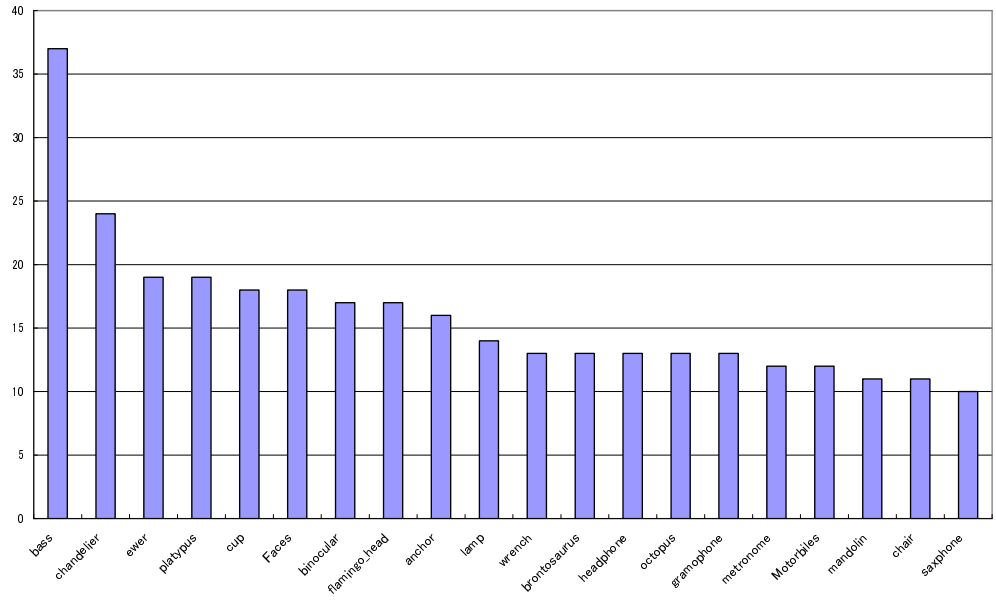


図 6.5: 分類精度下位の 2 値問題に属するクラス

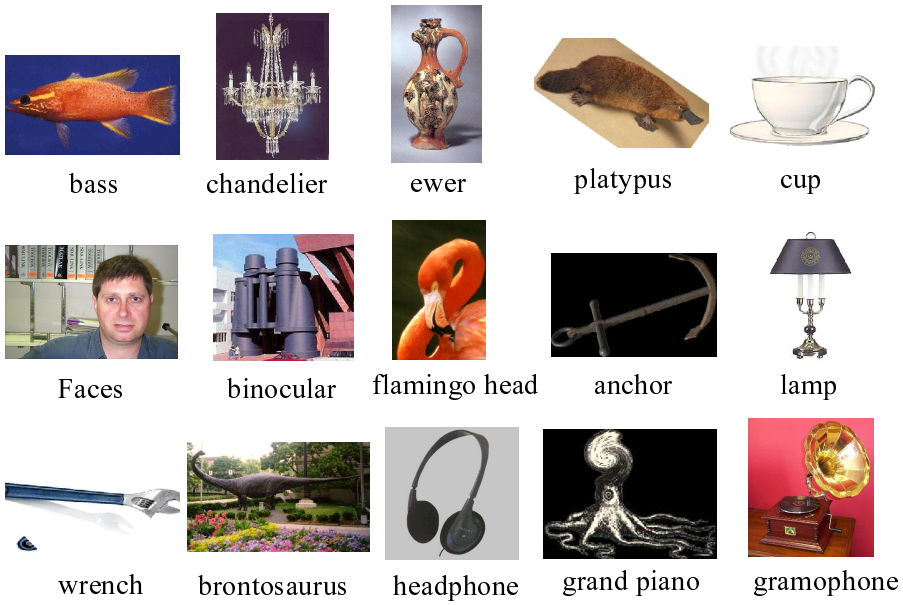


図 6.6: 分類困難な問題に属しやすいクラス

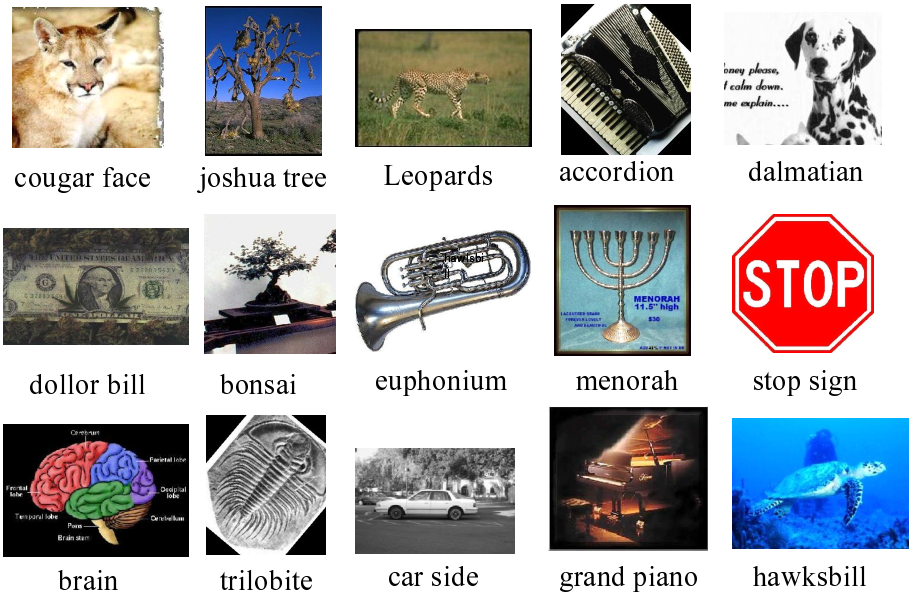


図 6.7: 分離困難な問題に属しづらいクラス

6.2.2 提案手法を適用した結果

今回の提案手法を適用した際の候補提示数に対する精度と，元の訓練データの絞込みを行わないでそのまま Bag-of-keypoints 表現を用いて SVM で分類した際の候補提示数に対する精度を SVM に適用するカーネルごとに比較したものが図 6.8・6.9 である．ここでは探索対象の 2 値問題の数は 125 と設定している．

どちらのカーネルでの比較を見ても，全体的な精度傾向が上昇している事がわかる．特に今回の実験では Spatial Pyramid Kernel の精度傾向が良くなっている．

また，以上を踏まえてクラスタ数 300 においても実験を行った．事前実験よりより高い分類精度の傾向が出る Spatial Pyramid Kernel を用いた．提案手法適用前後での精度を比較したのが図 6.12 である．クラスタ数 200 での実験結果よりも全体的な精度傾向が高くなっているのがわかる．精度 90%に収束する候補数は 17 から 15 に，精度 95%に収束する候補数が 27 から 25 になっている．

6.2.3 実験の考察

実験結果を見ると，提案手法を適用することで分類性能が向上していることがわかる．これは提案手法の有効性が示されているといって良い．また，計算時間も全体で学習データセットの最適化を行うよりは削減できることが出来る．計算時間については，基本的に探索問題数を増やすごとに単調に増加していくが，３０程度から増加問題数に対する増加時間は抑えられている．そして２５０程度からはその増加割合は更に抑えられる．これは，単純な特徴量だけで確実に分類できるような

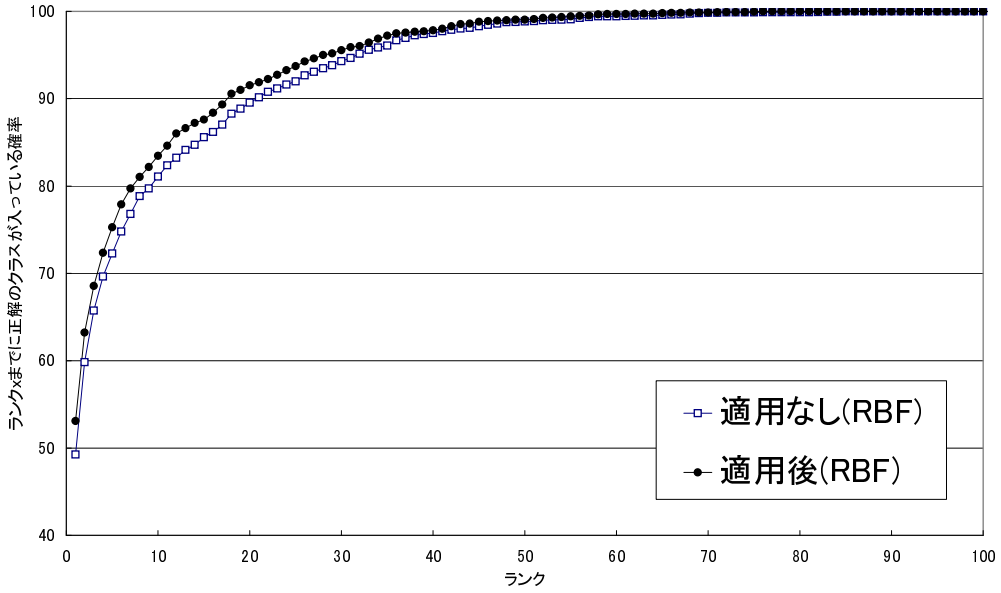


図 6.8: 提案手法適応前後での比較（RBF Kernel）

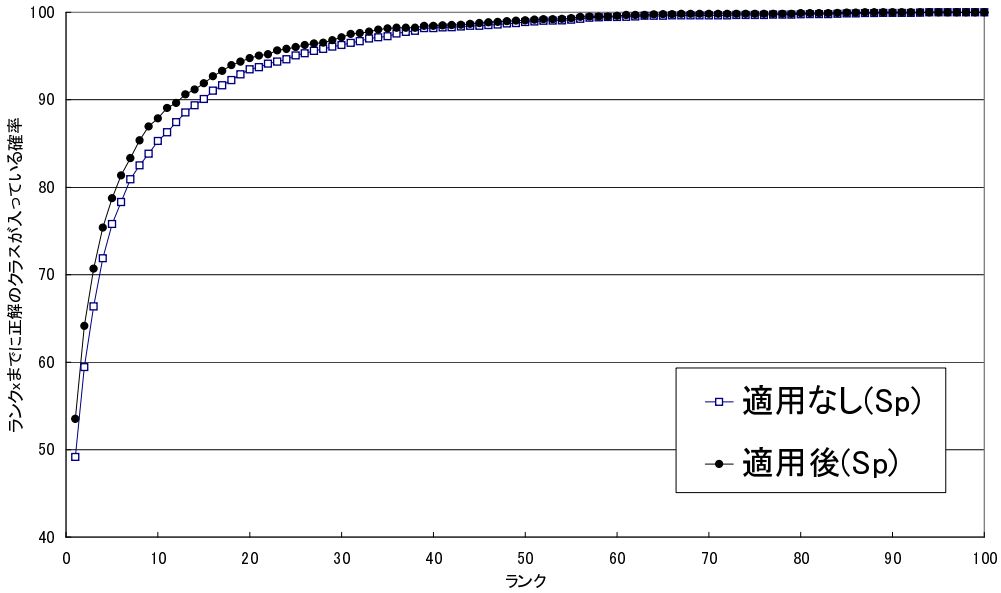


図 6.9: 提案手法適応前後での比較（Spatial Pyramid Kernel）

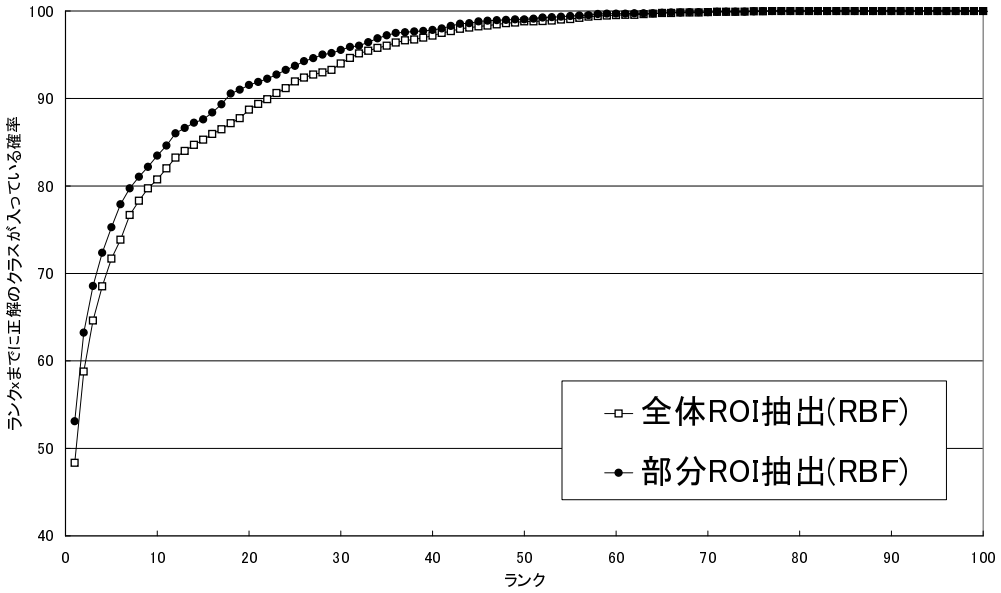


図 6.10: ROI を全探索した結果と部分探索した結果の比較（RBF Kernel）

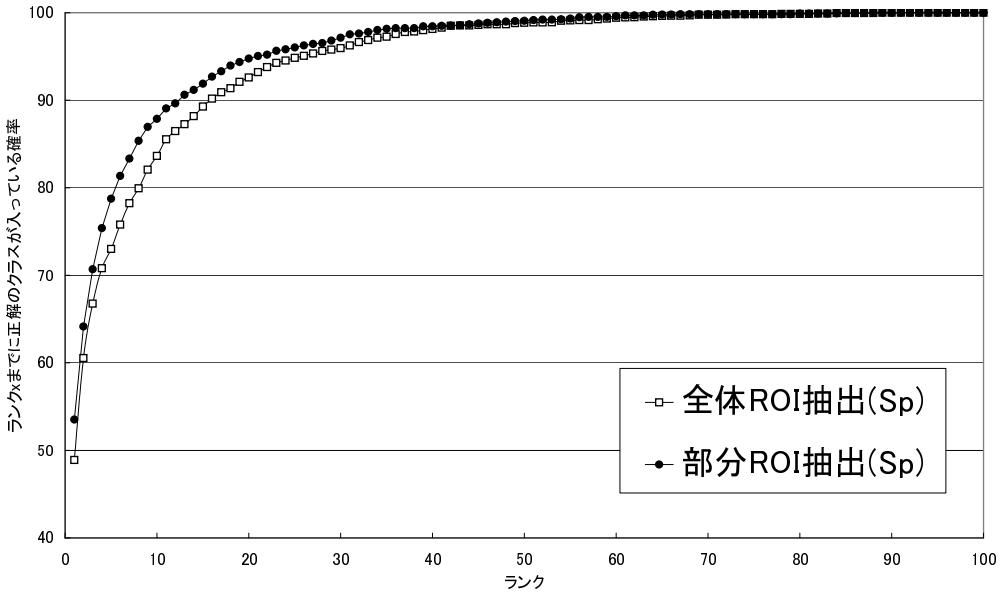


図 6.11: ROI を全探索した結果と部分探索した結果の比較（Spatial Pyramid Kernel）

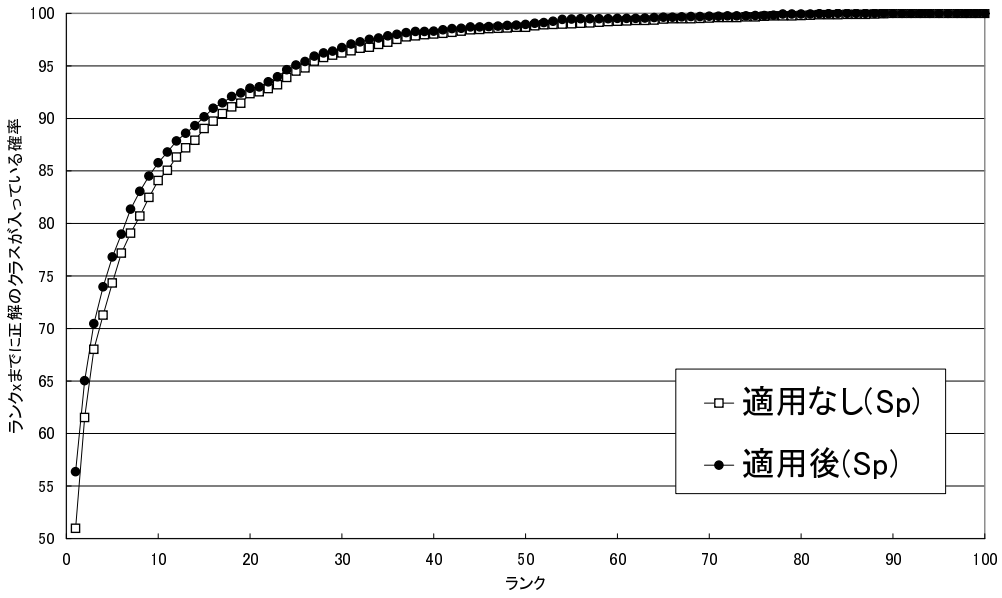


図 6.12: 提案手法適応前後での比較（クラスタ数 300・Spatial Pyramid Kernel）

クラスがデータセット中に 1 割強は存在しており，そのようなクラスはあまり領域探索をされる事がないからであると考えられる．

6.3 本章のまとめ

本章では第５章で提案した手法を適用した結果について元の手法（領域探索を行わないものと，全体のデータセットに対して Region of Interest の探索を行ったもの）と比較して精度的な上昇を確認した．計算量に関しても全体で Region of Interest を探索した場合と比較して，妥当な閾値で約 40〜60%程度で可能である．

次章で本研究のまとめとして結論と今後の課題について述べる．

第7章　終わりに

7.1　結論

本研究では、一般物体認識の複数候補提示下における精度を上昇するために，One-versus-One で構成される Support Vector Machine を用いて，交差検定によって分類困難な２値問題を発見し必要クラスの選択を動的に行い，学習データ中の Region of Interest の領域学習をして，各々識別平面を生成する分類手法を提案した.

提案手法を適用した結果として，候補提示数に対する全体的な精度の上昇を確認した．訓練データセット全体に対して Region of Interest の探索を行ったものと比較しても，精度・計算量の両観点から見て良い結果を得た.

7.2 今後の課題

今後の課題として以下のことが挙げられる．

より複雑な特徴表現や大量の特徴量を組み合わせた構成の考慮 今回使用した局所特徴量は SIFT の中でも最も基本的な形であり，その最終的な特徴表現も簡易な形のものを用いている．これは数多くの既存研究で様々な特徴量を使用した手法が提案されているが，そもそも単純な特徴量でどこまで分類可能性があるのかという事をまず知見として得たかった動機からきており，今回の報告では，最終的な提案手法でも単純な特徴量をそのまま使用するに至った．しかし，やはり分類をより高精度のものにするためには，現在用いている特徴量では限界があり，そもそも考えられうる大量の特徴量を全て入れてその調整を考慮に入れた分類の構成になってくると考えられる．

学習器の考慮 今回の実験では，事前実験より，テストデータに対するクラスを一意に推定する精度で SVM が優れた能力を発揮していたので，２クラス分類 SVM の構成を前提とした手法を提案してきた．また，汎化能力が秀でている事も SVM を選択した理由である．本研究は実用性を考慮して，複数候補提示下での分類性能に対して主眼に置いている．よって，実際の距離関数がダイレクトに順位付けに作用するような学習器を用いた構成にしていった方がより妥当なものになる可能性もある．また，多クラスの識別器を直接構成する Multiclass-SVM にそのまま本手法を組み込むなどの方法が考えられる．

実装の並列化 今回提案した手法部分には並列化可能箇所が多く存在する．特に各クラスに対する領域探索，交差検定，ペアの識別関数を構成する段階などは処理が完全に独立なので容易であると考えられる．また，今回効果的でないと判断して最終的な実装に見送ったパラメータの動的決定や，特徴量の更新などの処理も，並列化は可能であり，それを前提として更に検証を行えば分類性能向上の可能性も出てくると考えられる．

参考文献

[1] H. Zhang, A.C. Berg, M. Maire, J. Malik. *SVM-KNN: Discriminative Nearest Neighbor Classification for Visual Category Recognition*, Proc. of IEEE Computer Vision and Pattern Recognition,pp.2126-2136, 2006.

[2] S. Lazebnik, C. Schmid and J. Ponce. *Spatial Pyramid Matching for Recognizing Natural Scene Categories*, Proc. of IEEE Computer Vision and Pattern Recognition,pp.2169-2178, 2006.

[3] G. Wang, Y. Zhang and L. Fei-Fei. *Using Dependent Regions for Object Categorization in a Generative Framework*, Proc. of IEEE Computer Vision and Pattern Recognition,pp.1597-1604, 2006.

[4] K. Grauman, T. Darrell. *Pyramid Match Kernels: Discriminative Classification with Sets of Image Features*, Proc. of IEEE International Conference on Computer Vision,pp.1458-1465, 2005.

[5] J. Mutch, D.G. Lowe. *Multiclass Object Recognition with Sparse, Localized Features*, Proc. of IEEE Computer Vision and Pattern Recognition,pp.11-18, 2006.

[6] A. Bosch, A. Zisserman, X. Munoz. *Image Classification using Random Forests and Ferns*, Proc. of IEEE International Conference on Computer Vision, 2007.

[7] M. Varma, D. Ray. *Learning The Discriminative Power-Inveriance Trade-off*, Proc. of IEEE International Conference on Computer Vision, 2007.

[8] L. Fei-Fei, R. Fergus and P. Perona. *One-Shot Learning of Object Categories*, Proc. of IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006.

[9] A. Holub, M. Welling, and P. Perona. *Combining generative models and fisher kernels for object recogntion*, Proc. of IEEE International Conference on Computer Vision,pp.136-143, 2005.

[10] T. Serre, L. Wolf, and T. Poggio. *Object recognition with features inspired by visual cortex*, Proc. of IEEE Computer Vision and Pattern Recognition,pp.994-1000, 2005.

[11] L. Fei-Fei, R. Fergus and P. Perona. *Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories*, Proc. of IEEE CVPR Woekshop of Generative Model Based Vision, 2004.

[12] C. Cortes, V. Vapnik. *Support-Vector Networks*, Machine Learning, 20, pp.273-297, 1995.

[13] M.O. Stitson, J.A.E. Weston, A Gammerman, V.Vork, V. Vapnik, *Theory of Support Vector Machines*, Technical Report CSD-TR-96-17, Department of Computer Science, Royal Holloway College, University of London, Dec. 1996.

[14] Caltech 101 image dataset. *http://www.vision.caltech.edu/Image_Datasets/Caltech101/*

[15] D. G. Lowe. *Object recognition from local scaleinvariant features*, Proc. of IEEE International Conference on Computer Vision, pp.1150-1157, 1999.

[16] T. Hastie, R. Tibshirani. *Classification by pairwise coupling*, The annals of Statics,26,pp.451-471, 1998.

[17] T.G. Ditterich, G. Bakiri. *Solving multiclass learning problems via error-correcting output codes*, Journal of Artifical Intelligence Research,2,pp.263-286, 1995.

[18] E.L. Allwein, R.E. Schapire and Y. Singer. *Reducing Multiclass to Binary: A Unifying Approach for Margin Classifiers*, Journal of Machine Learning Research,2000.

[19] F. Aiolli, A. Sperduti *Multiclass Classification with Multi-Prototype Support Vector Machines*, Journal of Machine Learning Research,2005 pp.817-850.

[20] 藤吉弘亘 「Gradient ベースの特徴抽出 - SIFT と HOG - 」, 情報処理学会 Computer Vision・Image Media 研究会,160,pp.211-224, 2007.

[21] J. Weston, C. Watkins. *Multi-class Support Vector Machines*, Technical Report CSD-TR-98-04, 1998.

[22] C.D. Manning, H. SchFutze. *Foundation of Statistical Natural Language Processing*, The MIT Press, 1999.

[23] M. Gonen, A. Gonul and E. alpaydm. *Multiclass Posterior Probability Support Vector Machines*, Proc. of IEEE Transaction on Neural Networks, 2008.

[24] 柳井啓司 「一般物体認識の現状と今後」, 情報処理学会 Computer Vision・Image Media 研究会, 2006.

[25] P. Duyhulu, K. Barnard. *Object Recognition as Machine Translation: Lerning a Lexicons for a Fixed Image Vocabulary*, Proc. of European Conference on computer Vision,pp. :97-112, 2002.

[26] G. Csurka, C. Bray, C. Dance, L. Fan. *Visual categorization with bags of keypoints*, Proc. of European Conference on computer Vision,pp.1-22, 2004.

[27] R. Fergus, P. Perona and A Zisserman. *Object Class Recognition by Unsupervised Scale-Invariant Learning*, Proc. of IEEE Computer Vision and Pattern Recognition,pp.264-271, 2003.

[28] T. Kadir, M. Brady. *Scale, Saliency and image description*, International Journal of Computer Vision,pp. :97-112, 2001.

[29] J. Weijer, C. Schmid. *Coloring Local Feature Extraction*, Proc. of European Conference on computer Vision,pp. :334-348, 2006.

[30] J. Willamowski, D. Arregui, et.al. *Categorizing Nine Visual Classes using Local Appearance Descriptors*, CPR 2004 Workshop Learning for Adaptable Visual Systems Cambridge, 2004.

[31] Y. Ke, R. Sukthankar. *PCA-SIFT: A More Distinctive Representation for Local Image Descriptors*, Proc. of IEEE Computer Vision and Pattern Recognition,2004.

[32] A.C. Berg, T.L. Berg and J. Malik. *Shape Matching and Object Recognition using Low Distortion Correspondences*, Proc. of IEEE Computer Vision and Pattern Recognition,pp.26-33, 2005.

[33] A.C. Berg, J. Malik. *Geometric Blur for Template Matching*, Proc. of IEEE Computer Vision and Pattern Recognition,pp.607-614, 2001.

[34] Weka 3 - Data Mining with Open Source Machine Learning Software in Java *http://www.cs.waikato.ac.nz/ ml/weka/index.html*

発表文献

[1] 栗田哲平, 近山隆.　多クラス Support Vector Machine を用いた一般物体認識での複数候補提示下における分類性能の傾向.　情報処理学会研究報告 CVIM-165 (PRMU・MVE 共催), pp.251-258, 11 2008.

[2] 栗田哲平, 三輪誠, 近山隆. 将棋盤を対象とした画像情報を用いた自動局面認識手法.　第 12 回ゲームプログラミングワークショップ 2007，pp.172-179 , 11 2007.

他，共著 1 件

謝辞

本研究を進めるにあたり，多くの方にお世話になりました．

近山隆教授には数多くの研究への助言を頂きました．田浦健次朗准教授には、プレゼンテーションの際などに多くのご指導、ご教示を頂きました．また，IRT 研究機構の横山大作特任助教，近山研究室博士の三輪誠さにも，度々研究を進める際に相談に乗って頂きました。

同期の皆様とのコミュニケーションは研究を進める際に良い頭の切り替えになりました．他の近山・田浦研究室の皆様にも公私にわたり多くの助言を頂きました。

ここに、心より感謝の意を表します。

平成 21 年 2 月 3 日