

位置情報を利用した情報ハンドリング技術に関する考察

山口高康 高畑 実 本郷節之

NTT ドコモ マルチメディア研究所 〒239-8536 神奈川県横須賀市光の丘 3-5

E-mail: t.yama@mml.yrp.nttdocomo.co.jp, takahata@mml.yrp.nttdocomo.co.jp, hongo@mml.yrp.nttdocomo.co.jp

あらまし モバイル環境における位置情報を利用した情報ハンドリング技術を導入した、モバイルカメラによる撮影画像の対象判別を利用するマルチメディア情報検索技術について考察する。システム実現のために検討した2つの要素技術、すなわち撮影画像を用いた対象判別技術、位置情報を利用する対象候補の絞り込み技術について述べる。提案するシステムのプロトタイプを作成して屋外での対象判別実験を行った。モバイルカメラによって10対象(102画像)を撮影して対象判別を行い、90%以上の正解率を得た。

キーワード マルチメディア情報検索、位置情報、モバイルカメラ、対象判別

The Information Handling Technology using Location Information Takayasu YAMAGUCHI, Minoru TAKAHATA and Sadayuki HONGO

Multimedia Labs. NTT DoCoMo, Inc. 3-5 Hikarinooka, Yokosuka-shi, Kanagawa, 239-8536 Japan

E-mail: t.yama@mml.yrp.nttdocomo.co.jp, takahata@mml.yrp.nttdocomo.co.jp, hongo@mml.yrp.nttdocomo.co.jp,

Abstract This paper discusses the multimedia information retrieval and the presentation system using object distinction of a photographic image with a mobile camera. We apply location information to the system. Two essential techniques, object distinction using photographic images, extraction of candidates of planar advertisement signs using location information, are examined for the proposed system. We created the prototype of the object distinction system to perform outdoor experiments of object distinction techniques. Object distinction was performed with ten objects, and 90% or more of precision rate was achieved.

Keyword Multimedia Information Retrieval, Location Information, Mobile Camera, Object Distinction

1. はじめに

1.1. 研究背景

現在、我々はモバイルマルチメディアの世界の創造を掲げ、ユビキタスネットワークの構築、パーソナルサービスの実現、シームレス化に取り組んでいる。サービスとインフラストラクチャの高度化により、我々は各個人が携帯端末からいつでもどこからでもインターネットや LAN などの膨大な情報源にアクセスできる環境を提供している。さらに第3世代国際標準規格である IMT-2000 を開発・導入して、高速ワイヤレスアクセス環境を整えつつある。我々はこれらを核にシームレス化とユビキタスネットワークの構築を強力に推し進めている。

パーソナルサービスの実現については、近年のサービスとインフラストラクチャの高度化につれて、個人が持ち運びできる携帯端末で、実世界やネットワーク内に混在する情報の中から個人にあった情報のみをタイムリーに取り出す事の必要性がますます高まっている。その要素技術としては、モバイル環境において簡便で効果的なヒューマンインターフェース技術や情報

ハンドリング技術が挙げられる。

本研究ではこれらの背景を踏まえ、新たにモバイル環境特有の位置情報を利用した情報ハンドリング技術の一つとして、モバイルカメラによる撮影画像の対象判別を利用するモバイル環境下でのマルチメディア情報検索技術とその動作検証実験の結果について報告する。

1.2. モバイル環境下でのマルチメディア情報検索

近年、Webに限らずデジタルテレビや i-mode コンテンツ等において、マルチメディア化が進んでおり、ネットワークを通じて取り出すことのできる情報は文字や画像、動画、音声が複数混在して構成されている。これらのマルチメディアコンテンツは XML などの記述言語によって、データとその分類(タグ)に分けて記述されている。分類項目(タグ付けの手法)の統一の難しさを残しつつも、RDF 等の変換手法をはじめ情報の分類法に関する今後の方向性が見えてきている。

我々の生活の中においても、情報を効率的に管理して利用しやすくするという事が行われており、我々は分類という作業を行って情報を整理する。しかし、実

世界から取り入れられる多種多様な情報には、それらの情報を分類するためのタグはついていない。ある実世界の対象について詳しく知ろうとすれば、人が能動的に情報を引き出す必要がある。対象に関する知識が不足する場合には、必要に応じて書物や機器類を利用するなどの労力が強いられる。携帯端末に取り込まれた情報をいかに上手く分類する(タグを付ける)のかが、タグを付けた情報が行き交っている計算機ネットワークと、携帯端末を持つユーザとの親和性を高めて、個人にあった情報サービスを提供する為の課題である。

近年の携帯端末の進化はめざましく、モバイルカメラを搭載した機種の登場により、画像をひとつの自己表現として利用するコミュニケーションが進展する一方で、今後、モバイルカメラを周囲の情報を取得するセンサとして利用する方法も期待できる。モバイルカメラで撮影した画像から対象を判別し、対象の情報へ簡単にアクセスできれば、単に手間を軽減するだけではなく、革新的かつ高付加価値サービスの実現を通して、様々な対象とのコミュニケーションを喚起し、より豊かな生活が可能になるであろう[1]。センサとしてのモバイルカメラの利用は、これまで行ってきたキー入力の作業を軽減させ、操作性を大幅に向上させる。このような夢へ向けた一つの実現手段として、本稿で提案する撮影画像を用いた対象判別技術が位置付けられる。

本稿ではまず、高次元ベクトルの判別を高速に処理することに適しており、汎化性能に優れている Support Vector Machine(SVM)[2]を利用した看板・ポスター等の平面物体の判別手法を提案する。SVMはその高速判別処理の能力と汎化性能を活かして、画像などの高次元ベクトルを高速に判別処理する手法が報告されており[3][4]、本テーマである看板・ポスター等の平面物体の判別にも適していると考えられる。対象の判別には学習と判別の処理のトータルコストを踏まえて、対判別を用いる。

次に、本稿で用いる対判別を行うことによって、対象数増加に伴い処理速度が低下する問題をとりあげ、位置を利用することによってこの問題を改善する手法を提案する。対象を限定するには、位置以外にもユーザの趣味・趣向や、目的等によって絞込みを行うことができるが、モバイルカメラで撮影できる対象はユーザの近くにあるので、位置での絞り込みは有効であると考えられる。

最後に、実際に本システムをモバイル環境でテストした対象判別実験と結果について述べ、まとめを行う。

2. 撮影画像を用いた対象判別技術

2.1. 問題設定

本稿ではモバイルカメラによる撮影画像の対象判

別を利用するモバイル環境下でのマルチメディア情報検索技術について述べる。本システムでは、予め人間が判別対象の画像と位置情報とその他関連する情報を本システムに登録して学習させておく。ユーザがカメラ付きの携帯端末で対象を撮影した場合に、撮影した対象画像と登録・学習してある対象画像とを比較して最も撮影対象に類似している対象を判別する。

モバイル環境においてユーザが判別する対象としては様々なものが考えられるが、本研究では看板・ポスター等の平面物体を判別対象とする。理由は以下のように考えられる。

- ① 看板で宣伝されている情報はネット上に関連情報が多く存在
- ② ユーザが判別した看板は、ユーザの興味や利用目的と多く合致
- ③ 多くの看板は地理的な位置が固定
- ④ モバイル環境での情報検索サービスの一つとして有用

近年 SVM と呼ばれる新たな分類手法が提案され、各種の分類問題に適用され、効果をあげている。判別の手法としては、SVM、隠れマルコフモデル、ニューラルネットワーク、決定木など様々な手法があるが、SVMは汎化性能に優れており、高次元ベクトルの判別を高速に処理することに適している。そこで我々は一つのアプローチの方法として、この SVM を利用して対象判別を試みる。

SVMは1960年代にVapnikらによって提案された超平面による特徴空間の分割法に起源をもつ[2]。1992年に非線形の判別関数を求めるアルゴリズムが提案された。非線形の変換関数を用いて特徴ベクトルを高次元の特徴空間に変換し、そこで線形判別を行うというアルゴリズムである。1995年に訓練データが線形関数で分離できない場合に対する判別関数構成のアルゴリズムが考案され、SVMに対する基本的なアルゴリズムができあがった。SVMの研究の進展に伴い、SVMの特性の理論的考察、アルゴリズムの高速化の研究、各種問題への応用が行われ、多くの成果が得られている。[2],[3],[4],[5],[6]

本システムでは、モバイル環境における取得対象画像の色空間写像特徴量をキーとして対象を判別する。対象の登録は、対象の撮影画像から対象に基づく色空間写像特徴量を抽出し、ユーザの撮影位置の特徴の変化を含めて学習することによって行う。対象の判別は、予め位置情報によって判別対象候補を絞り込み、絞り込んだ対象候補のみを高速な判別アルゴリズムによって演算処理することで対象を判別する。後述する位置を用いて判別対象学習パラメータと情報を分散管理する手法と組み合わせることで、素早いモバイル環境で

の対象判別を実現する。

本提案方式は判別の一手法であり、従来の認識の手法（「不変特徴量と特徴空間」、「部品と構造的記述」、「アラインメント法」など）とは性質の異なる手法である。画像認識の分野では、前提条件や背景知識に基づいて画像中で何かしらの不変量を抽出し、対象の部分画像を見つけてから認識処理を進める手法がある。しかし、モバイル環境下で撮影した対象の画像は様々に変化し、全ての看板に対して共通する事柄は限られる。予め対象に大きさの前提条件を定めるのは難しいので、本稿では「対象は画面全体に写っている（ユーザは対象が画面全体に写るように対象を撮影する）」ことを前提条件としてユーザから入力を求める。計算機が撮影画像の不変特徴量や部分を計算して認識を行うのではなく、取得した画像をそのまま色の接合・包括に基づく画像特徴量に変換して登録・学習・判別する。モバイル環境下での撮影によって対象画像が変化しても、ユーザに使用条件の制約を与えて、対象と対象画像との相関値を高めることで、対象判別の正解率を向上させる。

対象が画面全体に写っていても、カメラの撮影位置により発生する射影歪の他、モバイル環境下での撮影には様々な外的要因があり、全てのケースに必ずしも良好な結果が得られない。本稿では以下、射影変換の補正を施さない場合（もしくは、上手く補正できない場合）として議論を進める。

射影変換の補正を施さない場合には、画像中の対象に移動・回転・変形が生じる。画像中から部分を切り出さずに判別を行うには、射影変換の歪が生じて大きく変化しない特徴量を画像全体から抽出する必要がある。画像全体から抽出可能な特徴量としては、濃淡、色、周期性、形状、テクスチャ、エッジ、テクスチャエッジなど様々なものがある。

本稿では、これら様々な特徴量の中から、一つの判別手法として配色や色の配置に関わる特徴を抽出して判別を行う手法を示す。看板は人工物であり、看板の作成者は他者に情報を伝える為に、看板を色で塗り分けることが多い。また、交通等の標識については一定の規格がある[7][8]。この配色と色の配置に関わる情報を判別に利用する。

一般に、撮影画像の対象の色と実世界での色を完全一致させることは難しく、まして撮影画像の色と作成者の意図した色を一致させることも難しい。そこで、まず大幅な減色を行って画像中の色を大別してから、大別した色の配置の特徴を抽出することを試みる。

配置関係としては、1項で位置、方向、大きさなど、2項間で包含関係、隣接関係、相対位置（距離・方向）など、3項間で包含関係、隣接関係、相対位置（距離・

方向、2つの間）などがある[9]。

本稿では、これら様々な配置関係の中から、色の包括と接合の配置関係の特徴量として用いる。

包括と接合を擬似的に抽出するアルゴリズムの詳細は次節以降で述べるが、これを本稿では色空間写像特徴量と呼ぶ。色空間写像特徴量は対象に基づく特徴量であり、この色空間写像特徴量を判別に用いる。色空間写像特徴量はユーザの撮影位置の変化によって変化する特徴量であり、一般に認識手法で用いられる不変特徴量とは異なる。

2.2. 色空間写像特徴量抽出法

色空間写像特徴量の抽出は、減色、領域分割、疑似接合量、疑似包括量、色空間写像ベクトルの作成の処理によって行う。

減色には従来から提案されている手法を用いた[9]。本稿ではRGBの中央値を閾値として黒、青、赤、紫、緑、水、黄、白の8色(n色)に減色した。

疑似接合量を求めるには、 $B_1 \times B_2$ [pixel]のボックスを撮影画像中で移動させ、そのボックスにかかる範囲にある色の組み合わせをカウントして疑似接合量とする。n色の組み合わせの疑似接合量を $\{(n^2-n)/2\} + n = 36$ [次元]ベクトルで表現する。

疑似包括量を求めるために、まず従来から提案されている手法を用いて領域分割を行う[9]。次に基点を領域分割を行った画面中で移動させ、その基点から発散するように複数の方向へ向けてラベリングされた領域をサーチする。複数の方向全てに同じラベルの領域を検出した場合には、それに基点のラベルの領域が疑似包括されているとする。疑似包括する領域の色と疑似包括される領域の色の組み合わせをカウントして疑似包括量とする。n色の組み合わせの疑似包括量を $n^2 = 64$ [次元]ベクトルで表現する。

得られた疑似接合量と疑似包括量にそれぞれ重みの係数を乗算したものを一つのベクトルとみなし、本稿ではこれを色空間写像特徴量と呼ぶ。また、上記の一連の処理を行うプログラムを、本稿では画像特徴写像演算エンジンと呼ぶ。

2.3. 登録・学習手法

図1に示すように、対象 $m=1,2,3$ の画像を登録・学習する手法について述べる。ハンバーガーショップの看板(対象: $m=1$)の学習用画像を3枚撮影し、レストランの看板(対象: $m=2$)の学習用画像を4枚撮影し、バイクショップの看板(対象: $m=3$)の学習用画像を2枚撮影する。色空間写像演算エンジンによりそれぞれの画像に対応する色空間写像特徴量を抽出して、写像ベクトルを作成する。これらの対象ごとの写像ベクトルの組み合わせをさらに高次元空間に写像する。高次元空間において写像ベクトルを分割する境界を SVM

により求めて、判別ベクトルとバイアスを作成する[2]。判別ベクトルとバイアスと対象の位置情報とサービスエリアの情報をデータベースに格納する。

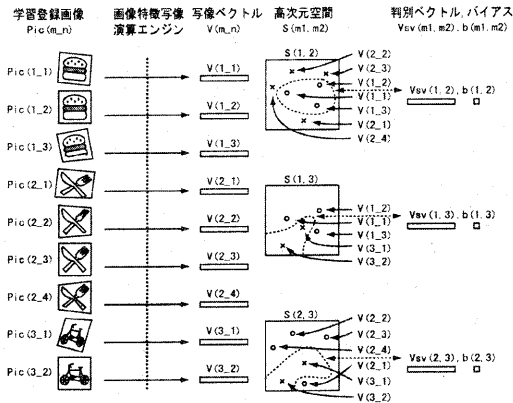
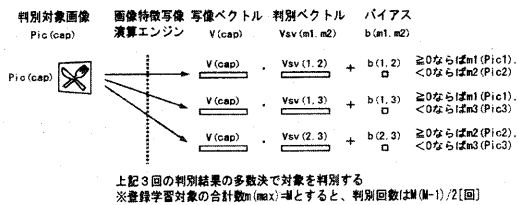


図1 対象画像の学習

2.4. 判別手法

図2に示すように、未知の判別対象画像が対象 $m=1,2,3$ のどれに該当するか判別する手法について述べる。色空間写像演算エンジンにより撮影した判別対象画像から色空間写像特徴量を抽出して、写像ベクトルを作成する。判別対象画像が撮影された位置をサービスエリアに含む判別ベクトルと写像ベクトルとの内積をとってスカラ値とし、これにバイアスを加算して、2値判別を行う。判別結果の多数決で最も票の多い対象を解とする。



上記3回の判別結果の多数決で対象を判別する
※登録学習対象の合計数(max)をMとすると、判別回数は $M(M-1)/2$ [回]

図2 対象画像の判別

2.5. SVM利用の問題点

SVMで2値判別を行い、上記のアルゴリズムのように判別を行う場合には、対象数が n のとき $(n^2-n)/2$ [回]の判別を行う必要があり、処理速度は O^2 となる。対象数が20で190回の判別を1秒で行うとすると、図3に示すように対象数の増加に応じて、判別時間の増大が見込まれる。

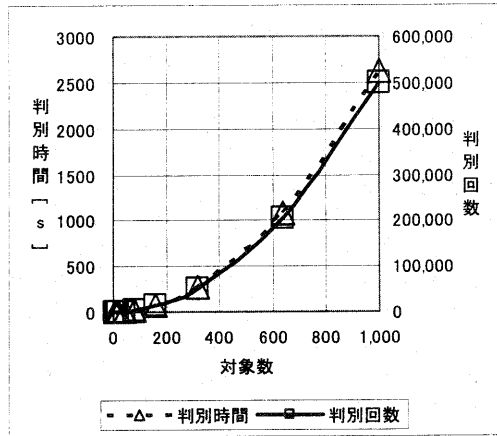


図3 対象数に対する判別時間・回数の変化

3. 位置情報を利用した対象候補の絞り込み技術

モバイルカメラで撮影する対象はユーザの近くにあるので、判別を行う対象をユーザの位置の近くの位置にある対象に絞り込む。

システムを図4に示すように、携帯端末、基地局、ネットワーク網、サーバ、データベースによって構成する。登録・学習データは予めデータベースに格納しておく。携帯端末は、携帯端末もしくは移動通信網による測位情報と共に、撮影画像の情報をサーバへ送信して処理する。サーバは携帯端末の位置情報を元に対象候補を絞り込み、撮影画像の特徴量を元にさらに対象候補を絞り込む。サーバは判別結果を携帯端末に送信する。携帯端末はサーバから受信した判別結果をユーザに提示、もしくは携帯端末自身が判別結果を元にさらに処理を行う。

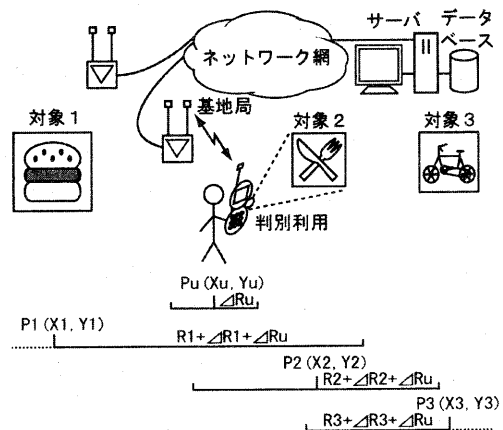


図4 位置情報を利用した対象判別

2章で述べた SVM の問題点を回避するために、位置情報を利用する対象候補の絞込み技術を新たに提案する。位置情報を利用して対象候補を絞り込むことで判別回数を減少させる。本稿では、位置情報を利用する具体的な対象候補の絞込みの方法と、広域なサービスエリアで切れ目の無い対象判別を提供する為に必要な対象の登録・学習情報の分散管理方法について述べる。

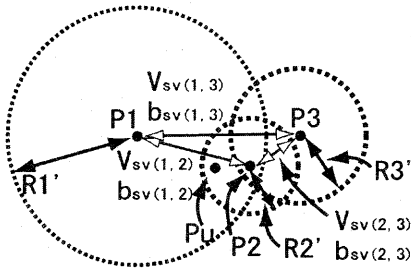


図5 位置情報を利用した対象判別ベクトルの管理

図5に示すように、個々の判別対象の位置(P1,P2,P3)を中心に個々の判別対象のサービス範囲を設定する(看板であればサービス範囲は一般に扇形になるが、簡単の為に円形のサービス範囲のモデルを用いる)。学習は、判別対象のサービス範囲が重なる部分が存在する場合に、両者を判別するベクトル $V_{sv}(m1,m2)$ とバイアス $b(m1,m2)$ を作成する。サービスエリアが重ならない判別対象の組み合わせについては、判別ベクトルとバイアスを計算する必要は無い。判別対象の追加または削除は、追加または削除を行う対象のサービス範囲と重なる部分だけを更新すればよい。これによって切れ目の無い柔軟なサービスエリアを構築する。

登録・学習時に測位した判別対象の位置の誤差を ΔR_m とし、判別時に携帯端末、または携帯端末と基地局が測位する位置の誤差を ΔR_u とする。位置測位の誤差は、基地局のサービスエリアを示すカラーコードのみを取得した場合に数 km ~ 数百 m、GPS 座標を取得した場合に数十 m ~ 数 m である。登録・学習時または判別時に用いた測位法により、誤差を見込んでサービスエリアを拡張する。

カメラの画角を ϕ [rad] とする。看板 P_m のサイズを W_m [m] とすると、カメラの画面全体に看板が写る距離

$$R_m = R_m = \frac{W_m}{2 \tan(\phi/2)} \text{ [m]} \text{ である。看板のサービスエ}$$

リアの範囲は P_m を中心として半径 $R_m' = R_m + \Delta R_m + \Delta R_u$ [m] の円の内部になる。

ここで、二つの看板 $P_{m1}(X_{m1}, Y_{m1})$ と $P_{m2}(X_{m2}, Y_{m2})$

のサービスエリアの重なりについて調べる方法について述べる。 P_{m1} と P_{m2} の間の距離 $D_{m1,m2}$ は

$$D_{m1,m2} = \sqrt{(X_{m1} - X_{m2})^2 + (Y_{m1} - Y_{m2})^2} \text{ [m]} \text{ である。}$$

このとき P_{m1} と P_{m2} の看板のサービスエリアの範囲が R_{m1}' と R_{m2}' であるとすると、 $D_{m1,m2} \leq R_{m1}' + R_{m2}'$ であ

れば P_{m1} と P_{m2} のサービスエリアは重なる(又は接している)。この式を用いて、それぞれの看板のサービスエリアの重なりを調べ、重なりがある場合には、判別ベクトル $V_{sv}(m1,m2)$ とバイアス $b(m1,m2)$ を計算してサーバに登録する。

4. 対象判別実験

位置を利用することによって、対象の学習と判別をより素早く行うことができるかを検証する。

YRP野比駅周辺とYRPセンタ周辺の2つのエリアでそれぞれ10対象ずつ、合計20対象の対象判別実験を行った。YRP野比駅周辺では学習用と判別実験用に102枚の画像を撮影し、YRPセンタ周辺では学習用と判別実験用に99枚の画像を撮影した。いずれも屋外で撮影を行い、天候は晴れであった。対象を画面全体に写るように撮影し、モバイルカメラを対象の正面と対象の正面から左右に45度程度まで角度を変えて撮影した。モバイルカメラで撮影した対象の画像をノートパソコン(Pentium3, 600MHz)に取り込み、画像サイズ320×240[Pixel]のAVIファイルで一旦記録した。そのAVIファイルを同じPCで再度処理して、登録・学習・判別の実験を行った。実験結果として、図6に対象数に対する学習時間と判別時間の関係を、図7に位置の適用による学習サンプル数と正解率の変化を示す。

5. 実験結果

図6の実験結果により、判別にかかる時間については、対象数が20の場合の判別時間は1秒程度であったが、位置によって判別対象の候補を絞り込むことで判別回数を抑え、10対象にした場合では判別時間を0.25秒程度で行うことができた。

学習にかかる時間については、図7に示すように1対象あたりの登録・学習画像数が多い場合に、処理する画像数を減少させる効果がある。対象の学習に必要な画像の枚数は対象に依存し、より正確な判別を行うためにはある程度の枚数の画像を学習する必要がある。

位置を利用することによって判別時間の短縮と、看板データ変更の影響を局所にとどめることで、効率良く学習が行えることを確認した。

また、図7の実験結果により、位置を利用することによって他のエリアにある対象の候補を排除し、対象判別の正解率が向上することを確認した。

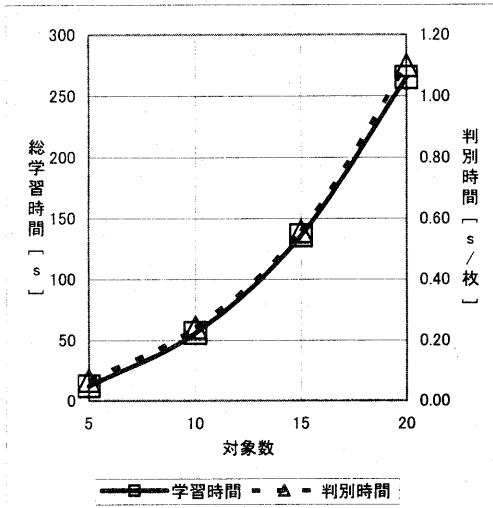


図 6 位置の適用による学習サンプル数と正解率の変化

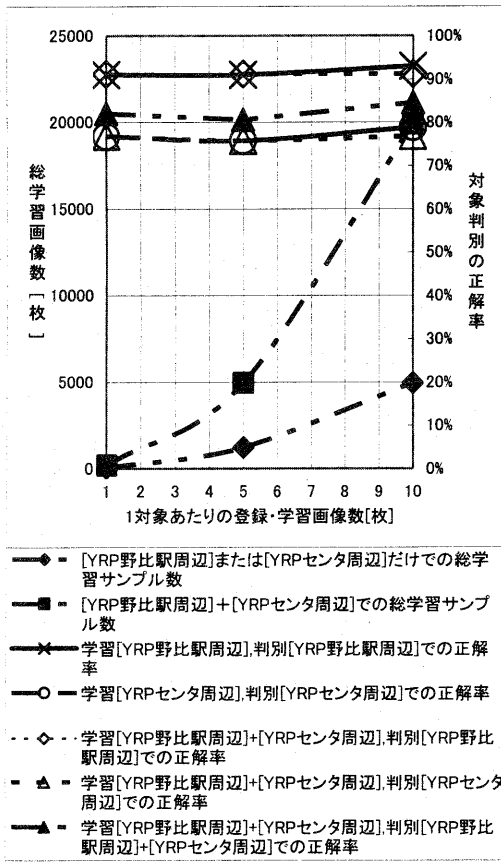


図 7 位置の適用による学習サンプル数と正解率の変化

6. おわりに

本稿では SVM による対判別を利用した看板・ポスター等の平面物体の判別手法を提案した。撮影画像を用いた対象判別技術と位置情報を利用する対象候補の絞り込み技術を提案し、屋外にある看板画像について対象判別実験を行った。

その結果、位置を利用することによって判別時間の短縮と、看板データの変更の影響を局所にとどめることで、効率良く学習が行えることを確認した。また、位置を利用することによって他のエリアにある対象の候補を排除し、対象判別の正解率が向上することを確認した。

情報提示対象画像の登録時に予め対象の関連情報やハイパーリンクなどを付加しておけば、対象の関連情報をタイムリーに提示することもできる。様々な条件下においてさらに正確で高速な対象判別を実現できれば、モバイル環境下でのマルチメディア情報検索に関する有用な要素技術となりうるであろう。

今後はモバイル環境の様々な条件下で有効な対象判別の手法と、対象の絞り込みの手法について研究をする予定である。

文献

- [1] 渡辺靖彦, 岡田至弘, 他, "シーン中のテキストを対象にした機械翻訳," 自然言語処理シンポジウム「実用的な自然言語処理に向けて」, 1997年
- [2] Corinna Cortes and Vladimir Vapnik, "Support-Vector Networks," Machine Learning, 20, pp.274-297, Mar. 1995.
- [3] Marti A. Hearst, "Support Vector Machines," IEEE Intelligent Systems, pp.18-28, July/August 1998.
- [4] Anuj Mohan, "Object Detection in Images by Components," A.I. Memo No.1664, C.B.C.L Paper No.178, Jun. 1999.
- [5] Edgar E. Osuna, Robert Freund and Federico Girosi, "Support Vector Machines: Training and Applications," A.I. Memo No.1602, C.B.C.L Paper No.144, Mar. 1997.
- [6] Steve Gunn, "Support Vector Machines for Classification and Regression," University of Southampton ISIS(Image Speech and Intelligent Systems group) Technical Report pp.1-52, Jul./Aug. 1998.
- [7] 村上祥司, "屋外広告物の現状と条例規定 -神奈川県の場合スタディー," 第31回日本都市計画学会学術研究論文集, pp.625-630, 1996年
- [8] 日本色彩学会編, "新編 色彩科学ハンドブック," 東京大学出版会, 1998年
- [9] 長尾 真, "デジタル画像処理," 近代科学社, 1978年