

モバイルカメラで情報を 検索する対象判別技術

ユーザが移動端末などの携帯端末に搭載されたカメラで撮影した街中にある広告・看板などの画像を用いて簡単に情報検索できる技術について検証した。

一般に、画像を用いた判別技術はさまざまな要因によって判別の正解率が大きく低下するという問題や、画像の比較的大きな情報量を取り扱うために処理時間が大きいという問題がある。しかし、今回の提案手法ではより正しくかつ高速に判別の処理を行えるように工夫を凝らしている。

やまぐち たかやす あおの ひろし ほんごう さだゆき
山口 高康 青野 博 本郷 節之

1. まえがき

移動端末の普及によって、モバイルインターネットを利用した情報検索が広く行われている。インターネットには、広告や宣伝をはじめとする実世界の対象物に関連する情報が多数存在するため、インターネット上に公開された情報を実世界の対象物に関連付けて検索することができればユーザの利便性は向上する。

一方、移動端末をはじめとする携帯端末の高機能化が進み、カメラを搭載した携帯端末が多数市場に出ている。これらの携帯端末に搭載されたカメラを、ここではモバイルカメラと呼ぶ。モバイルカメラは写真撮影のほか、2次元コード読み取りなど、新たな情報入力手段として利用されており、キー入力する方法と比べてユーザの利便性が高い。

以上のような背景から、今後モバイルカメラは周囲の情報を取得するセンサとして利用することが期待される。モバイルカメラで撮影した画像から対象物を判別し、その対象に関連するマルチメディア情報を検索することができれば、単に検索の手間を軽減するだけでなく、さまざまな検索アプリケーションへの応用によって利便性の高いマルチメディア情報検索が可能になると考えられる。モバイル環境において、ユーザが判別の対象とするものにはいろいろと考えられるが、以下のような理由から、看板、ポスターなどの平面物体を判別の対象とすることとした。

- ・ユーザがその場で受けられるサービスを知る手段として、看板が広く用いられている
- ・ユーザが知りたい情報に関連した看板を撮影することは直感的であり、サービス性が高いと考えられる

- ・看板で宣伝されている情報と、ネット上に存在する情報の相関が高いと期待できる

本稿では、従来技術の問題と提案手法の前提条件を述べたうえで、「画像の表現[1]」、「学習・判別手法[2]」、「位置情報の活用[3]」の3点に注力した新しい対象判別技術について述べる。また、この対象判別技術の適用を想定して構築したプロトタイプシステムの概要と今後の展開について述べる。

2. 対象判別技術

2.1 従来の判別手法での問題

既存の判別手法には2次元コードや電子透かし^{*1}などがあるが、これらの判別手法を用いて看板を判別しようとすると、以下のような問題がある。

(1) 2次元コード

遠方からの撮影に対応するため、大きな2次元コードを看板に貼り付ける必要があり、看板のデザインを損ねる場合がある。

(2) 電子透かし

街中ではさまざまな角度から看板が撮影されるが、判別対象の設置位置とユーザの撮影位置との間の角度に対して敏感である場合が多く、正面から撮影できない看板を判別することは難しい。

(3) RFID (Radio Frequency IDentification)

一般に、RFIDタグで読み取れる距離は制限される場合が多く、高所に設置してある看板などを判別することは難しい。

(4) 文字認識

特殊なフォントを使用している看板や、商品画像のみで文字がない看板も多く、想定したフォントの文字が含まれていない看板を判別することは難しい。

一方、画像処理を用いて画像を判別する手法であれば、既存の看板には手を加えずに、対象の画像をそのまま登録・判別できるというメリットがある。しかし、一般に画像処理は時間がかかり、屋外に設置されている対象の画像(以下、屋外画像と呼ぶ)は天候や時刻などにより刻一刻と変化するので、正確に判別することが困難である。

2.2 前提条件

今回の提案にあたっては、タウンガイドなどへ対象判別技術を適用することを想定している。このため、一般的な認識対象へ本技術の適用を目指しつつも、まずは判別対象

を看板・ポスターなどの平面物体に限定することとする。これまで、撮影画像に写っている対象を判別するという処理に対して、さまざまな手法が提案されているが、屋外では照明条件に大きな変化があるために、いずれの手法でも屋外の対象物を正確に判別できるとは限らない。そこで、照明条件などの環境変動に対する頑健性に重点をおいた汎用的な対象判別技術を確立することを目指す。

一口に店舗の看板・ポスターなどの平面物体といっても、物体自体の色、形状などの違いが存在する。そのため、ここで対象判別する物体は、必ずしも特定の色や形状を持った物体に限らず、多種多様な看板やポスターを判別対象とする。判別対象となる看板・ポスターなどの平面物体については、その種類、設置数、設置状況により判別の難しさが変わる。そこで、東京都渋谷区の渋谷センター街にある看板を調査した結果から、設置数の多かった「看板を照らす光源を持たない通常の看板」と「看板にライトを当てて看板を照らす看板」と「看板の内部に蛍光灯などの光源を設置して看板自体が発光する看板」とを対象とした。ただし、同じ看板でも昼と夜に照らされた場合とでは見かけの違いが大きいことから、昼と夜に撮影した看板の画像は別々の特性を持つデータのグループ(ここではクラスと呼ぶ)として扱うこととする。

提案した技術では、ユーザが一度だけ撮影する静止画を基に看板を判別することを想定している。時間の経過とともに光源や看板自体が変化する看板(電飾看板、スクロール式看板、テレビジョン看板など)は対象としない。看板の撮影にあたっては、物体の形状や撮影する角度によって見えない部分が発生する問題を避けるために、先に述べたように判別する対象を平面物体に限定し、見えない部分がないように撮影する。また、撮影画像に写る対象の位置や大きさを揃えるため、撮影の際には対象を画面中央におき、対象を画面からはみ出さない範囲で極力大きく(対象の縦方向の長さを画面の縦の長さに合わせる、もしくは対象の横方向の長さを画面の横の長さに合わせる)撮影することを条件とする。また、判別装置は判別の参考とする学習画像をあらかじめ登録しておく必要があり、店主が店舗の看板を撮影して登録を行うことを想定している。登録の手間を少なくするために学習画像の枚数は1つの看板当たり少数の画像で行う。

ユーザへの判別結果の提示に関しては、移動端末の画面には複数の判別候補の表示が可能であることから、10個の対象候補を順位付けて画面に表示する。

*1 電子透かし:画像や動画、音声などに、画質や音質にはほとんど影響を与えずに情報を埋め込む技術のこと。電子透かし検出ソフトにより埋め込まれた情報を取り出せる。

3. 画像の表現

3.1 問題の分析と要求条件の抽出

ユーザに判別結果を提示する際、特徴空間^{*2}における各クラスの分布の重なる度合いが大きくなると、必然的な誤り（バイズ・エラー）が発生するため、目的の判別対象が上位に表示されないことがある。より正しく判別するためには、特徴空間において各クラスの分布が分離するように画像の特徴を表現することが必要である。本稿では、分布の分離度を高めるための要求条件を以下のように設定する。

- (1) 前提としている撮影条件を守っても避けられない背景の影響を極力取り除いて表現し、環境変動に頑健なこと。
- (2) 看板特有の明るさや色、テキストチャ（模様）に着目し、看板の画像をよりよく表現すること。なお、明るさや色、テキストチャといった特徴は撮影を行う環境によって変動するが、今回取り扱う問題においては、看板の画像をよりよく表現することによって各クラスの分布の分離度を高め、看板画像を判別する性能を高める効果が期待できる。

3.2 表現方法

背景の影響を極力取り除くという要求条件を踏まえて、撮影条件を守って撮影した場合の表現方法を検討する。前提条件で述べたように、ユーザは対象を画面中央に極力大きく撮影することとしているので、判別対象とする看板は撮影画像の中央に存在する。そこで、撮影画像の中央領域と周辺領域に重みをつけて、撮影画像の中央領域を強調して表現する。また、看板特有の明るさや色をよりよく表現するという要求条件を踏まえて表現方法を検討する。画像を格子状に分割した際のそれぞれの領域における画像は $L^*a^*b^*$ ヒストグラムで表現する。

今回、新たに提案した 7×7 領域分割中央部強調 $L^*a^*b^*$ ヒストグラムと、既存の代表的な表現方法である $L^*a^*b^*$ ヒストグラム、ローパスグレー、周波数、モルフォロジを用いて画像を表現して評価を行う [1]。色を用いた簡易な表現手法の1つである色ヒストグラムは、照明条件の変化に弱いという欠点があるが、看板の色は心理的な効果を狙って配色されている場合も多く、看板を判別するのに大きな手がかりとなる色の特徴を表現する手法である。本稿では、心理的に同じ色違いに見える色同士の距離が均等な色空間である均等色空間、具体的には、均等色空間の代表的な色空間である $L^*a^*b^*$ 色空間を用いる。

色の変化を考慮しない方法としてローパスグレー、周波

数、モルフォロジを用いる方法がある。グレー画像の画素値を特徴とする表現は、ある程度高次元の特徴ベクトルとなるという欠点があるが、近年の計算機の高速化はめざましく、数千次元程度の特徴ベクトルであれば、取り扱うこともできるようになってきている。画像の周波数を特徴とする表現は、ある程度高次元の特徴ベクトルとなるという点と、高速フーリエ変換を用いて周波数成分を取り出すために、画像の縦横のサイズに制限があることが欠点である。また、看板画像にモルフォロジ処理を行い、画像中のオブジェクトの特徴を取り出して表現する方法がある。モルフォロジ処理にはさまざまな操作があり、看板画像からよりよい特徴を得るには、具体的な抽出の方法について、吟味する必要がある。

3.3 表現方法の評価基準

表現方法の評価は、2.2節の前提条件で述べたように、10個の候補を移動端末の画面に表示することを想定している。1位に画面に表示された場合は10点、2位に画面に表示された場合は9点となり、10位以内に入らずに画面に表示されなかった場合は0点とする。

3.4 表現方法の評価

渋谷センター街の渋谷駅から近い5つの交差点の周囲25m以内にある、520個の看板・ポスターなどの平面物体を対象とした。それぞれの表現方法について、 k -NN (k -Nearest Neighbor) の $k=1$ で判別した結果を図1に示す。なおNN法は、少ないサンプル数で高次元のデータを分離する場合に有効であるといわれている [2]。

昼・夕・夜の平均点の比較を行ったところ、提案方式が最も高い得点となり、判別の際の分離性能が良いことが分かった。今回提案した方式は、その次に性能が良かったローパスグレー方式の結果と比べて、約2.8点向上させること

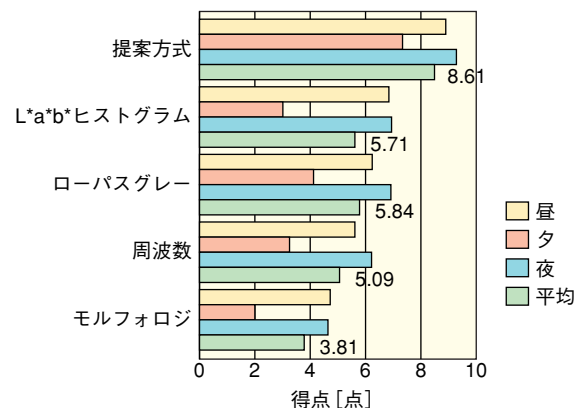


図1 表現方法の評価結果

*2 特徴空間：特徴（本稿では明るさや色、テキストチャなど）を数値化して座標軸を用いて構成した空間。

ができた。得点が8点以上となったことは、移動端末の画面の平均上位3位以内に、ユーザが検索する対象を表示できることを示している。

4. 学習・判別手法

4.1 良い学習・判別手法の観点と要求条件

2.2節で述べた前提条件のもとで、以下の3つの観点から、具体的な要求条件を設定する。

(1) 正解を上位に表示

ユーザが対象の候補の中から、目的の対象を簡単に選び出せるようにするために、画面の上位に正しい判別結果を表示する。最低でも平均で上位5位以内、可能であれば平均で上位3位以内に目的の対象を表示する。

(2) 学習時間が短い

店主が店の看板の登録・学習を行った際に、直ちに当該看板が正しく登録されたかを確認できるようにする。学習時間は最低でも1分以内、可能であれば1秒を切る。

(3) 判別時間が短い

街角で対象を撮影したユーザに、即座に対象に関する情報を提供する。判別時間は最低でも1秒以内、可能であれば100msを切る。

4.2 学習・判別手法

3章で提案した画像の表現を用いて、看板画像の判別を行う。提案した画像の表現に適した学習・判別アルゴリズムを明らかにするために、本稿では既存の代表的な学習・判別アルゴリズムである、FLD (Fisher's Linear Discriminant)、部分空間法、LVQ (Learning Vector Quantization)、SVM (Support Vector Machine)、 k -NN、NB (Naive Bayes) を用いて[2]評価を行う。FLDでは、クラス間変動のクラス内変動に対する比をフィッシャーの評価基準と呼び、これを最大にすることで2クラスの分離を行う。FLDは二値判別の手法であり、ある事象とその余事象を用いて判別する単一判別の方法と、クラスの組合せの数だけ判別を繰り返して判別する対判別の方法とがある。本稿では2クラスずつ次元圧縮を行って対判別を行う方法を試みた。部分空間法には複雑な分離面を構成できるなどの多くのバリエーションが知られているが、最も代表的な部分空間法であるCLAFIC (CLAss-Featuring Information Compression) 法を用いる。CLAFIC法はクラスごとにKL展開^{*3}によって部分空間を作成し、未知データが各クラスの部分空間にマッチングする度合いによって、未知データのクラスを判別する

手法である。LVQは入力データベクトルと結合重みベクトルにカテゴリを割り振り、それらのカテゴリを比較し、一致していれば入力と結合重みベクトルの距離を近づけ、一致しなければ遠ざける操作を繰り返し学習する手法である。SVMは2クラスのデータの間のマージンを最大にする判別面を構成する手法である。カーネルを導入することにより、高次元のデータを高速に分離することができる。本稿ではSVMを単一判別で適用し、判別面からの未知データまでの距離で各クラスの順位付けを行うこととした。 k -NNは、未知データと学習用のプロトタイプとの間の距離を基に多数決で判別を行う手法である。 k は多数決に用いる学習用のプロトタイプの数であり、 k の値は学習用のプロトタイプを判別するのに最適な k を推定する方法と、 k を固定する方法の両方がある。本稿では1クラス当りの学習のプロトタイプ数が少なく、おおむね k の値を固定にしても正しく判別できると考えられることから、 k の値を固定にする方法を試みた。NBは学習データを基に事後確率を最大化するように各クラスの分布のパラメータを学習しておき、未知データの分布と各クラスの分布を照らし合わせて判別を行う。多項分布のハイパーパラメータ (ξ) は、学習サンプルを最もよく分離するように、それぞれのクラスごとに調整する方法を用いる。

4.3 学習・判別手法の評価基準

各学習・判別手法を評価するにあたり、3.1節の要求条件に基づき、表1に示す評価基準を用いる。学習時間は渋谷センター街の1つのエリア内にある看板すべてを学習するのに要した時間とする。判別時間は未知の看板が撮影されている1枚の画像を判別するのに要する時間とする。以後、1枚の画像を判別する時間を、1対象当りの判別時間と呼ぶ。

4.4 学習・判別手法の評価

効果的に比較実験ができる1交差点の周囲 (半径20m) 内にあった145個の対象を選び出し、渋谷センター街のエリアでの性能について評価を行い、表2に示すような評価結果を得た。

k -NN (k 固定) は平均得点が8.8点 (上位2.2位に表示)、学習時間が無く、判別時間が94msである。また、NB (ξ 調整あり) は平均得点が8.8点 (上位2.2位に表示)、学習時

表1 学習・判別手法の評価基準

評価	145対象の得点	145対象の学習時間	1対象当りの判別時間
○	8点以上	1秒未満	100ms未満
△	8点未満、6点以上	1秒以上、1分未満	100ms以上、1秒未満
×	6点未満	1分以上	1秒以上

*3 KL展開: Karhunen-Loève展開。統計的な性質を利用して関数展開を行う手法。

表2 学習・判別手法の評価結果

判別手法	145対象の平均得点	145対象の学習時間	1対象当りの判別時間
FLD	○8.3点	×5.7時間	△454ms
部分空間法	○9.0点	△24秒	○6ms
LVQ	△7.6点	×1分53秒	○7ms
SVM	○9.1点	×3分44秒	△280ms
NB	○8.8点	○0.4秒	○46ms
k-NN	○8.8点	○学習なし	○94ms

間が0.4秒、判別時間が46msである。よって、上記2つの手法は、4.3節で述べた3位以内に目的の対象を表示し、1秒以内の学習時間と100ms以内の判別時間という要求条件を満足する。

5. 位置情報の利用とプロトタイプシステム構築

モバイル特有の情報である位置情報を活用し、検索候補をユーザの周囲にある少数の対象に限定することで、正確で高速な対象判別を実現する方法[3]を提案した。

対象判別のプロトタイプシステムはサーバと移動端末により構成する。サーバはデータベースと判別エンジンを持ち、移動端末から送られてくる位置に応じて、対象の判別・登録・修正と対象に関連する情報提供を行う。移動端末はユーザからの操作を受けて、サーバに対して位置情報を送り、対象の判別・登録・修正と対象に関連する情報を要求する。対象判別後は、i-modeを経由して対象に関するホームページにジャンプすることができ、文字、画像、音声、動画を駆使した情報提供を、スムーズに行うことができる。対象判別プロトタイプシステムの構成を図2に示す。

FOMA端末を使った対象判別の判別時間は、位置情報取得の機能を持たないため測位時間を考慮していないが、通信時間を含めても10s程度である。将来、移動端末の記憶容量が増えると、例えば、テーマパークに入場する際に学習データを移動端末に一括ダウンロードし、テーマパーク内での判別の際にはサーバに画像を送らずとも、移動端末の中だけで判別できる可能性もある。

6. あとがき

本稿では、屋外に設置されている看板を取り上げ、既存の判別手法での問題を解決するべく開発した新しい対象判別技術の核となる、画像の表現、学習・判別手法、位置情報の活用について述べた。さらに、これらの提案技術を用いて、iアプリと、インターネットに接続するサーバとを組み合わせたプロトタイプシステムを構築した。今後、さらなる判別性能の向上、測位精度の向上、測位時間の短縮、通信時間の

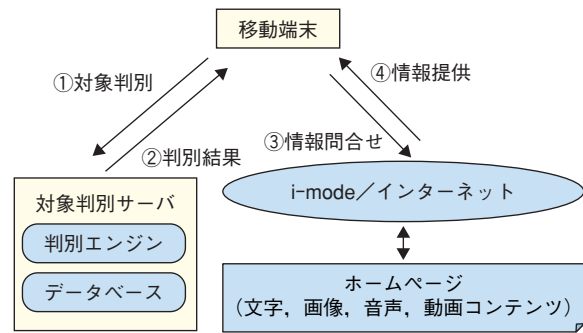


図2 対象判別プロトタイプシステムの構成

短縮、移動端末の高機能化を実現することによって、携帯端末を用いた街中での手軽な情報検索や新しい広告、ゲーム、イベントなどへと、ビジネス展開させていきたい。

文献

- [1] 山口 高康, 青野 博, 本郷 節之: “モバイルカメラで撮影した看板画像の特徴量に関する考察,” 信学技報PRMU2004-105, Vol. 104, No. 448, pp. 1-6, 2004.
- [2] 山口 高康, 青野 博, 本郷 節之: “モバイルカメラで撮影した看板画像の学習・判別手法に関する考察,” 信学技報PRMU2004-106, Vol. 104, No. 448, pp. 7-12, 2004.
- [3] 山口 高康, 高畑 実, 本郷 節之: “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,” 情処研報MBL02021017, Vol. 2002, No. 49, pp. 101-106, 2002.

用語一覧

CLAFIC : CLAss-Featuring Information Compression
 FLD : Fisher's Linear Discriminant
 k-NN : k-Nearest Neighbor
 LVQ : Learning Vector Quantization
 NB : Naive Bayes
 RFID : Radio Frequency IDentification
 SVM : Support Vector Machine