

画像内オブジェクトの自動抽出を使った画像検索システム

ExSight

- 写真(PhotoDisk)への適用 -

赤間 浩樹 紺谷 精一 三井 一能 串間 和彦

NTT情報通信研究所
〒239 横須賀市光の丘 1-1
akama@acm.org

あらまし 我々は画像内のオブジェクトの自動抽出技術を使った画像検索システムを研究開発している。我々のシステムでは、「この画像内のこのオブジェクトに似たオブジェクトを含む別の画像を探す」といった高度な検索が高速に実行できる。既に我々は、その第1ステップとして日本画に適用したシステムの詳細をデータ工学ワークショップ DEWS'97^{III}において報告した。本稿ではそれに補足する形で、その討論の中で会場から出た2つの質問に対する肯定的な回答を示す。

(質問1) オブジェクト自動抽出は写真に適用できるのか。

(質問2) 写真から抽出したオブジェクトに対して形状のマッチングは有効なのか。

キーワード 画像検索、写真検索、内容検索、オブジェクト認識、部分形状マッチング

ExSight

Image Retrieval System based on Automatic Object Extraction applied to Photographs (PhotoDisk)

Hiroki AKAMA Seiichi KONYA Kazuyoshi MII Kazuhiko KUSHIMA

NTT Information and Communication Systems Laboratories.
1-1 Hikarinooka Yokosuka-Shi Kanagawa 239 Japan
akama@acm.org

Abstract We are developing an image retrieval system based on an automatic object extraction method. This system can process queries such as "Search all images that include an object like this." We reported on its architecture and some techniques as applied to Japanese paintings at DEWS'97. In this paper, we answer the following two questions that were posed at this workshop.

(Q1) Can the system extract objects automatically from photographs?

(Q2) How effective is object retrieval by shape when the object is extracted from a photograph?

key words image retrieval, content based retrieval, object recognition, retrieve by partial shape

1 はじめに

デジタルカメラ等のデジタル機器やインターネットといった流通経路の急速な普及に伴って個人が発信/蓄積する画像の量が急増し、それらのDBとしての管理が求められている。また、出版社などの大量コンテンツ所有者もDBによる効率的な管理、および、そのインターネット上での流通に強い関心を抱いている。このような画像DBの高度化のニーズに対し、画像に対する内容検索を可能にした商用DBMSが登場している。しかし、これらの内容検索は、あくまで画像全体の特徴(色、色の配置、テキストチャ)などを利用するのに留まっている。

一方、我々は、画像内オブジェクトの認識技術を利用するアプローチをとり、画像DBに対する検索の高度化と高速化を狙っている。そして最終的には、画像内オブジェクトの単語化を行い、従来のテキスト検索技術の各種手法(辞書、FASTM、VSMなど)を適用し、より柔軟な画像検索を可能にすること目標としている。

既に我々は、その第1ステップとして日本画に適用した画像内オブジェクトを使った検索システムをデータ工学ワークショップ DEWS'97TMにおいて報告した。本稿では、その討論の中で出た以下の2つの質問に対する回答を示す。

- (1) オブジェクト自動抽出は写真に適用できるのか。
- (2) 写真から抽出したオブジェクトに対して形状のマッチングは有効なのか。

2 ExSight の概要

2.1 機能的特徴

ExSight は画像内オブジェクトを自動的に切り出し、その画像内オブジェクトの検索を可能にするシステムである。例えば以下のような検索ができる。

「このオブジェクト(物体)と似たオブジェクトを含む画像を検索せよ」

従来のテンプレートマッチングと呼ばれる手法が同等の機能を提供する。しかし、それは検索実行時に画像内をスキャンする方式であるため、画像数が多い場合に対処できない。また、従来の内容検索可能な画像DBシステムは画像の構造(色配置やテキストチャ配置など)に索引を付与することで大量画像に対する高速検索を可能とするが、画像内オブジェクトに対する上記の検索には対処できない。

2.2 構成

ExSight は、画像からオブジェクトを抽出する部分、画像やオブジェクトから特徴量を抽出部分、類似検索を行う部分、検索インタフェースの部分からなる。類似検索部では 3.1 節で述べる大規模性に対処するため、ベクトル空間に対する検索エンジンとして HyperMatchTM を利用し高速化を図っている。

2.3 技術的特徴

DEWS'97 で報告した ExSight の主要な技術は以下の3点であった。

(1) 自動的なオブジェクト抽出法

不完全なオブジェクト抽出を前提とし、大量のノイズを許容し、複数の認識候補を許すことで検索精度を向上させる。詳細については3章で述べる。

(2) 形状情報の特徴量の取得法

画像内オブジェクトを扱うシステムでは従来に比べ形状に対する類似検索の処理が重要になる。そこで、外接円を用いた形状情報の抽出手法でそれを実現した。詳細については4章で述べる。

(3) 類似尺度の計算法

特徴ベクトル間の類似度の計算処理において、非対称類似尺度を用いた手法を提案した。さらに、再現率-適合率グラフでユークリッド距離に比べ効果が高いことも確認した。

3 画像内オブジェクトの自動抽出

3.1 基本的な考え方

現状の認識技術を前提とし、それが不完全(認識結果にノイズと漏れを大量に含む)であるという認識のもとに、各種DB技術との組合せによって、それを実用的検索システムに応用する。検索精度の向上のための基本的な考え方は以下の通りである。

- ・もし、複数の認識候補があれば、全てをキーとして採用することで再現率を向上させる(ノイズが大量に含まれることを許す)。
- ・検索キーと投入データに同一の画像内オブジェクト認識手法を用い、抽出ノイズも含めたオブジェクト間の相対位置を利用することで、適合率を向上させる。

3.2 日本画版

我々が ExSight (日本画版) で採用した画像内オブジェクトの自動抽出の流れを図1に示す。

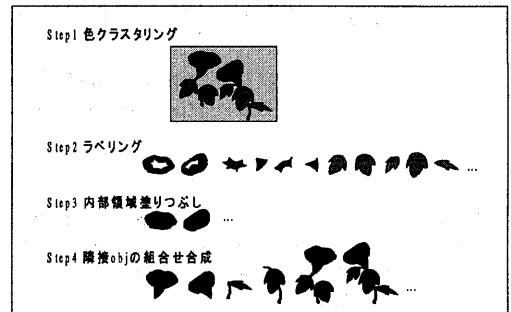


図1 日本画版での画像内オブジェクト抽出の流れ

第3のステップで行う塗りつぶしは、オブジェクトの内部領域を埋める操作で、オブジェクトを意味的

にまとめ、検索精度（再現率）を高めるといった効果がある。さらに第4ステップで隣接するオブジェクト同士のテンプレート形状を使った組合せによるオブジェクト合成を行い、分離しすぎたオブジェクトを再結合する。なお、これらのステップは意味的にまとまりのある画像と共に、大量のノイズ画像を生成する。

3.3 写真への適用の問題点

我々が日本画版 ExSight で適用したコンテンツは中村岳陵画伯の素描であった。それらは和紙上に描かれた後に写真撮影した画像であるため、認識対象としては背景が単色であるクリップアートなどよりは難しい。よって、前記の手法は、コンテンツ毎に人手でパラメータを調節せずに、日本画の素描より複雑度の低いクリップアートやアニメーション、背景がそれほど複雑でないカタログ等にはそのまま適用することができ、同程度の検索精度が期待できる。

しかし、我々は適用コンテンツとして、より市場規模の大きい一般写真を狙っている。そこで今回、実際の写真（著作権フリーの写真 PhotoDiskTM）に前述の手法の適用を試みた。この実験によって得られた主な問題点を以下に示す。

- (1) 色差がない場合に従来の手法は全く適用できない。例えば、図2のリング等、同一色のオブジェクトが密集して存在する場合には1つもオブジェクトが抽出できない。実は日本画においても同様な問題が存在した。つまり、朝顔の花の抽出は可能だったが、葉の一枚一枚まではとれていなかった。しかし、日本画の場合には、検索キーとして指定するオブジェクトは画像中で目立つ（色の）オブジェクトであることが多いため、この問題はそれほど大きくは感じない。しかし、写真の場合にはより本質的な問題となっている。
- (2) 共通のしきい値が存在しない。ここで、しきい値とは、画像内のオブジェクト間の区切り値を指す。例えば、屋外と室内、農村と海辺と町中などの写真内のオブジェクト群に共通のしきい値を設けることは難しい。さらに、1つの画像に限定しても、通常、1つのしきい値による領域抽出では過剰分割と分割不足が同時に発生する。
- (3) 複数のしきい値を用い、可能な限り抽出を行うと、逆に1枚の写真から冗長なオブジェクトが無数に抽出されてしまう。ここで、特に冗長であると目立つものは2種類存在した。1つは見た目にほとんど同じオブジェクト(3a)、もう1つは複数の領域の無駄な組み合わせで発生したオブジェクト(3b)。よって、3.1節の基本的な考え方は維持しつつもオブジェクト抽出精度の向上が欠かせない。

これらの課題への一般的な対処法として、写真毎や領域毎にしきい値を人手で微妙に変えながらオブジェクトを抽出する手法や、輪郭線に近い点をいくつか指定することでオブジェクトを精度良く抽出す

る手法が存在する。我々のシステムも商用化時にはそういった手法との統合も必要であると考えるが、我々の当面の目的からは外れるため今回の検討からは除外した。

3.4 写真版

今回我々は ExSight（写真版）のために、前述の課題に以下のアプローチで対処する手法（しきい値の全域調整によるオブジェクト抽出法）を考案した。

- ・色差も明度差もその他の情報も階調エッジ情報として統合して利用する。
- ・全範囲のしきい値からオブジェクトを抽出する。
- ・一定以内の類似のオブジェクト群からはその中の代表のみの抽出を行う。

以下にそのアルゴリズムを、図2に写真に適用した時の一連の流れを示す。

- 入力：階調エッジ画像（画素値は0～MAX_{edge}）
- (Step1)階調エッジ画像の0の値の画素に対してラベルづけを行いラベル画像を作成
 - (Step2)画素値 $e = 1 \sim \text{MAX}_{\text{edge}}$ に対して、以下のStep3を繰り返す。
 - (Step3)画像上で値 e をとる各連続領域 E_e に対して、以下のStep4、Step5を繰り返す。
 - (Step4)階調エッジ画像上の E_e に対応するラベル画像上の領域 L_e を求める。
 - (Step5)ラベル画像上で L_e に隣接する領域の個数 C を求め、その値によって処理を分岐する。
 - $C=0$ の場合：
 - 新たなラベル番号 L_{new} を生成し、ラベル画像上の領域 L_e のラベル番号を L_{new} とする。
 - $C=1$ の場合：
 - その隣接領域を $N1$ としたとき、ラベル画像上の領域 L_e と隣接領域 $N1$ を統合する。
 - $C \geq 2$ の場合：
 - その隣接領域を $N1 \sim Nc$ としたとき、各領域 $N1 \sim Nc$ のそれぞれをオブジェクトとして出力し、ラベル画像上の領域 L_e と隣接領域群 $N1 \sim Nc$ を統合する。

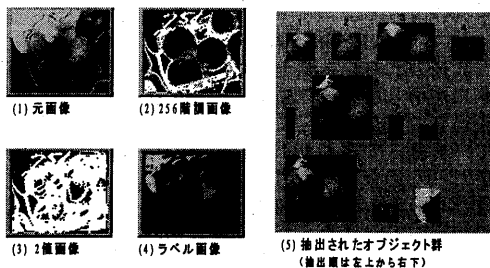


図2 アルゴリズムの適用例

元の画像から256階調エッジ画像を作成し、さらに、0の領域にラベリングを行い、上記 Step2以降の繰り返しを行う。従来手法では各々のリングは1つの領域に過剰統合されてしまっていたが、本アルゴリズムによって、1つ1つをオブジェクトとして抽出できている。さらに、256段階の全エッジ使用したにも関わらず(3a)タイプのノイズは少ない。なお、実際には日本画版のときと同様に100dot以下の領域は出力対象にしていない。

さらに、今回のオブジェクト抽出アルゴリズムは、抽出したオブジェクト間に図3に示すような関係のOE-tree (Object Extraction tree)を構成する。そこで、この tree 上で枝刈りを行うことで更に(3b)タイプのオブジェクトを削除し抽出精度の向上を行う。今回使用したルールを以下に示す。

- Rule1: ルート (=全体画像) は抽出対象から除く。
- Rule2: オブジェクトの外接矩形が同一の場合は抽出対象から除く。

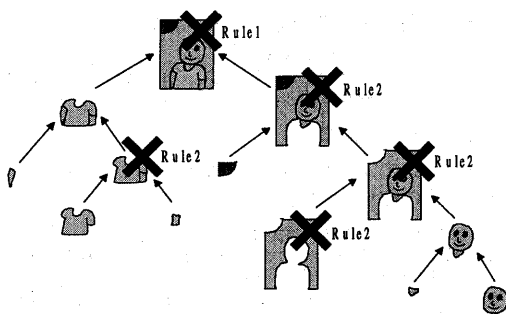


図3 OE-treeの例

今回のオブジェクト抽出手法は日本語版に比べ、技術的にはクラスタリングベースのオブジェクト領域抽出から、階調エッジベースのオブジェクト抽出に大きく様変わりした。一般的な画像処理技術であるセグメンテーション技術やクラスタリング技術との違いは、領域の重複する候補を積極的に認める点にある。

3.5 評価と考察 (その1)

上記アルゴリズムで抽出されるオブジェクトの精度は階調エッジの精度に大きく依存する。今回の実験ではHSIをベースとした単純なものを用いた。

- (1) 日本画 (中村岳陵画伯の素描) での比較
素描のうち、文献1に示したような朝顔の画像数点に適用した。その結果、朝顔の花の切り出しは従来通りできた。さらに、本アルゴリズムの効果として、文献1で述べたように従来では組合せ合成が必要だった場合でもその対処が不要になった。また、重なりのある朝顔の葉でも8割以上は1枚1枚の葉の抽出が可能になった。ただし同時に、ノイズとなるオブジェクトも2倍程度に増加した。

- (2) 色差の少ない写真¹⁴⁾での比較
従来の手法で全く対処できなかった画像 (1912番: 果物店の店頭)に並ぶリンゴの画像)について詳細に比較した。まず、従来の手法では全体として1つの大きな赤い領域となり、丸くとれるオブジェクトはない。一方、今回の手法を適用すれば、綺麗にとれたリンゴが17個 (図4)、検索に利用できる程度の品質であれば合わせて39個抽出できた。ただし、影と分離できなかったリンゴ、青い部分と赤い部分が分離したリンゴもいくつか存在した。

- (3) 一般写真¹⁴⁾での比較
一般写真としての代表的な写真を13点 (図4&5: 観点としては、同一オブジェクトの繰り返し、海辺、室内で暗い、明るい、構造物のみ、構造物と人、文房具、太陽光が強い、...)を選んだ。従来法と比較した時 (今回の手法の精度を1.0と正規化)の有意なオブジェクト数の増加を表1に示す。従来法でオブジェクトが抽出できなかった場合(N19121)を除いても、今回提案した手法は平均で2.1倍の有意オブジェクトが抽出できた。なお、この表には現れていないが、抽出されたオブジェクトの輪郭の質も今回の手法の方が全般に勝っていた。表2に今回の手法の抽出精度 (有意とノイズ) および OE-tree によるノイズの削減効果を示す。今回の OE-tree に対する2つルール適用により、平均12.4%のノイズ削減効果があった。ただし、それでもなお有意オブジェクトの平均11.6倍のノイズが存在している。

(矩形内は高品質に抽出できた物体)

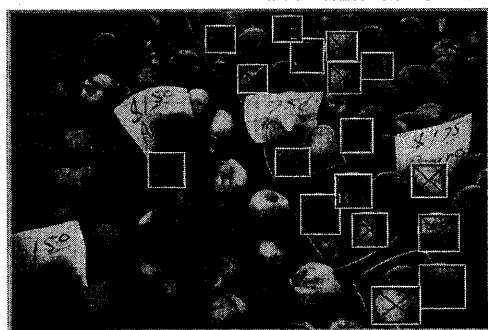


図4 色差の少ない写真への適用例



図5 実験に使用した一般写真の一覧

4 オブジェクト形状のマッチング

4.1 外接円を使った特徴量抽出

画像の中のオブジェクトを利用する我々のアプローチでは、形状による比較が従来の全体画像に対するアプローチに比べ重要になってくる。我々は、DEWS'97において、形状情報の新たな表現手法とし

表1 有意なオブジェクト数

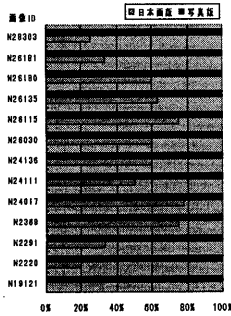


表2 オブジェクト抽出精度

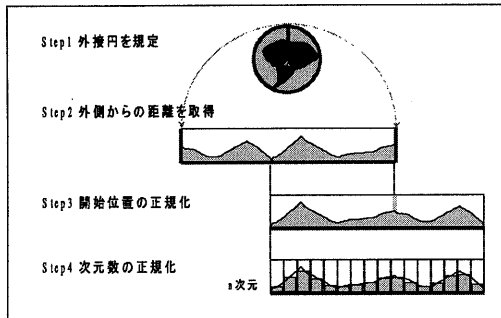
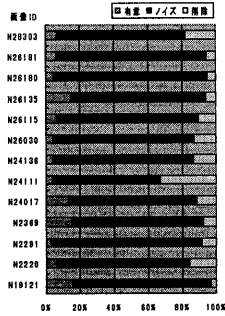


図6 形状情報のベクトル化の流れ

て外接円を使った形状情報のベクトル化手法を報告した(図6)。

この手法は、1つの基準で形状情報を表現することが可能で、色と形の両方の特徴を用いて検索する場合などに、その比率の調整が容易であるという特徴がある。また、文献1でも述べた通り、内部空間を持つ情報の取得も容易な拡張として実現できる。

4.2 写真への適用の問題点

形状のマッチングは写真版においても精度の向上に対して有効である。例えば、肌色の領域を指定した検索では、顔の他に、手、足はもちろん、砂地など、多くのノイズを含む。よって、検索精度を向上させるためには、色や大きさや位置だけでなく、その形状を指定することが有効な状況が多くある。

しかし、DEWS'97での質問でもでた通り、形状のマッチングに対してはより本質的な問題が存在する。

利用者側の問題：

- 必ずしも全体の形状を指定できるとは限らない。
- 例えば、人の横顔であることは指定したいが、髪型については自由度を与えた検索をしたい。

システム側の問題：

- 自動抽出したオブジェクトの精度に限界がある。特に、影や光沢の部分の抽出精度が悪い。
- もともとオブジェクト同士の重なりが多く、形状データとして不完全である場合が多い。

4.3 部分形状のマッチング

そこで、我々は部分形状でもマッチングできる手法を開発した。4.1節で述べた形状情報は外接円の円周上の各点からオブジェクトまでの距離を使用しているため、部分的な形状の情報を表現することも可能である。よって、全体形状のマッチングに以下の改良を加え部分形状マッチングを実現した。

- 入力エリアとして、重心を中心とした外接円(または最小外接円)を表示し、利用者にはその内部に部分的な形状を指定する。
- 部分的な形状の指定と共に、その形状の存在角度を指定できる。複数の部分形状が存在する場合にはそれらの間の角度も指定できる。

4.4 評価と考察(その2)

いくつかの検索キー画像(部分的スケッチ)による検索実験を行った。コンテンツとしては、正解の判定しやすさから、クリップアート集^[9]の900画像を用いた。なお、円の中心は重心とし、同一形状のものは除き、形の揺れの範囲は±20度以内とした。実験に用いたキーとその検索結果の上位5件を図7に、上位10件と20件内の正解率を表3に示す。平均正解率は共に74%と高く、方式としての有効性が確認できた。

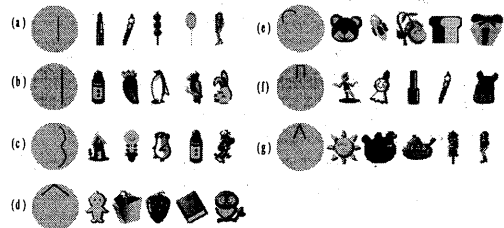
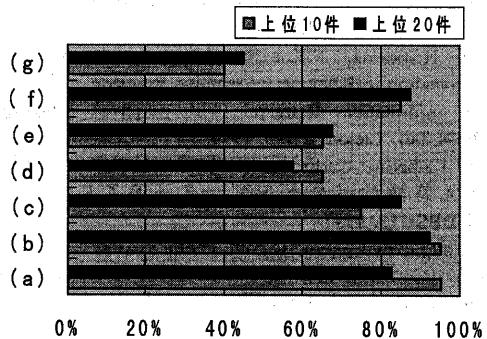


図7 部分形状による検索結果例

表3 部分形状による検索精度



一方、以下のような問題点も明らかとなった。

- ・入力領域の中心を重心とすべきか、最小外接円の中心とすべきかの判断が難しい。今後、統合または使い分けの基準の提示が必要である。
- ・精度の良い検索キー形状(a,b,f)と悪い検索キー形状(d,g)が存在する。今後、その原因の調査が必要である。
- ・部分形状での検索は、通常、他の情報(色など)と組合せて利用する。その組合せ方法をユーザインタフェースも含めて考えなければならない。なお、現時点の部分画像入力インタフェースは ExSight とは統合できていない。

5 まとめ

ExSight を写真に適用する場合に生じた問題、および、それを解決するために新たに開発した技術について報告した。オブジェクトの抽出においては、しきい値の全域調整によるオブジェクト抽出アルゴリズムと OE-tree を使ったノイズ削減による精度向上を行い、不完全形状オブジェクトへの対処には部分形状でのマッチング手法を適用する。より大量データセットでの評価については今後報告したい。

謝辞

本画像 DB システムを作成するにあたり、画像処理技術に関する多くのアドバイスをいただいた NTT ヒューマンインタフェース研究所 メディア応用システム研究部の倉掛正治氏、新井啓之氏、および画像処理の基本処理ライブラリの提供をしていただいた桑野秀豪氏に心から感謝します。

データ工学ワークショップ DEWS'97 の発表に対し、ご討論戴いた皆様に感謝します。

参考文献

- [1] 赤間, 三井 "画像内オブジェクトの自動抽出を使った画像検索システム ExSight", 第8回データ工学ワークショップ(DEWS'97), pp107-112, 1997.
- [2] H.Akama, F.Konishi "Application of Frame Association to Japanese Information Retrieval", Workshop on Information Retrieval with Oriental Languages (IROL'96), Taejon, Korea, 1996.
- [3] Katherine Curtis, 谷口 "画像内容検索のための多次元特徴量ベクトルデータ高速アクセス法", 97-DBS-113, 1997.
- [4] <http://www.photodisc.com>, PhotoDisk, Inc.
- [5] インプレス編集部, "Internet ホームページ用素材集 HOT&プライベート編", インプレス, 1997.