

縁に注目した画像のタイプ分けを利用する 検索画像へのテキスト付加 Image Annotation According to the Types of Images Derived by Regarding Color-Histograms of the Borders

中島 誠† 高嶋 修† 伊藤 哲郎†
Makoto Nakashima, Osamu Takashima, Tetsuro Ito

1. はじめに

デジタルカメラやスキャナ等デジタル機器の普及により、写真画像を含んだ大規模な画像データベースが構築されるにつれ、その中から所望の画像をうまく取り出せる検索手法への要求が高まっている。多くのユーザは、画像の視覚的な内容より、例えば“春の雰囲気を伝える画像が欲しい”といった、抽象的な要求に概念的に合致する画像を求め[3]、文献検索と同様思い浮かんだキーワードによる質問を使った検索手法を好むとされる[5]。このような手法を実現するには、その主題は何かを説明するテキストを画像に付加しておくことが不可欠となる。

これまでに、画像処理技術を用いて画像中のオブジェクトを特定し、テキストが付加された画像との比較において、自動的にテキスト付加を行う方法が提案されているが[2][10]、オブジェクト特定に十分な精度が得られていない。商用の画像検索エンジンで行われているように[1][4]、人手によるテキスト付加は、画像が増大するにつれそのコストが大きくなる。これに対し、データベース中の画像に複数のユーザが共同でテキストを付加できる Web ページを用意した試みがある[6]。ここでは、テキスト付加の対象を多くのユーザが WWW を介して検索した画像に広げ、これにテキストを付加しやすいよう支援することで、精度とコストの問題を軽減する方策を提案する。

ユーザにとって、漠然と画像の主題の説明を求められるのではなく、どのようなテキストを付加したらよいかの示唆がなされれば、その負担は少なくなる。ここでは、扱う主題の大まかな違いに合わせて画像をタイプ分けし、各タイプに共通のテキストの形式や既にタイプ分けされた他の画像の例を提示することで、テキスト付加をしやすくする。タイプ分けには、画像はその作成者により、写しだされている範囲より広い領域から必要な部分を切り出すことで得られたことを考慮にいれる[8]。すなわち、切り出し方が、画像の主題と深い関連があるとして、画像の縁の色分布に注目した方法を提案する。

以下では、ここで用いる画像タイプとこれを用いたテキスト付加の考え方を述べる。また、画像として四角形の写真画像を対象とし、縁に注目したタイプ分けの方法と、その有効性を示す。

2. テキストを用いる画像検索

画像へのテキスト付加も含めた画像検索の全体像を図 1 に示す。ユーザの質問は、まず画像のテキストおよび画像

が存在する WWW 上での位置を蓄えたデータベースへのポータルシステムに渡される。質問に合致した画像が見つからない場合には、ポータルシステムから商用の画像検索エンジンを通じて検索を行う。検索結果の中からユーザに自身の要求にあった画像を選んでもらい、主題を説明するテキストを付加してもらい、そして、画像の位置と共に格納しておく。これを、多くのユーザに行ってもらうことで、様々なユーザの見方で大量の画像にテキストが付加される仕組みとなる。

画像に付加されたテキスト中の語句はユーザによって様々になる。これには、分類体系やシソーラスなどの概念辞書を介して、概念的に関連の深い語句をまとめることでテキストを整理し、画像の概念的表現として格納する[11]。質問と画像の概念的表現とのマッチングに概念辞書を利用することで[12]、ユーザが思い浮かぶキーワードで質問をしても望む画像が取り出し易くなる。

ユーザがテキスト付加をし易いかどうか、どれだけ多くの画像へのテキストが得られるかに影響する。以下では、画像のタイプ分けをもとにテキスト付加を支援する方法について述べる。

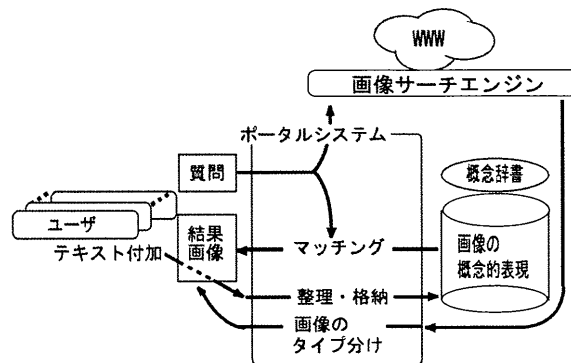


図1 画像検索の全体像

3. 画像タイプとテキスト付加

3.1 画像タイプ

画像のタイプ分けの目的は、タイプごとに、ユーザに付加すべきテキストについての示唆を与え、付加のコストを軽減することにある。従って、画像に付加されるテキストの形式が明白なタイプが望ましい。

ここでは、写真のタイプ分けを参考に、動植物や静物などのオブジェクト、風景、およびこれら以外で場の雰囲気を伝える3つの画像タイプを基本とする[8]。加えて、風景と雰囲気について、その中にオブジェクトを含む場合（それぞれ風景+オブジェクト、雰囲気+オブジェクトと書く）

†大分大学知能情報システム工学科
Dept. of Computer Science and Intelligent Systems,
Oita University

表3 画像タイプ別語句の概念

	物	場所	時	事象	物と場所	物と時	物と事象	場所と時	場所と事象	時と事象
オブジェクト	1.00	0.13	0.07	0.50	0.13	0.07	0.50	0.00	0.10	0.03
風景	0.13	0.88	0.33	0.17	0.04	0.04	0.00	0.25	0.13	0.04
雰囲気	0.73	0.43	0.27	0.60	0.30	0.13	0.47	0.10	0.20	0.13
風景+オブジェクト	1.00	0.74	0.17	0.48	0.74	0.17	0.48	0.04	0.39	0.09
雰囲気+オブジェクト	0.93	0.51	0.33	0.71	0.49	0.31	0.67	0.09	0.31	0.24

表1 タイプ別画像

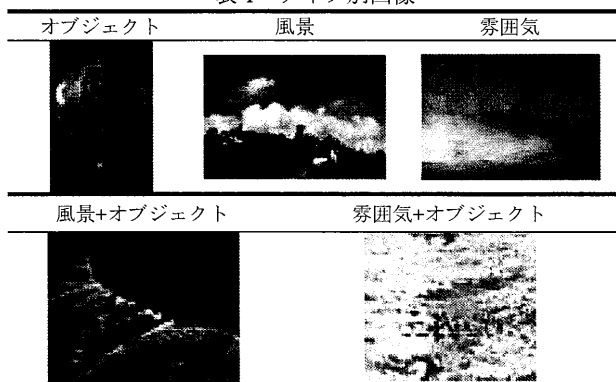


表2 タイプ別テキスト

オブジェクト (30)	象, 白鷺, マーチソン滝, 二葉葵の葉, 説教する修道士
風景(24)	エチオピア高原, マサイマラ動物保護区の夕暮れ, 色づく豆畑, 瀬戸内の段々畑
雰囲気(30)	コスモス満開, 鮎つり, 光るのりひび, 紅葉一色の庭, 金曜日正午過ぎの合同礼拝
風景+オブジェクト(23)	タナ湖の葦船, 砂漠に流れ入った青ナイル, 青い海と釣り舟,
雰囲気+オブジェクト(43)	朝霧に煙るカフラーのピラミッド, 池面を流れる花筏, 仲秋名月と芒

を考慮し、5つのタイプを用いる。表1に5つのタイプの画像例を示す。

設定した5つの画像タイプについて、画像に付加されるテキストの形式が明白かどうかを、作成者によりタイトルが画像に付加されている写真集[7][9][13]に掲載されている150画像を対象にして調べてみた。これらの画像を手でタイプ分けした後、各画像のタイトルをテキストとして抽出した。タイプ別の画像数とタイトルの例を表2に示す。

抽出したテキストの中の助詞、助動詞、接続詞以外の語句を、それらが表す概念を EDR 概念体系辞書を用いて調べ、動植物、静物、事柄などの「物」、高原、湖などの「場所」、正午や仲秋などの「時」、流れる、満開などの「事象」の4つに分けた。そして、タイプごとに各概念の語句が用いられている頻度(タイプごとの総画像数で正規化した)を調べた。結果を表3に示す。

オブジェクト、雰囲気、風景+オブジェクト、雰囲気+オブジェクトの4タイプの画像には、高い頻度で「物」に関する語句が用いられていた。風景では「場所」に関する語句のみの頻度が高く、風景+オブジェクトになると、「物」と「場所」両方の頻度が高い。雰囲気と雰囲気+オブジェクトでは、「事象」の語句が多く、後者では、

「物」と「時」、「物」と「事象」、あるいは、「物」と「事象」がそれぞれ同時に現れる頻度が他のタイプに比べて高かった。まとめると、オブジェクトには「物」、風景には「場所」、雰囲気には、「物」あるいは「事象」、風景+オブジェクトには、「物」と「場所」の両方、そして、雰囲気+オブジェクトには、「物」に加えて「時」あるいは「事象」の語句、または「時」と「事象」の両方の語句を付加しておけばよいことがわかる。

3.2 テキスト付加

上で調べたように、タイプごとのテキストの形式は明らかで、それぞれをユーザに示すことが、付加するテキストについての示唆となるといえる。タイプごとの他の画像に付加されたテキストを提示すると、形式に合う語句の選定も容易になる。

画像へのテキスト付加の例をみてみよう。表1に示した画像と上で調べたテキストの形式を示し、複数の大学生にそれぞれのタイプに合わせてテキストを付加してもらった。例として、オブジェクトと雰囲気+オブジェクトの結果を示す。

オブジェクト：バルーン, 気球, 熱気球
 雰囲気+オブジェクト：花見をする家族,
 春の桜を楽しむ人たち,
 花見を楽しむ人たち

オブジェクトの画像には、表記の違う「物」に関する語句が用いられている。雰囲気+オブジェクトには、「物」に関する家族、桜、人たち、花見と「事象」に関する花見する、楽しむといった語句の組み合わせになった。2番目のテキストには、「時」に関する語句、春、も付けられた。これらを概念辞書を介して整理することで、画像の概念的表現が得られる。

4. 縁に注目した画像のタイプ分け

4.1 画像のタイプ分け

図1に示したように、多くのユーザに検索画像へのテキスト付加を行ってもらうことで、そのコストの軽減が図れる。しかしながら、画像のタイプ分けを手で行なうのではコスト軽減の利点が薄れてしまう。これへの対処として、自動的に画像のタイプ分けをする方法を述べる。

タイプ分けの方法は大きく2段階からなる。まず、画像から抽出した縁の色分布に注目し、オブジェクト、風景、雰囲気の3つの基本タイプの何れかを定める。その後、風景および雰囲気は、縁と異なる色分布の領域が内部に存在すれば、オブジェクトを含む画像として、それぞれ、風景+オブジェクト、雰囲気+オブジェクトとする。

基本タイプが決まれば、画像内部に向けて縁と同じ色分布の領域を徐々に辿っていけば、縁とは違う領域を見つけることができ、オブジェクトの存在がわかる。問題は基本タイプの決め方である。画像の構成は、作成者によって様々で、色の分布の仕方も異なる。ここでは、風景の画像では、縁の色は大きく2種あるいは3種になるとする。オブジェクトおよび雰囲気画像では、両者とも大まかな縁の色は一樣であるが、雰囲気画像では、細かな色の違いが存在する。以上の考えをもとにした、基本3タイプへの画像の分け方を以下に示す。

一定の幅を持たせた画像の縁を抽出して、四角形の画像の4辺をそれぞれ N 分割し、分割された領域 R_1, \dots, R_{4N} を色のヒストグラムを基にした類似度に従ってクラスタリングする。縁の分割例を図2に示す(図中では、各辺を5分割している)。クラスタリングには、各領域 R_i を節点とし、2つの領域 R_i と R_j の類似度を節点間の線重みとしたグラフ上での最大生成木を求める。重みが閾値 τ 未満の線を除去した場合に連結されている節点に対応する領域が1つのクラスタをなすとする。

領域中の各ピクセルの色は、RGB値をもとに、日本色彩研究所の系統色体系(最も細かな小分類で230色)のうち基本分類の16色に置き換えて減色しておく。得られるクラスタ数 C_{num} をもとに以下の条件にしたがって対応する基本タイプを決める。

(i) $C_{num} = 1$ の場合

オブジェクトあるいは雰囲気画像とする。両者の区別には、各領域 R_i のピクセルの色を系統色体系の中でより細かな小分類の230色に減色させ、分割数 N と閾値 τ を増加させて、クラスタリングを行う。新たに得られたクラスタ数が1ならばオブジェクト、それ以外は、雰囲気画像とする。

(ii) $2 \leq C_{num} \leq 3$ の場合:

風景の画像とする。

(iii) $C_{num} \geq 4$ の場合:

雰囲気画像とする。

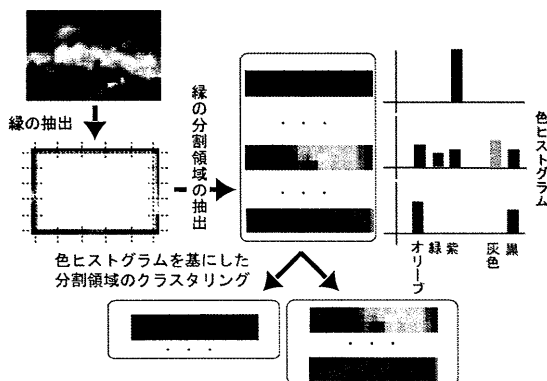


図2 画像の縁の分割とクラスタリング

4.2 タイプ分けの評価

Web ページデザイン等のための市販の素材集にある70枚の写真画像を対象に、基本3タイプへのタイプ分け方法の有効性を調べた。各画像は、その主題ごとに名前がつけられたフォルダにまとめられている。ここでは、オブジェクトの画像として、“植物”と“オブジェ”，風景の画像

として“自然風景”，雰囲気として，“日本の雅”と名づけられたフォルダ内の画像をそれぞれ30枚(計90枚)ランダムに抽出して用いた(画像サイズは 640×427 あるいは 427×640 ピクセルであった)。

各画像の上下と左右の縁の幅は、それぞれ画像の縦、横のサイズの $1/50$ とした。クラスタリングの閾値 $\tau=0.65$ とし、分割数 N を3から11に種々変えたときの、タイプ分けの精度を表4に示す。 N の各値についてのオブジェクトと雰囲気の精度は、クラスタ数 C_{num} が1の場合に、 $\tau=0.75$ として減色する色数と N の値を増やし($N=11$ のときは $N=13$ とした)てオブジェクトか雰囲気かを決定した時の最良の結果を載せてある。

表4 タイプ分けの精度

分割数 N	オブジェクト	風景	雰囲気	平均
3	0.73	0.83	0.43	0.67
5	0.80	0.83	0.70	0.78
7	0.76	0.87	0.50	0.71
9	0.76	0.73	0.63	0.71
11	0.76	0.70	0.67	0.71

平均でみると、 $N=5$ の場合が最もよく、このとき、オブジェクトか雰囲気かの決定は $N=11$ に増やしたときが最も良かった。オブジェクトと風景では、 N が小さすぎても大きすぎても結果がよくなく、特に風景に関しては、 N が大きくなる程その傾向が強かった。オブジェクトは、 N が大きくなると精度に変化がない。縁の色分布が一樣であったためである。また、雰囲気については、最初の分割数によって結果が上下した。縁の色分布の仕方が画像ごとに異なっていたためである。

6. おわりに

今後は、画像検索エンジンにより検索した画像を受け付けて、タイプ分け結果とテキストの形式やタイプ分けされた他の画像のテキストを提示するWebサーバーを構築し、ブラウザ上でユーザにテキストの付加を行ってもらい、そして、キーワードによる質問を受け付けるポータルシステムの運用を通じてここでの方法の有効性を確かめる。

参考文献

- [1] <http://www.altavista.com/>
- [2] K. Barnard and D.A. Forsyth, “Learning the semantics of words and pictures,” Proc. IEEE ICCV’01, pp.408-415, Vancouver, Canada, 2001.
- [3] H. Feng and T.-S. Chua, “A bootstrapping approach to annotating large image collection,” Proc. ACM SIGMM MIR’03, pp.55-62, Berkeley CA, USA, 2003.
- [4] <http://www.google.co.jp/>
- [5] J. Jeon, V. Lavrenko and R. Mannmatha, “Automatic image annotation and retrieval using cross-media relevance models,” Proc. ACM SIGIR’03, pp.119-126, Toronto, Canada, 2003.
- [6] 小泉大地, 柘植 寛, 獅々堀正幹, 北 研二, “テキストと画像のクロスメディア情報検索に向けた画像キーワード登録システムの開発,” 情処データベースシステム研究会報告, pp. 105-112, 2002.
- [7] 前田真三, 前田真三写真美術館3ふるさと, 講談社, 東京, 1999.
- [8] 三木淳 他(監修), 写真大事典, 講談社, 東京, 1984.

- [9] 水野克比古, “こころの京都,” 主婦の友社, 東京, 1999.
- [10] Y. Mori, H. Takahashi and R. Oka, “Image-to-word transformation based on dividing and vector quantizing images with words,” Proc. MISRM’99, Orland, USA, 1999.
- [11] 中島 誠, 伊藤哲郎, “容易な情報アクセスのための文献クラスの概念的表現,” 情報科学技術フォーラム情報技術レターズ, vol.2, pp.107-108, 2003.
- [12] M. Nakashima, K. Sato, Y. Qu, and T. Ito, “Browsing-based conceptual information retrieval incorporating dictionary term relations, keyword association, and a user’s interest,” JASIST, vol.54, no.1, pp. 16-28,2003.
- [13] 野町和嘉, “The NILE,” 情報センター出版局, 東京, 1989.