

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

小泉 大地 柘植 覚 獅々堀 正幹 北 研二

徳島大学 工学部

従来の情報検索技術は、文書から文書といった同メディア間の検索技術が主流であった。しかし、マルチメディアデータの急速な増加により、画像から音声、音声から映像といった異なるメディア間での情報検索(クロスメディア情報検索)の技術が必要とされてきている。我々は、最初のステップとして、文書から得られる言語情報と、画像から得られる可視特徴量との統合を図り、文書と画像間のクロスメディア情報検索技術について研究を進めている。本研究を行う上で、言語情報と可視特徴量を統合するためのメタデータとして、キーワード付けされた画像データベースを作成することが必要である。本稿では、既存の画像データベースに対して容易に言語情報を登録することができる画像キーワード登録システム「ぴぼっとくん」を紹介する。本システムの特徴として、以下の点があげられる。(1) ネットワークを通じて複数ユーザが共同して作業可能である。(2) 通常のキーワード以外に感性語、自由記述文書を登録可能である。(3) 複数キーワードによる高速な画像検索が可能である。(4) 類似画像検索技術を用いることにより、各画像に対する推薦キーワードの提示が可能である。

Development of Image Keyword Registration System for Cross-media Information Retrieval between Text and Images

Daichi Koizumi Satoru Tsuge Masami Shishibori Kenji Kita

Faculty of Engineering, Tokushima University

Multimedia information retrieval has been addressed as one of the most promising solutions for accessing rapidly growing multimedia data. Traditional multimedia information retrieval systems have dealt with the same kind of media data. With diversification of multimedia data, however, it becomes necessary to develop multimedia information retrieval technology among different media data. In order to achieve the purpose above, we are studying cross-media information retrieval technology between text and images. Based on linguistic information provided by associated text and visual information provided by image features, an image retrieval system capable of efficiently searching appropriate image data from huge image database will be developed. In this paper, we introduce keywords tagging support system "PIVOT" as well as an image retrieval system based on color distribution and tagged keywords. This system has the following characteristics: (1) several users can work jointly through the network, (2) sensibility words and free descriptions can be attached in addition to system pre-defined keyword, (3) we can quickly and efficiently retrieve target images from the given keywords, (4) based on the content-based image retrieval technique, the system is able to propose keywords suitable to a given image.

1 はじめに

通信のブロードバンド化にともない、テキスト(文書)、画像、映像、音声、音楽等のデータがすべてデジタル化され、オフィスや大学あるいは家庭においても、膨大なマルチメディア・コンテンツを蓄積する時代が近づいてきている。これにともない、膨大なマルチメディア・コンテンツの中から、ユーザの多様化する情報ニーズに適合するような情報を検索する手段が求められている。このためには、テキスト、画像、音声などの各メディアごとの検索技術の高度化に加え、質的に異なるコンテンツを統合的な観点から扱うことのできる情報検索技術が必要である。

我々は、意味的な表現と感性的な表現の2つをマルチメディア・コンテンツに持たせることにより、質的に異なるコンテンツを相互に検索することのできる知的な情報検索技術について研究を進めている。意味的な表現を用いることにより、たとえば画像コンテンツ中の対象物をキーワードや自然言語文により検索することが可能となり、感性的な表現を用いることにより「明るい絵」や「綺麗な絵」などの検索要求に応えることが可能であると考えられる。また、さまざまなメディア・コンテンツに対し同一の統一的な表現形式を用いることにより、与えられた画像や映像に適した音楽を検索する等のクロスメディア情報検索(cross-media information retrieval)も実現可能となる。現時点では、テキストと画像を対象にした研究を進めているが、将来的には、映像や音声、音楽等のメディア・コンテンツも取り入れる予定である。

本研究を行う上では、画像に意味的あるいは感性的な情報が付与された大規模なメタデータベースが必要である。しかし、膨大な画像に対し一つ一つ人手でキーワードを付与するのは多大な労力・時間を要する。そこで我々は、ネットワークを通じて複数のユーザが共同で作業することが可能な画像キーワード登録システム「ぴぼっとくん」を開発した。本システムの特徴としては以下の点をあげることができる。(1) ネットワークを介し、複数のユーザが共同で作業することが可能である。(2) システム側であらかじめ用意されているキーワードだけでなく、感性語、自由キーワードといった情報を個別に登録管理することが可能である。(3) 複数のキーワードを用いて高速な画像検索が可能である。(4) 類似画像検索技術を用いることにより、各画像に対する推薦キーワードの提示が可能である。「ぴぼっとくん」により作成されたキーワード付き画像データをもとにし、キーワードから得られる意味的特徴量と画像データから得られる画像特

徴量との確率的相関関係を求めることにより、画像の類似検索とテキスト(キーワード)による画像検索を複合的に組み合わせたクロスメディア情報検索システムを実現することが可能になる。

2 クロスメディア情報検索

クロスメディア情報検索とは、異なるメディア間のコンテンツを検索する手法である。キーワードやテキストによる画像検索の例としては、Google [1]、Lycos Multimedia Search [2]、AltaVista Image Search [3] などがある。これらのシステムはページ中の画像近傍の単語を使って画像の索引化を行っている。また内容に基づく画像検索技術 [4] [5] [6] も従来より研究されている。クロスメディア情報検索の例としては、フィードバックにより半自動的に画像にキーワードを付与することが可能なシステム [7] やフィードバックを用いたコンテンツに基づく画像検索システム [8] などがある。

クロスメディア情報検索を実現する1つの方法として、異なるメディア・コンテンツを統一的な表現形式に変換する方法がある。従来、キーワードなどから文書を検索する場合には、文書内の各単語の意味的な情報を特徴量として使い、画像から類似画像を検索する場合には、画像から得られる可視情報を特徴量として用いている。これらのメディア間での表現形式の相違を吸収するためには、キーワードやテキストから得られる意味情報と画像から得られる可視情報の確率的相関関係を利用する方法が考えられる。

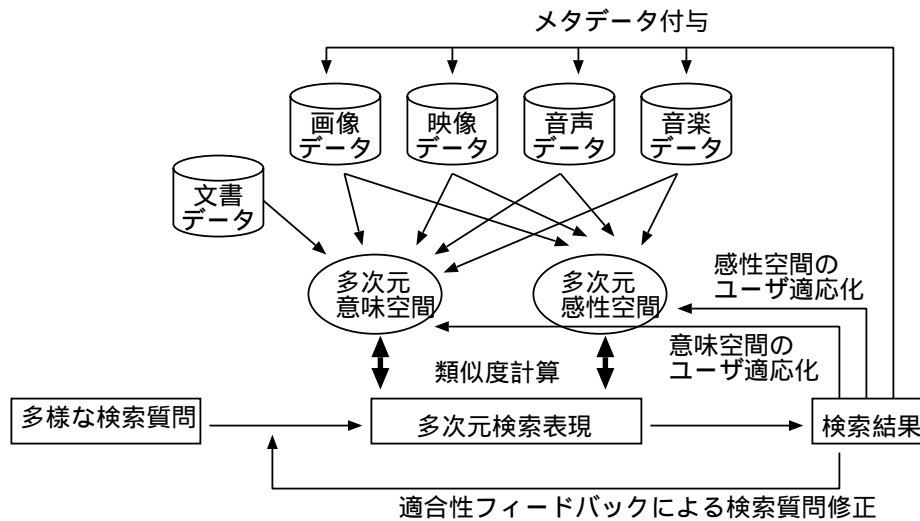
我々が研究開発中のクロスメディア情報検索システムの概念図を図1に示す。まず文書、画像、映像、音楽等のさまざまなメディアに対し、意味情報および感性情報が付与されたメタデータを作成し、そのメタデータから意味空間および感性空間を構築する。以上のように各メディアのコンテンツを同一の空間上に配置することにより、異種メディア間の検索を統一的な枠組みで扱うことが可能となる。

3 画像キーワード登録システム

「ぴぼっとくん」

3.1 「ぴぼっとくん」の概要

テキストと画像情報を対象にし、クロスメディア情報検索を実現するためには、人間が持つ知識を反映したメ



- ・キーワード・自然言語による検索
- ・コンテンツの類似検索
- ・クロスメディア検索
- ・.....

図 1: クロスメディア情報検索システム

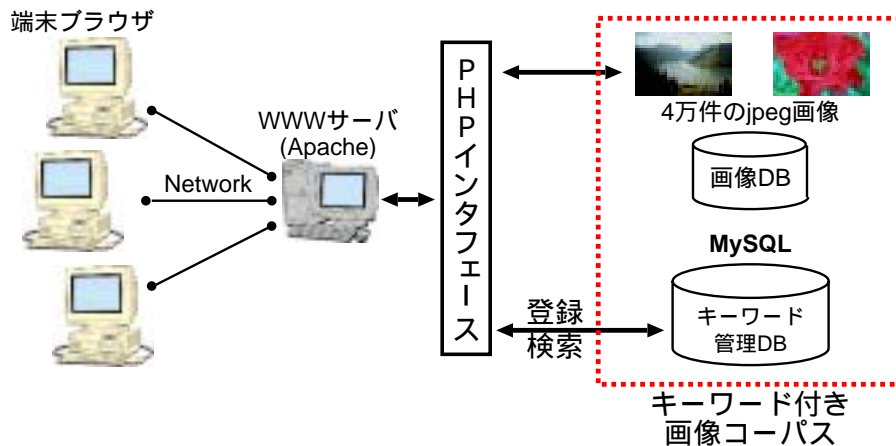


図 2: 画像キーワード登録システム「ぴぼっとくん」のシステム構成

タデータとして、各画像から連想されるキーワードや感性語が付与された大量の画像データが必要になる。しかし、既存の画像データに人手で一つ一つキーワードを付与するには、多大な労力を要し、また、少人数で作成した場合には、多大な日数が掛かってしまう。そこで、我々は、労力の軽減を目的とし、使いやすいインタフェース機能とネットワークを介して複数のユーザがデータベースに高速にアクセスできる機能を有した画像キーワード登録システム「ぴぼっとくん」を開発した。

「ぴぼっとくん」のシステム構成を図 2 に示す。「び

ぼっとくん」は、ブラウザ上からユーザが入力したキーワードをネットワークを介してサーバ側に送信し、サーバ内のデータベースに登録する。また送信されたキーワードをもとに画像を検索することも可能である。登録するキーワードが思い浮かばない場合には、画像データベース内で類似画像検索を行い、類似した画像に既に付与されているキーワードに基づき推薦キーワード一覧をユーザに提示する機能を持っている。我々が使用した類似画像検索技術の概要は次節で説明する。「ぴぼっとくん」はブラウザとデータベースとのインタフェースに

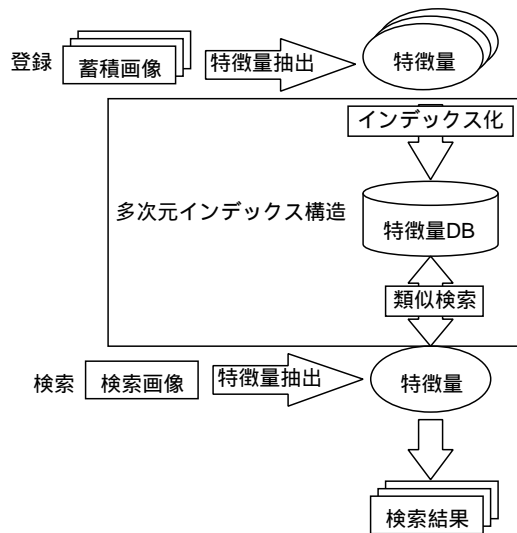


図 3: 類似画像検索処理の概要

PHP(Hypertext Preprocessor) を用いており、データベースには PHP と相性の良い MySQL を用いている。

3.2 類似画像内容検索

類似画像内容検索は、画像から計算で得られるさまざまな情報、つまり特徴量をもとに検索を行う技術である。用いる特徴量には、画像全体の色、模様、構図、画像内の物体の色、形、大きさ、模様、位置、配置などさまざまなものがある。これらの特徴量を多次元のベクトルで表現することにより、各特徴ベクトル間の距離の近さを類似性とみなし検索を行う方法が類似画像内容検索である。類似画像検索処理の流れを図 3 に示す。

我々が開発したシステムでは、色情報を表す特徴量として RGB ヒストグラムおよび HSI 尺度、また、形状情報を表す特徴量として Wavelet 変換を用いた形状特徴量を用いている。また、SR-tree [9] を多次元インデックス構造に用いることにより、検索の高速化を図っている。

さらに、色情報分布に基づく高精度な類似画像検索を可能にするため、Earth Mover's Distance [10] を用いた類似画像検索システムも開発している。Earth Mover's Distance は、2 つの分布間の距離尺度であり、一方の分布を他方の分布に変換するための最小コストにより距離を定義する。我々のシステムでは、まず平均移動アルゴリズム (mean shift algorithm)[11] により画像を粗く複数の領域に分解し、その後、各領域の中心、領域の広さ、領域の平均的な色を求める。この際、色間の類似性をと

らえるために、人間の知覚を反映しているとされている CIE-Lab 表色系で色を表現している。2 つの画像中の領域に対し、領域の中心点の (x, y) 座標上での距離を $\Delta x, \Delta y$ 、CIE-Lab 系での色差を $(\Delta L, \Delta a, \Delta b)$ とするとき、一方の領域内の画素を他方の領域内の画素に変換するためのコストを次のように定める。

$$\sqrt{(\Delta x)^2 + (\Delta y)^2 + (\Delta L)^2 + (\Delta a)^2 + (\Delta b)^2}$$

このとき、一方の領域を他方の領域に変換するための最小の総コストにより、2 つの画像間の距離を定義している。

3.3 「ぴぼっとくん」によるキーワード登録処理

本項では「ぴぼっとくん」によるキーワード登録処理の概要を登録手順の流れに従って説明する。図 4 に示すように、ユーザが画像データが格納されたフォルダを左フレームから選択すると、右下フレームにフォルダ内の画像のサムネイルが表示される。なお一覧形式の各画像上にマウスを移動することにより、マウスが指す画像が右上フレームに拡大表示される。各画像の選択は、個々の画像をクリックすることで可能であるが、同じキーワードを複数の画像に付与したい場合には、グループ登録機能も有している。

画像を選択すると、図 5 に示すキーワード登録画面に移る。キーワード登録の際には、システム側があらかじめ用意しているキーワード集合 (表 1 参照) から適当なものを選択してもよい¹、もし定義済みキーワード集合の中に適当なものが見つからなければユーザが自由にキーワードを与えることもできる。また、画像から受けるイメージは、感性語・印象語キーワードとして選択することができる。さらに、自由記述欄を設け、キーワードでは表現できない詳細な記述文を画像に付与することも可能である。

登録するキーワードが思い浮かばない場合には、推薦キーワード一覧表示を選択すると、図 6 に示す推薦キーワード一覧画面に移る。ここでは、画像データベース内で類似画像検索を行い、類似した画像に既に付与されているキーワード上位 5 件を結果として出力する。類似画像検索にはすでに説明した Earth Mover's Distance に基づく画像検索システムを使用している。

次に、構築したデータベースに対する検索機能につい

¹ 11 個の大項目 (カテゴリ) と約 450 個の小項目 (キーワード) が用意されている。



図 4: 「ぴぼっとくん」の画像選択画面



図 5: 「ぴぼっとくん」のキーワード登録画面

表 1: 定義済みキーワード項目

大項目	小項目 (約 450 個)
風景	空, 山, 海, 町, 森, 村, 公園, 島, 雪, 流氷 … (総数 53 個)
地理	アフリカ, フランス, アメリカ, 日本, エジプト, 中国, 京都, ニューヨーク, ドイツ … (総数 56 個)
建築物	橋, 家, 城, 遺跡, ビル, 万里の長城, 寺, 灯台, ドーム, 神社 … (総数 39 個)
人工物	家具, 絵画, コップ, ドア, 看板, 椅子, 気球, 切手, 時計, 彫刻 … (総数 59 個)
乗物	車, バス, 列車, カヌー, 自転車, 飛行機, 船, トラック, バイク, スポーツカー … (総数 20 個)
食物	料理, リンゴ, 野菜, ジュース, トマト, ビール, デザート, ケーキ, ワイン, 洋食 … (総数 51 個)
人間	女性, 群衆, モデル, 日本人, 足, 手, 赤ん坊, 美人, 金髪, 老人 … (総数 33 個)
スポーツ	野球, フットボール, ゴルフ, ダンス, 水泳, アイスホッケー, スキー, サーフィン … (総数 34 個)
動物	鳥, 犬, 喋, 魚, 牛, 馬, 麒麟, 像, 熊, サイ … (総数 79 個)
植物	花, 木, 海草, 果実, 盆栽, 薔薇, シダ, ヤシ, サボテン, キノコ … (総数 13 個)
鉱物	石, 岩, 宝石, 水晶, 鍾乳石, メノウ … (総数 6 個)



図 6: 推薦キーワード提示

て説明する．図 7 に示すように，本システムでは付与されたキーワードに基づき画像検索を実行する機能も有している．さらに，図 8 に示すように画像を選択し，各画像に付与済みのキーワードとその頻度分布をユーザに提示する機能も有している．この機能を用いると，他のユーザが各画像にどのようなキーワードを付与しているか確認することが可能となる．以上のように，本システムはキーワードの登録機能だけでなく，キーワードもしくは画像からの高速な検索機能も持ち合わせている．

3.4 画像キーワードデータベース

「ぴぼっとくん」を実際に使ったのキーワード付与は，2002 年 2 月末から行われており，2002 年 4 月時点で約 17,000 件の画像にキーワードが付けられている．2003 年 3 月には，約 4 万件の画像に対するキーワードデータベースが完成する予定である．



図 7: キーワードによる検索画面

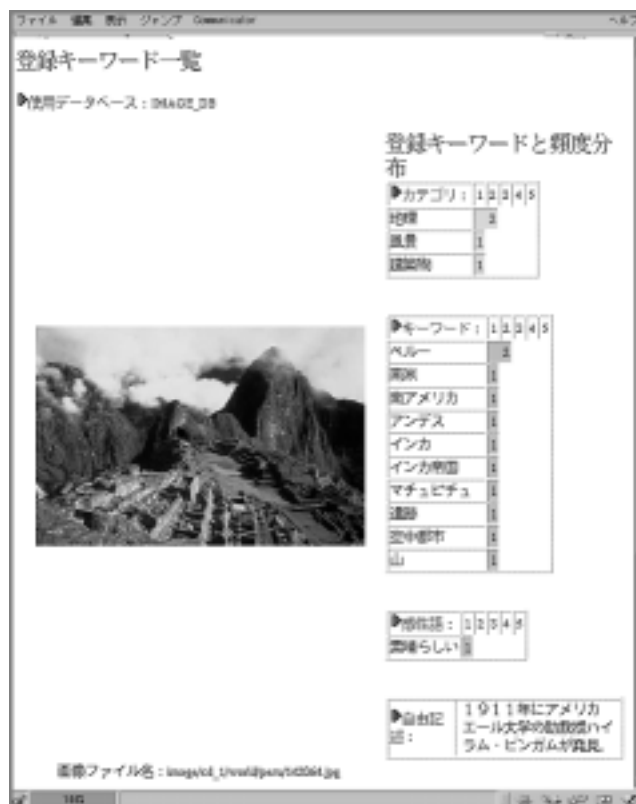


図 8: キーワードの付与された画像の例

4 まとめ

本稿では、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「ぴぼっとくん」について述べた。今後、キーワード付き画像データ数の拡大、付与されたキーワードを学習データとして用いるクロスメディア検索システムの開発などを行っていく予定である。また、キーワード付き画像データベースは、画像検索システムの評価にも有効であると考えられるため、データベースの公開についても検討していきたい。

謝辞：画像キーワード登録システム「ぴぼっとくん」の開発を手伝ってくれた現ダイワボウ情報システム株式会社の木内純氏に感謝する。本研究の一部は、財団法人放送文化基金の援助による。

参考文献

- [1] Google homepage. <http://images.google.com/>.
- [2] Lycos Multimedia Search homepage. <http://multimedia.lycos.com/>.
- [3] Alta Vista Image Search homepage. <http://images.altavista.com/>.
- [4] 呉君錫, 金子邦彦, 牧之内顕文, Sang-Hyun Bae. Wavelet-som に基づいた類似画像検索システムの設計・実装と性能評価. 情報処理学会論文誌データベース, Vol. 42, No. SIG 1(TOD 8), pp. 1-11, 2001.
- [5] 上川伸彦, 岩崎一正. 画像の類似検索に向けた多次元インデックス手法. 情報処理学会論文誌データベース, Vol. 42, No. SIG 1(TOD 8), pp. 140-147, 2001.
- [6] 小早川倫広, 星守, 大森匡. ウェーブレット分解係数の階層的相関関係を用いたテキスト類似画像検索. 情報処理学会論文誌データベース, Vol. 42, No. SIG 1(TOD 8), pp. 12-20, 2001.
- [7] Liu Wenyin, Susan Dumais, Yanfeng Sun, HongJiang Zhang, Mary Czerwinski, and Brent Field. Semi-automatic image annotation. In *Human-Computer Interaction Interact '01*, Hirose, M. (Ed.), IOS Press, 2001.
- [8] Bernt Schiele and Julia Vogel. Vocabulary-supported image retrieval. *Proceedings of the First DELOS Network of Excellence Workshop on Information Seeking, Searching and Querying in Digital Libraries*, 2000.
- [9] 片山紀生, 佐藤真一. SR-tree: 高次元・データに対する最近接検索のためのインデックス構造の提案. 電子情報通信学会論文誌, Vol. J80-D-1, No. 8, pp. 703-717, 1997.
- [10] Y. Rubner, C. Tomasi, and L. J. Guibas. A metric for distributions with applications to image databases. *Proc. of IEEE International Conference on Computer Vision*, pp. 59-66, 1998.
- [11] D. Comaniciu and P. Meer. Robust analysis of feature spaces: Color image segmentation. *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 750-755, 1997.