

第 8 章

WWW からの画像収集方法の提案

本章では、実世界画像を大量にかつ手軽に収集する方法として、キーワード入力のみで WWW(World-Wide Web) から自動的に大量の実世界画像を収集する方法について提案する。WWW 空間中に存在する画像は現在数十億枚と言われ、様々な画像が存在している。WWW 空間中に存在する画像はその多くが画像内容を表すテキスト情報を伴っているので、テキスト情報を解析することによって、ユーザが入力したあらゆるキーワードに関連した画像を WWW から収集することが可能である。実験では、最初にキーワードを与えるだけで、5 分程度の処理時間で 100 枚以上の画像を収集することが出来ることを示す。

8.1 はじめに

近年の WWW(World-Wide Web) の普及によって、我々がインターネットを通してアクセスできる画像データの量は爆発的に増大している。WWW 上の画像は容易にアクセスすることが可能であるので、WWW を巨大な画像データベースであると見なすことが出来る。ただし、WWW 上の画像データは無秩序な形で存在しており、一般の画像データベースの様に画像内容によって分類されていたり、適切なキーワードが付けられていることは少ない。

そうした WWW 上の無秩序な画像データに対して、テキストデータに対するのと同様に、WWW 上の画像データに対してキーワードに基づく検索を可能としている検索エンジンが存在する。例えば、Google Image Search[155], Ditto[156], Yahoo! Picture Gallery[157], Lycos Multimedia Search[158], AltaVista Image Search[159] などは、その代表例である。しかし、これらの検索エンジンは、画像の内容の解析をせずにテキスト検索の場合と同様にページに含まれるキーワードのみを手がかりにした検索を行っており、無関係な画像が検索結果として返されることが多く、検索精度は高くない。また、検索結果として上位数百件程度の画像しか返さないことが多く、大量の画像を収集するのには向いていない。

一方、画像データに対する検索技術として、内容に基づく画像検索 (Content-Based Image Retrieval, CBIR)[124, 125] が従来より研究されている。キーワードに基づく画像検索では、人手によって予めすべての画像データにキーワードを付けておく必要があるが、CBIR では各画像データ

から画像特徴を自動抽出して画像間の類似度を判定し、ユーザが指定した画像に類似した画像を検索する。

そこで、本章では、キーワードに基づく WWW 上の画像検索と、内容に基づく画像検索を組み合わせることによって、ユーザの指定したキーワードが表わす画像を WWW から自動的に大量に収集するシステム Image Collector を提案し、システムの設計、実装について述べる。Image Collector は、(1) 商用のテキスト検索エンジンを利用するため、予め大規模なインデックスを作ることを必要とせず、(2) 画像内容に基づく画像のクラスタリングを行うことによって、ユーザとのインタラクティブなやりとりなしに、最初に 1 度キーワードを指定するだけで大量の画像を収集できるという、従来の WWW に対する画像検索システムにはない 2 点の新しい特徴を備えている。

画像認識や CBIR などの研究においては大量の実験画像が必要になる場合があるが、大量の画像を研究者自身が集めることは困難なことであり、通常は市販の画像コレクションを利用することが多かった。けれども、市販の画像コレクションはプロの写真家が撮影した整った画像のみを集めているので、多様性に欠けるという欠点があった。一方、WWW 上に存在する WWW 上の画像は、様々な人が様々な目的で撮影した画像であり、実世界にある画像の多様性をそのまま反映していると考えられ、WWW 上の画像の自動収集は画像認識や CBIR の実験画像の収集にも役立つと思われる。特に近年、大量の画像を学習画像として画像認識を行う example-based による一般物体認識 (generic object recognition) の研究が盛んに行われるようになっており [61, 160, 161]、WWW から自動収集した大量の画像を画像認識のために利用するという応用も可能である [162]。

8.2 関連研究 と その問題点

本章での研究で目的とする、キーワードに基づく画像検索と内容に基づく画像検索を組み合わせた WWW に対する画像検索についての研究は、既にいくつか提案されている。その中でも代表的な研究が、WebSeer[163]、WebSEEk[164]、Image Rover[165] の 3 つである。

これら 3 つの研究は、検索時にユーザが欲しい画像に関係するキーワードを入力するという点ではほぼ同じで、検索に用いるキーワードについても、「WWW 上の画像は HTML 文書にそのファイル名が記されてリンクされており、画像を説明するキーワードが HTML 文書に含まれていることが多い」という経験則を利用して、通常の HTML 文書の検索手法と同等の方法を用いて、画像をリンクしている HTML 文書からキーワードを抽出して、その画像を検索する場合のキーワードとしている。

WebSeer[163] はキーワード及び、画像の特徴として、画像の大きさ、主要な色、写真か図であるかの区別、人の顔の有無などを検索時に指定する。WebSeer では、画像内容については、Face Detector[42] などの既存の画像認識のモジュールを利用することによって利用している。このシステムでは、画像同士の類似度は用いられていない。

WebSEEk[164] は初めにキーワードのみによって画像を検索し、さらにその検索結果の画像群の中からユーザが望む画像を指定して、さらに CBIR の手法を用いて、指定画像に類似している画像を検索するという、2 段階の検索が前提となっている。図 8.1 に WebSEEk での検索手順と検索例を

示す。なお、類似画像の検索には、同じグループで研究されている VisualSEEk[166] の技術が用いられている。

Image Rover[165] は基本的には WebSEEk と検索方法は同じであるが、2 段階目の類似画像検索において、画像にリンクしている HTML 文書が含む単語の出現頻度のベクトルと、画像特徴のベクトルを統合して、検索を行っている点が WebSEEk とは大きく異なる点である。同様の手法による研究は、キーワードが予め付与された画像データベースに対する画像検索システムに対しても行われている [167, 168]。



図 8.1 WebSEEk における画像検索。最初にキーワード検索を行い、その結果からユーザが望む画像を選択し、次に画像特徴量に基づく類似画像検索を行う。2 段階のインタラクティブな検索になっている。

WebSEEk と Image Rover では、初めにキーワードによる画像検索を行って、さらにその検索結果の画像からユーザが必要とする画像を指定して、その類似画像を検索する (図 8.1) という様に、本研究とは異なり、インタラクティブに画像検索を行うことを前提としている。そのため、少量のユーザの希望に合った画像を検索するには向いているが、大量に自動的に収集するような目的には適していない。また、WebSeer では、Face Detector のような対象に依存したモジュールを利用しているため、人物画像以外では画像内容の利用が十分ではない。

一方、我々の研究では「検索」ではなく、例えば、「ライオン」「富士山」のようにあるカテゴリーに属する画像を大量に「収集」することとするために、ユーザとのインタラクティブなやりと

りを考えず、初めにユーザが指定を行うだけで画像が収集可能になるシステムの構築を目指す。そのために、画像のクラスタリングを利用することによって、処理途中でユーザによる画像指定を不要とする。

また、3つのシステムとも、実用的なシステムとするには、予め巨大な画像検索専用のインデックスを作成しておくことが必要であり、そのためには大規模な計算機資源と多くの時間が必要になってくる。一方、我々の研究では、既存のキーワードベースのテキスト検索エンジンを利用することによって、予め巨大なインデックスを作成しておく手間を不要にするという方法を採用する。

他にキーワード検索と CBIR を Web 画像検索に対して組合せて適用したシステムとしては、S.Mukherjea らによる Web 画像検索システム AMORE [169] がある。これも WebSeek と Image Rover と同様にインタラクティブな検索であり、また同様に予め画像を大量に収集してインデキシングしておく必要がある。一方、N.Bianchi ら [170] や R.Inder ら [171] による K-DIME では、商用の画像検索エンジンにキーワードを与えて、その検索結果を予め学習された感性語と画像特徴の対応関係を利用し、感性語による画像フィルタリングを行うことによって、「ロマンチックなハワイ」といったような通常の Web 画像検索エンジンでは検索が難しい画像を検索することが可能となっている。

WWW に対する画像検索において、以上述べた代表的な 3つの研究以外にも、キーワードのみ基づく画像検索や、Web ページのリンク構造を考慮して画像検索を行う研究などが提案されている。

キーワード解析のみを用いる方法としては、H.T.Shen ら [172] による研究がある。この研究では、画像タグ周辺のテキストを ChainNet という表現によって表し、単なるキーワードマッチングによる方法よりも検索精度を改善している。単語の重要度の重み付けをユーザとのインタラクションから学習し精度を改善できる。

Web ページ同士のリンク構造を用いて Web 画像と検索キーワードとの関係評価を行うシステムとして、Web 画像検索システム PicAShow [173, 174] がある。この研究では、J.M.Kleinberg による HITS アルゴリズム [175] と、検索エンジン Google で用いられているアルゴリズムである S.Brinn と L.Page による Page Rank [176] を用いた 2通りの方法で、画像検索の実験を行っている。また、HITS アルゴリズムにおける、検索トピックに関する重要なページである authority page, authority page へのリンクを多く持っている hub page の概念を拡張して、画像検索における、検索トピックに関する画像を多く含んでいる container page, container page へのリンクを多く持っている hub page の概念を提案している。また、C.C.Yiu ら [177] は PicAShow に対して、いくつかの改良されたリンク構造解析方法を適用し、その検索精度の改善を提案している。

8.3 画像収集システムの概要

本研究におけるシステム Image Collector では、ユーザの与えたキーワードが表現する画像（以下、「目的画像」と記す）をできるだけ多く収集することを目的とする。Image Collector は、従来の WWW に対する画像検索システムにはない以下の 2 点の新しい特徴を備えている。

- (1) 事前の巨大インデックスの作成不要。
商用のテキスト検索エンジンを利用するため、事前に WWW ロボットを動かして大規模なインデックス作成をすることが不要で、手軽にシステムを利用できる。しかも、テキスト検索エンジンは世界中に多数存在するので、全世界から画像を収集可能。
- (2) 画像特徴を利用したクラスタリングによる検索結果の自動的な絞込。
従来の画像検索システムと異なり、処理途中でのユーザの介入が不要で、初めに欲しい画像のキーワードを指定するのみで、大量の画像を収集できる。そのため、例えば、深夜に空いている計算機とネットワークを利用して、ユーザの指定したキーワードに関する画像を大量に集めることができる。

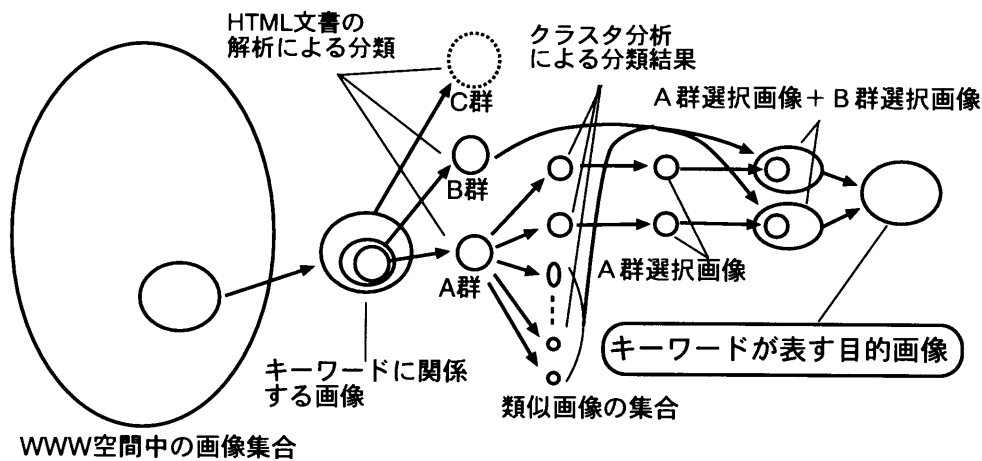


図 8.2 画像自動収集の方法.

次に画像収集システムの処理の概要について述べる。最初に WWW からキーワードに関する画像を集め、次に集めた画像を画像特徴量に基づいて選別してノイズを取り除き、最終的にユーザの与えたキーワードに対応する目的画像を収集する (図 8.2)。以下では、キーワードのみを用いて WWW から収集された画像を「収集画像」、画像特徴量に基づいて収集画像から目的画像として選ばれた画像を「選択画像」と呼ぶこととする。

第一段階の、キーワードに関する画像の WWW からの獲得は以下の手順で行う。

1. 既存のテキスト検索エンジンを利用し、ユーザの与えたキーワードに関する Web ページの URL(Universal Resource Locator) を集める。

2. 集めた URL が示す Web ページにアクセスして、各 Web ページの HTML 文書を獲得する。
3. 各 HTML 文書に対して HTML タグに基づく解析をすることによって、HTML 文書からリンクされている画像ファイルとキーワードとの関係の強さについての評価を行い、評価の高いものから順に A 群、B 群、C 群に分類する。
4. A 群、B 群に該当した画像ファイルのみを実際に WWW から収集する。C 群に該当した画像ファイルはキーワードとは無関係の画像と見なして収集しない。

以上の処理は、「WWW 上の画像は HTML 文書にそのファイル名が記されてリンクされており、画像を説明するキーワードが HTML 文書に含まれていることが多い」という従来の WWW 画像検索エンジンでも用いられている WWW 上の画像の一般的な性質に着目して行っており、初めにテキスト検索エンジンによって、ユーザの指定したキーワードを含む HTML 文書を獲得し、解析している。そして、HTML 文書からリンクされている画像ファイルがあるかどうか調べ、画像ファイルへのリンクタグやタグ周辺のテキストに指定キーワードがどの程度含まれているかどうかを調べることによって、画像ファイルとキーワードとの関係の強さを評価し、画像ファイルを A 群、B 群、C 群のそれぞれ以下のような意味を持つ 3 つのグループに分類する。なお、評価の方法としては、WWW 画像検索システムにおいて一般的に用いられている方法を利用する [163, 165, 178]。

A 群 キーワードとの関係が強く、キーワード評価のみでほぼ正解画像と見なせる画像。

B 群 キーワードとの関係が中程度で、さらに画像内容の評価が必要な画像。

C 群 キーワードとの関係が低く、正解画像である可能性が低い画像。

A 群は画像内容を説明する `alt` タグや画像ファイル名にキーワードが含まれているという厳しい条件によって選別される。そのため、選ばれる枚数は少ないが、HTML 文書の解析のみでほぼ正解画像であるとみなすことが出来る。B 群は、画像ファイルへのリンクタグの周辺のテキストやタイトルタグにキーワードが含まれるという、A 群より緩い条件によって選ばれる。そのため、通常、A 群よりも選ばれる枚数は多くなるが、A 群と比べると正解画像の含まれる割合は低くなり、WWW からの収集後に、さらに画像内容による選別が必要となる。C 群は、A 群、B 群以外であり、HTML 文書中で画像ファイルへのリンクタグと離れた位置にキーワードが存在するようなほとんど見込みのない画像ファイルがこれに該当する。そのため、C 群の画像ファイルは実際には WWW からは収集しない。

抽出した画像ファイルの URL を以上の様に 3 つに分類する理由は、C 群を収集しないことによってコストかかる画像ファイルの収集量を減らし、さらにその後の画像内容の解析フェーズで、キーワード評価のみでほぼ正解画像とみなせる A 群を質問画像として、画像内容の評価が必要な B 群から、画像特徴を用いた CBIR による画像検索を行い、画像選択を行うためである。

CBIR において画像特徴を利用して画像検索を行う場合、検索の手がかりとなる画像、つまり質問画像の画像特徴が与えられる必要があるが、Image Collector では、他の WWW 画像検索システムと異なり、正解画像をユーザがインタラクティブに指定することはない。また、どのような

キーワードが与えられるかは事前には分からないので、予めキーワードと画像特徴の対応付けをし置くことも不可能である。そこで、A 群画像を CBIR での質問画像とみなして、それらと画像特徴が類似している画像を B 群の画像の中から選択する。ただし、A 群画像も少数の無関係の画像が含まれていることがあるので、盲目的にすべてを正解画像であるとみなすことはせずに、画像内容が他の A 群画像に比べて著しく特異である画像を除去することとする。

第二段階の画像内容の解析による A 群、B 群からの目的画像の選択は、以下のように行う。

1. WWW より収集した A 群、B 群の各画像について画像特徴を計算し、画像特徴ベクトルを作成する。
2. 画像特徴ベクトルを利用して、A 群の画像間の距離を計算し、階層的クラスタ分析 [179] による類似画像同士のグループ化を行う。階層的クラスタ分析によって生成された画像クラスタのうち、メンバー数がある一定数以上になったクラスタに含まれる画像を選択して目的画像とし、それ以外のクラスタの画像を消去する。
3. B 群および消去されたクラスタに属する A 群の画像から、選択された A 群の各クラスタの平均特徴ベクトルに特徴ベクトルの距離がある一定値より近いものを選択し、それらを A 群からの選択画像と併せて目的画像としてシステムの最終結果とする。

以上の画像内容の解析に関する処理では、小クラスタの画像の除去によって、A 群画像中で画像内容が他と大きく異なる画像のみを除外して、残りの画像を正解画像であるとみなしている。そして、さらに、各クラスタの平均画像を質問画像とみなして、B 群画像および消去されたクラスタに属する A 群画像から類似画像を検索していることになる。

以上の処理中での、小クラスタの画像の除去は、画像内容が他の A 群画像に比べて著しく特異である画像を除去することであり、「**類似画像が多い一般的な画像の方が、類似画像の少ない特異な画像よりも、キーワードを表す画像としてふさわしい場合が多い**」という経験則に基づいて行っている。これは、多数決によって目的画像を選択しているといえる。

従来より画像特徴に基づく画像のクラスタリングをユーザへの画像提示に利用した研究 [180] は行われているが、大量画像収集のための検索結果の自動的に絞込みに利用することは、本研究における新しい提案である。

8.4 システムの実装

画像収集システム Image Collector は、現在の実装では、入力されたキーワードに基づいて WWW から HTML 文書や画像ファイルを収集する画像収集部と、画像内容を解析することによって収集部で収集された画像から画像選択を行う画像解析部の 2 段階に分かれている (図 8.3)。以下では、画像収集部、画像解析部それぞれについて説明する。

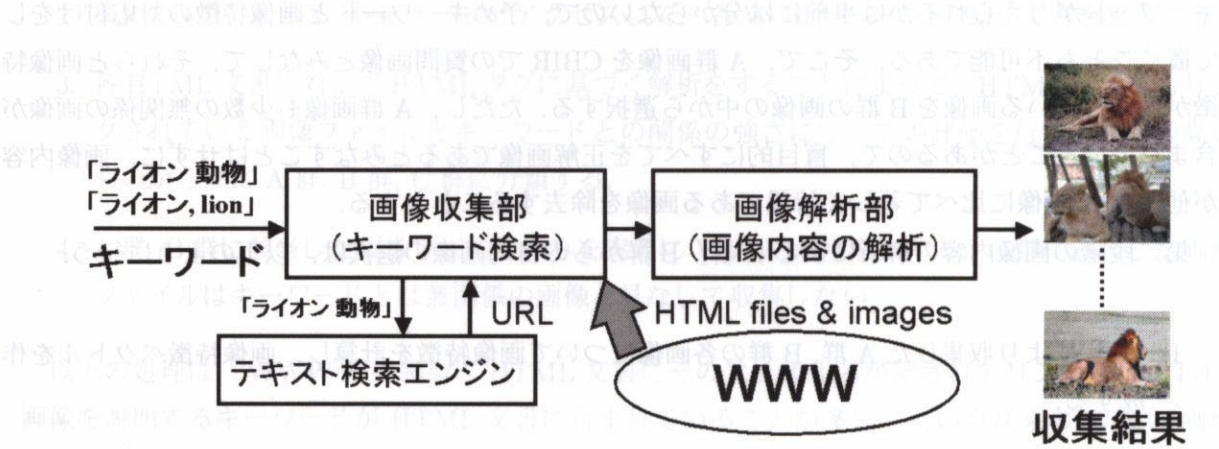


図 8.3 画像収集システムの構成.

8.4.1 画像収集部

収集部は、図 8.4 に示すように、メインモジュールと多数の収集プロセスからなる。メインモジュールは、収集する HTML 文書や画像ファイルの URL リストの作成、収集した HTML 文書の解析などを行い、収集プロセスは、URL リストに基づいて WWW から HTML 文書及び画像ファイルの収集を行う。

収集部では、以下の手順で処理を行い、テキスト情報のみからキーワードに関係のある画像を収集する。

1. ユーザが 3 種類のキーワード (画像分類用、検索エンジン用、ファイル名) をメインモジュールに与える。
2. メインモジュールは検索エンジンに検索エンジン用キーワードを与え、HTML 文書の URL を獲得し、URL リストを作成する。
3. 収集プロセスは URL リストから URL を 1 つずつ獲得し、WWW から HTML 文書または画像ファイルの収集を行う。収集した HTML 文書はメインモジュールへ、画像ファイルは解析部へそれぞれ送る。収集プロセスは、この処理を URL リストが空になるまで繰り返す。
4. メインモジュールが HTML 文書を受け取るとその HTML 文書を解析して、画像ファイルの URL を抽出する。そして、各画像ファイルについて、キーワードとの関係の強度に関する評価を 8.4.1 節で示す方法で行い、A, B ランクの画像の URL を URL リストに追加する。
5. 以上の処理で URL リストが空になって、ある一定数以上の収集プロセスが処理を終了した時に、収集部の処理を終了する。

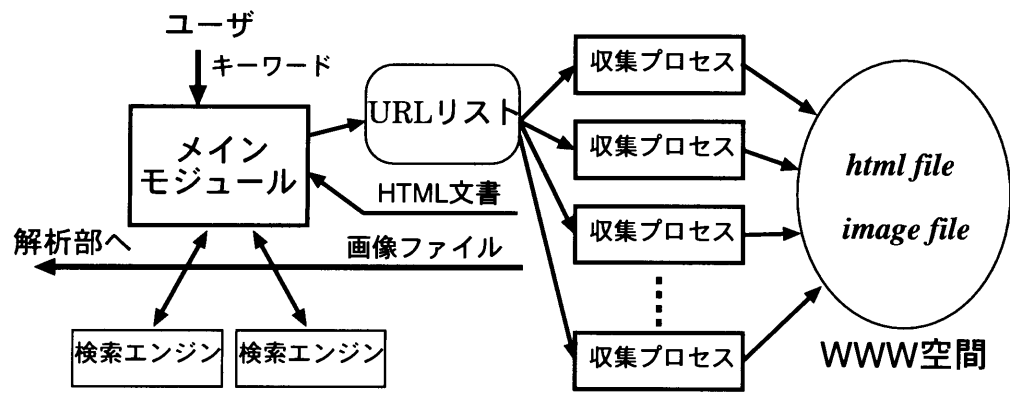


図 8.4 収集部のシステム構成.

キーワードの指定

ユーザは初めにメインモジュールに、**画像分類用キーワード**、**検索エンジン用キーワード**、**ファイル名キーワード**の3種類のキーワードを与える。キーワードは日本語または英語の単語であり、それぞれ複数与えることができる。

画像分類用キーワードは、収集したい画像を表すキーワードであり、例えば、ライオンの画像が収集したい場合は「ライオン」と指定する。

検索エンジン用キーワードは検索エンジンに検索キーワードとして与えるキーワードで、複数のキーワードを指定することによって、画像分類用キーワードよりも収集したい画像を限定する必要がある。動物のライオンの画像を収集したい場合は、「ライオン 動物」などとする。検索エンジン用キーワードを画像分類用キーワードと同じにしてしまうと、同じ「ライオン」でも、家庭用洗剤のメーカー、マンション、プロ野球の球団、自動車などの動物以外の「ライオン」が検索エンジンにヒットしてしまう可能性がある。

3つめの**ファイル名キーワード**は、HTML 文書解析時に、画像ファイル名を手がかりとして用いる時に必要となるもので、画像分類用キーワードを表現する英単語を通常は指定する。画像分類用キーワードが「ライオン」の場合は、「lion」と指定する。

なお、以上の説明では、日本語 Web サイトを対象とした検索の場合の例を挙げたが、英語 Web サイトを対象とする場合は画像分類用キーワード、検索エンジン用キーワード、ファイル名キーワードをそれぞれ、「lion」「lion animal」「lion」とすればよい。

検索エンジンからの URL の獲得

次に、メインモジュールは、検索エンジン用キーワードを既存のテキスト検索エンジンに与え、キーワードに強く関係する WWW 上の HTML 文書の URL のリストを求める。検索エンジン用キーワードが複数の時は、すべてのキーワードを含む HTML 文書の URL を返すように AND 指定をする。キーワードに関係する URL を取得するための検索エンジンとしては、Google、Lycos な

どのロボット収集型のテキスト検索エンジンを利用し、より多くの URL を集めるために 1 つではなく複数の検索エンジンを同時に用いる。

システムは HTTP(Hypertext Transfer Protocol) によるソケット通信によって、直接、検索エンジンにキーワードを与え、結果として返される HTML 文書から検索結果の URL を抽出する。キーワードの与え方、返される HTML 文書の書式は検索エンジン毎に異なるので、予め解析してシステムにその方法を与えておく。

複数の検索エンジンからそれぞれ上位数 100 個程度にヒットした URL を収集し、重複している URL を除去してから、すべての URL を 1 つの URL リストとしてまとめる。

HTML 文書および画像ファイルの収集

収集プロセスは、URL リストからそれぞれ 1 つずつ URL を獲得し、収集した URL が示す HTML 文書または画像ファイルをソケット通信で取ってくる。URL リストには、最初は検索エンジンから獲得した HTML 文書の URL しか含まれていないが、8.4.1 節で述べるように、やがて収集した HTML 文書から画像ファイルの URL が抽出されて、URL リストに画像ファイルの URL が追加されることとなる。

収集プロセスの実行中には通信の待ち時間が多くあるので、高速化のために収集プロセスは同時に複数プロセスを並行実行する。収集プロセスが HTML 文書を収集した場合には、HTML 文書はメインモジュールに送られる。また、画像ファイルを収集した場合は、収集画像ファイルがある一定の大きさ以上であるかチェックし、条件を満たしていれば、決められたサイズ (現在の実装では 240×180) に正規化してから、画像を解析部に送る。条件を満たさない一定の大きさ未満である小さい画像はそのまま捨てられる。収集プロセスは、この処理を URL リストが空になるまで繰り返す。

HTML 文書の解析

収集プロセスから送られた HTML 文書はメインモジュールにおいて、その内容が直ちに解析される。まずは、HTML 文書に含まれる画像を Web ページ中に張り込むタグ (`img src`) 及び画像に直接リンクを張るタグ (`a href`) から、そのページに含まれるもしくは直接リンクされる画像ファイルの URL を求める。そして、後述する画像ファイルの評価法によって評価を行い、A もしくは B ランクに評価された画像の URL を URL リストに追加する。もし、1 つも画像ファイルの URL が含まれていない時は、その HTML 中に同一 Web サイト内の HTML 文書へのリンクタグ (`a href`) があれば、そのリンク先の URL を URL リストに追加する。

URL を URL リストに追加する際には、同じ URL の HTML 文書や画像ファイルに 2 回以上をアクセスするの防ぐため、既に同一 URL を URL リストに登録したかどうかチェックしてから、未登録の場合のみ登録することとする。

次に HTML 文書の解析による、画像ファイルとキーワードの関連性の強さの評価法について述べる。まずは、HTML 文書中でキーワードがどの HTML タグに修飾されているか、もしくは画像

ファイルの URL の近くに存在しているかなどを以下に示す具体的な条件について調べることによって、画像ファイルとキーワードの関連強度の評価点を求める。

条件 1 以下の条件を 1 つ満たす毎に評価点に 3 点を加える。

- HTML 文書から抽出した画像ファイル名にファイル名用キーワードを含む (表 8.1(1)(2)).
- 画像が `img src` で参照されている場合は、画像内容を説明するタグである `alt` タグ内に画像分類用キーワードを含む (表 8.1(1)).
- 画像が `a href` タグで参照されている場合は `a href` と `/a` タグの間に画像分類用キーワードを含む。 (表 8.1(2)).

条件 2 以下の条件を 1 つ満たす毎に評価点に 1 点を加える。

- `title`, `meta name= "description"`, `meta name="keyword"` の各タグに画像分類用キーワードを含む (表 8.1(3)),
- 画像ファイルのタグの直前の見出しタグ (`H1`, ..., `H6`) に画像分類用キーワードを含む (表 8.1(4)).
- 画像ファイルのタグの前後 3 行に画像分類用キーワードを含む (表 8.1(5)).

評価点の合計が 3 点以上を A ランク, 2 点から 1 点を B ランク, 0 点を C ランクとする。なお, A ランク, B ランクに分類された画像をそれぞれ A 群画像, B 群画像と呼ぶこととする。

こうした HTML タグを利用したキーワードと画像ファイルの関係の強度に関する評価方法は、テキスト検索エンジンにおいてキーワードと HTML 文書との関係の強度の評価に用いられていた方法であり、WWW 画像検索システムにおいては一般的に用いられている手法である [163, 165, 178].

8.4.2 画像解析部

画像解析部は、逐次処理を行う単一のモジュールによって構成され、収集部で A 群画像, B 群画像として収集された画像から画像特徴量を抽出し、それを利用してキーワードに関係する画像を最終的に選択する。

画像特徴量の抽出

収集部で収集され正規化されたすべての画像について、画像特徴量を計算する。画像特徴量としては、テクスチャやエッジの方向、部分領域の形状や位置関係など様々な方法が CBIR の分野において研究されているが、現在の実装ではカラーヒストグラムのみを使用している。カラーヒストグラムは対象とする画像を限定しない手法で、簡単な処理である割に画像の特徴をよく表現することができ、画像検索の研究において一般的に用いられている手法である [15].

具体的には、収集された画像データを RGB データから人間の感性に近いと言われる均等色空間である Lu^*v^* 表色系に変換する [181]. 均等色空間とは、色空間における 2 点間のユークリッド距離が人間の感覚的な色の差に比例する色空間であり、同程度の距離であっても人間にとっては色の差

表 8.1 画像ファイルのキーワード評価の手がかりとなる HTML タグの一覧と例.

(1) 画像表示タグ	
(2) 画像リンクタグ	これがトラです
(3) 強調キーワードタグ	ページタイトル <TITLE> ZOO動物園のライオン</TITLE> キーワード <META name="keyword" content="動物園,ライオン" > 内容説明 <META name="description" content="トラのページ">
(4) 見出しタグ	<H1>ライオンの写真館</H1> <H4>トラのコーナー</H4>
(5) 画像表示タグ, 画像リンクタグに近いキーワード	 上の画像はトラの親子です.

が大きく異なることがある RGB や YCbCr 等の色空間よりも、色を用いて画像間の距離を計算するための画像特徴には適している色空間であると言われている [182]¹. そして、次に Lu^*v^* 色空間を各軸について 6 つに均等分割し、全体で $6 \times 6 \times 6$ の 216 の部分空間に分割する. 具体的には、各軸 0 から 255 の整数値をとる RGB 空間で表現できるすべての色を Lu^*v^* 空間に変換して、 Lu^*v^* 空間の各軸についてその最小値から最大値までの範囲を 6 つに均等に分割した. そして、それぞれの部分空間についての頻度を計算し合計 216 個のピンを持つヒストグラムを作成する. このようにして、収集した全画像について 216 次元の特徴量ベクトルを作成する.

A 群画像のクラスタリング

次に、A 群に分類された画像のみについて、各画像の特徴ベクトルを利用して、すべての画像間の距離を計算する. 距離は単純なユークリッド距離ではなく、以下の式に従って、色ヒストグラムの各ピンの平均色ベクトル間の距離 d_{ij} を考慮して計算した [183]. この距離は quadratic form distance と呼ばれている.

$$D_{ij} = (h_i - h_j)^t A (h_i - h_j) \tag{8.1}$$

ただし、

$$A = [a_{ij}] \tag{8.2}$$

¹ただし、本章の画像収集システムにおいては、RGB 色空間を利用した場合の実験結果と比べて、それほど大きな差が表われる訳ではなかった.

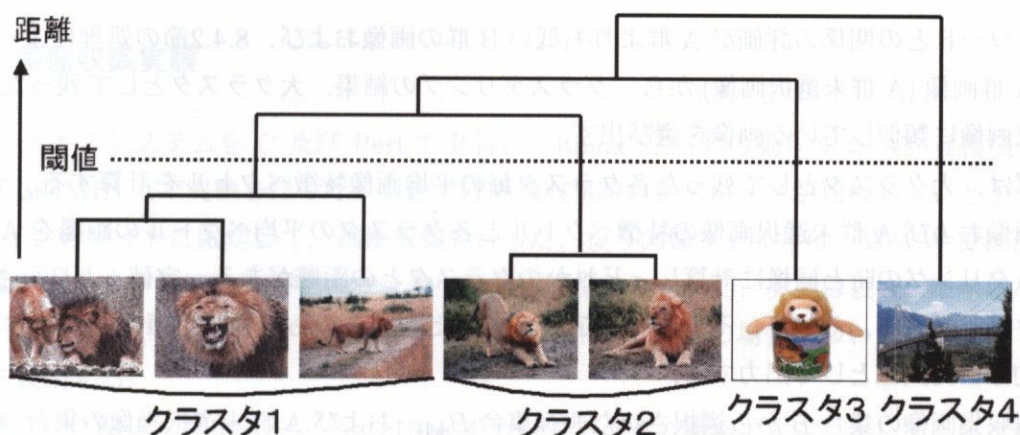


図 8.5 クラスタリングの結果の例.

$$a_{ij} = \begin{cases} 1 & (i = j) \\ e^{-cd_{ij}/d_{\max}} & (i \neq j) \end{cases} \quad (8.3)$$

$$d_{\max} = \max_{ij}(d_{ij}) \quad (8.4)$$

h_i, h_j はそれぞれ i 番目, j 番目の画像の特徴量ベクトルを表わし, c は定数, d_{ij} はヒストグラムの各ビンの平均色間のユークリッド距離を表す.

以上のようにして画像間の距離を求めたら, それを利用して, クラスタ分析 [179] による類似画像同士のクラスタリングを行う. クラスタ分析を行う際には, 異なるクラスタのメンバー同士の距離でもっとも大きい値をクラスタ間の距離とする FN 法 (farthest neighbor method) を採用し, 階層的クラスタリングを行う. ある一定距離以下の画像もしくはクラスタ同士を併合していき, すべてのクラスタ間の距離がある一定値以上になるまで, クラスタの併合を繰り返す. 図 8.5 に階層的クラスタリングの結果のデンドログラム (樹形図) を示す. 似ている画像が多くある場合は大クラスタになり, 他に似ている画像が少ない場合は, 小クラスタになる.

ここでは, 大クラスタに分類された画像がキーワードが表わしている典型的な画像で, 小クラスタに分類された似ている画像が少ない画像は特異な画像であると見なして, メンバー数がある一定数 m 以上になった比較的大きなクラスタの画像を正解クラスタとして残して, それ以外のクラスタの画像を消去する.

A 群収集画像の集合 A が n 個のクラスタ C_1, \dots, C_n にクラスタリングされて, $C_1, \dots, C_{n'}$ が $|C_j| \geq m$ ($|C_j|$ はクラスタ C_j に属する画像の枚数を表す) を満たすとする, A 群選択画像の集合 A_{sel} は, 次のように表すことができる.

$$A_{\text{sel}} = \{a_i \in A \mid a_i \in \bigcup_{j=1}^{n'} C_j\} \quad (8.5)$$

B 群画像からの選択

キーワードとの関係の評価が A 群よりも低い B 群の画像および、8.4.2 節の処理によって除去された A 群画像 (A 群未選択画像) から、クラスタリングの結果、大クラスタとして残った A 群からの選択画像に類似している画像を選び出す。

まずは、大クラスタとして残った各クラスタ毎の平均画像特徴ベクトルを計算する。そして、B 群の画像および A 群未選択画像の特徴ベクトルと各クラスタの平均ベクトルの距離を A 群画像のクラスタリングの時と同様に計算し、どれかのクラスタとの距離がある一定値 t より小さい画像をピックアップする。そして、そのピックアップされた画像を最も小さい距離のクラスタに加えて、最終的な目的画像として出力する。

B 群収集画像の集合 B から選択された画像集合 B_{sel} 、および A 群未選択画像の集合 A_{sel} から選択された画像集合 A'_{sel} は、それぞれ次の様に表すことができる。

$$B_{\text{sel}} = \{b_i \in B \mid \min_{1 \leq j \leq n'} d(b_i, \text{avg}(C_j)) < t\} \quad (8.6)$$

$$A'_{\text{sel}} = \{a_i \in A_{\text{sel}} \mid \min_{1 \leq j \leq n'} d(a_i, \text{avg}(C_j)) < t\} \quad (8.7)$$

ただし、 $d(a_i, a_j)$ は画像 a_i と a_j の特徴ベクトル間の距離、 $\text{avg}(C_j)$ はクラスタ C_j に属する全画像の特徴ベクトルの平均ベクトルを表す。

ここでは、A 群未選択画像からも B 群からと同様に画像選択を行っている。なぜなら、A 群選択のためのクラスタリングを FN 法で行い、クラスタの大きさを距離の閾値で定めているために、クラスタの分布領域のごく近傍の画像でも、大きなクラスタに入れずに小さなクラスタしか形成できない画像が存在することがある。そうした画像を救うのが、A 群未選択画像からの画像選択である。

以上の処理によって、ユーザが最初に入力したキーワードに対応する目的画像が最終結果として得られる。

8.5 実験

8.5.1 画像収集実験

プロトタイプシステムを C 及び Perl で実装し, Linux 2.2.14 の動作する AT 互換機 (Athlon 750Mhz, 384MB) 上で実験を行った. 実験では画像収集時のレスポンスが比較的よいと予想される .jp ドメインのサイトに限定して, 画像収集を行った. 扱う画像ファイルは, WWW では最も標準的な画像フォーマットである JPEG フォーマットの画像のみとした. 実験は表 8.2 に示した 8 つのキーワードについて行った. 以下では, 「ライオン」のキーワードの場合について, 各段階での処理の結果を説明する.

キーワードに関係する Web ページの URL を取得するための検索エンジンとして, いずれも日本向けの日本語テキスト検索エンジンである Google 日本語版 [184], Goo[185], Infoseek Japan[186], Lycos Japan[187], OCN Navi[188], Excite Japan[189] の 6 つの検索エンジンを用いた. その 6 つのテキスト検索エンジンから, 合計 2000 件の URL を獲得し, その検索結果の URL を重複を除いて統合した. 「ライオン」の画像収集においては, 検索キーワードは「ライオン 動物」とした. その結果, 6 つの検索エンジンから 301 件の重複を除いて 1699 件の URL を獲得した. すべての URL 獲得にかかった時間は, 実験時には 70 秒程度であった.

次に, 収集プロセスの数を 120 として, 獲得した URL に基づく HTML 文書データの取得, 解析を行った. その結果, 1979 個の HTML 文書を取得し, 全部で 1364 件の JPEG 形式の画像ファイルの URL を抽出した. このうち, A 群には 72 枚, B 群には 216 枚が分類され, 合計 288 枚の画像を収集した. 実行時間は 135 秒掛った. なお, 取得した HTML 文書の数が検索エンジンから取得した URL の 1699 件を越えるのは, HTML 文書中に画像ファイルへのリンクが無い場合に HTML 文書へのリンクをたどって, その HTML 文書も取得したからである.

収集した A 群収集画像 72 枚, B 群収集画像 216 枚について, それぞれ主観的な評価によって「ライオン 動物」に適合しているかどうか調べた結果, A 群画像 61 枚, B 群画像 57 枚が正解画像と判断された. 適合率を $(A \text{ 群収集画像中の正解画像の枚数}) / (A \text{ 群収集画像の枚数})$ とすると, A 群収集画像の適合率は 85% となった. 同様に計算して, B 群画像, 収集した全画像 (A 群+B 群) の適合率はそれぞれ 26%, 41% となった. なお, 収集画像の再現率は, $(\text{正解画像の枚数}) / (\text{WWW 空間中の正解画像の枚数})$ となり分母が測定不可能なので, 収集画像に対しては再現率は用いないこととする.

解析部では, まず, 収集した A 群の 72 枚の画像についてカラーヒストグラムによる画像特徴量を求めて, 互いの距離を計算し, クラスタリングの処理を行った. その結果, メンバ数が A 群収集枚数の 5% 以上²のクラスタだけ残すと, 9 クラスタ, 合計 60 枚の画像が残った. 実験では, 収集枚数が多いとそれだけノイズ画像の枚数も多くなり, 収集枚数が少ない場合に比べてメンバ数の大きいノイズクラスタができる傾向があったため, クラスタ選択の閾値として, 「A 群収集枚数の 5% 以

²実際には, $\lfloor (A \text{ 群収集枚数}) \times 0.05 \rfloor$ 枚以上. $\lfloor x \rfloor$ は x を越えない最大の整数を表す. この場合は, 3 枚以上ということになる.

表 8.2 実験結果. A 群, B 群, 全体 (A 群+B 群) それぞれの収集枚数, 選択枚数を示す. 収集枚数の () 内はキーワードのみによる収集結果の適合率 (%). 選択枚数の () 内は画像特徴を利用した処理結果の適合率 (%) と再現率 (%).

検索エンジン用 キーワード	HTML 文書数	A 群画像枚数		B 群画像枚数		全画像枚数 (A 群+B 群)	
		収集枚数	選択枚数	収集枚数	選択枚数	収集枚数	選択枚数
ライオン 動物	1979	72 (85)	62 (94,95)	216 (26)	66 (42,49)	288 (41)	128 (67,73)
りんご	2054	97 (86)	76 (95,87)	237 (50)	99 (72,60)	334 (60)	175 (82,71)
赤ちゃん	2031	85 (48)	73 (53,95)	528 (74)	272 (83,58)	613 (70)	345 (77,62)
机	2112	76 (90)	72 (92,97)	212 (50)	84 (71,56)	288 (60)	156 (81,72)
キーボード パソコン	2194	39 (95)	38 (95,97)	167 (60)	58 (73,43)	206 (67)	96 (82,57)
トラ 動物	2006	57 (71)	51 (75,95)	178 (33)	71 (42,50)	235 (42)	122 (56,69)
野茂 大リーグ	1778	38 (95)	34 (97,92)	28 (25)	14 (36,72)	66 (65)	48 (79,88)
富士山	1981	541 (71)	317 (91,75)	837 (42)	158 (66,30)	1378 (53)	475 (82,53)

上」³という相対枚数を用いた. この 60 枚の画像のうち主観的な評価による正解画像は 56 枚で, 適合率を (A 群選択画像中の正解画像の枚数) / (A 群選択画像の枚数), 再現率を (A 群選択画像中の正解画像の枚数) / (A 群収集画像中の正解画像の枚数) とすると, 適合率は 93%, 再現率は 92% となった. なお, 本来は再現率の計算における分母は (WWW 空間中の正解画像の枚数) となるべきであるが, 測定不可能であるので, ここでは, 画像収集後の画像選択処理だけの評価のために (A 群収集画像 (もしくは B 群収集画像, 収集画像全体) 中の正解画像の枚数) を分母とする再現率を用いることとする. 付録 A の図 8.6 に A 群より選択されたライオンの画像の一部を示す. 図中には上下にメンバー数の上位 2 つのクラスタの画像を示している.

最後に, B 群の画像および A 群未選択画像から, A 群で残った 9 つのクラスタのどれかに近い画像を選んだ結果, B 群の画像 216 枚から 66 枚の画像, A 群未選択画像 12 枚からは 2 枚が選択された. B 群選択画像 66 枚の適合率, 再現率を A 群選択画像の時と同様に求めると, それぞれ 42%, 49%であった. また, A 群未選択画像からの再選択によって選ばれた画像 2 枚は 2 枚とも「ライオン」の画像であり, 再選択画像を含めて A 群選択画像の適合率, 再現率を再計算すると 94%, 95% となった. 図 8.7 に B 群より選択された画像を示す. 図中には上下に 2 つのグループに近かった画像が示されており, 図 8.6 の上下のグループに対応している.

以上の解析部の処理によって, A 群, B 群から選択された画像の合計枚数が 128 枚でその適合率, 再現率は最終的に 67%, 73%となった. なお, 解析部の実行時間は全部で 70 秒程度掛り, 全体での処理時間としては合計 5 分弱程度になった.

参考のために, 実際の処理ではすべては収集しない, HTML 文書から抽出した画像ファイルの URL1364 件の指す画像を収集して, 主観評価を行った結果, そのうち正解画像の枚数は 138 枚で, 適合率は 10%であった. 1364 枚の画像に対する収集画像, 選択画像の再現率を (収集 (選択) 画像

³この 5%という値は事前に行った予備実験で平均的に良い結果が得られた値である.

中の正解画像枚数)/(全画像 URL 中の正解画像枚数) として求めると、それぞれ 86%, 62% となった。このことから、本システムにおける画像収集の決定のためのキーワード評価が有効に機能していることが分かる。

表 8.2 に「ライオン」を含めた 8 つのキーワードに対する実験結果を示す。獲得し解析した HTML 文書の数、A 群、B 群、全画像 (A 群+B 群) のそれぞれの収集枚数、選択枚数が示されており、収集枚数の後の括弧内はキーワードのみによる収集結果の適合率、選択枚数の後の括弧内は画像特徴を利用した選択処理結果の適合率と再現率をそれぞれ表している。なお、表中では、A 群未選択画像からの再選択画像の枚数は、A 群画像選択枚数の中に含めている。参考までに、図 8.8, 図 8.9 に「富士山」の A 群選択画像、B 群選択画像を示す。

図 8.10 は、表 8.2 中に記された 8 つのキーワードについての A 群画像と B 群画像を合計した選択画像枚数と、収集画像および選択画像の適合率を表したグラフとなっている。図 8.11 は、8 つのキーワードについての A 群画像、B 群画像、全画像の適合率、再現率をそれぞれ別々にレーダーグラフで表したものである。レーダーグラフ中の 3 つの線は、選択された画像の適合率および再現率、キーワードのみによって収集した画像の適合率をそれぞれ表わしている。

8.5.2 選択条件を変化させた時の実験結果

上記の実験と同様に「ライオン 動物」で画像収集を行い、A 群収集画像から目的画像を選択する時の選択されるクラスタの条件であるクラスタのメンバー数の下限値 m を変化させた時の実験結果を図 8.12 に示す。図 8.12 は、 m を 1 から 10 まで変化させた時の最終的に得られた目的画像の適合率、再現率を縦軸に適合率、横軸に再現率をとって表したグラフで、 m が小さい時ほど再現率が高く、グラフの右の方になり、 m が大きいほど再現率が低くなって、グラフの左の方になる。適合率の変化は全体的には大きくないものの、初めは m を大きくしていくにつれ、ノイズ画像が除去されるため、適合率は上昇し、 $m = 4$ の時に最大になる。しかし、さらに m を大きくすると、正解画像のクラスタまでもが除去されてしまい、適合率は減少する傾向を示している。

次に、B 群から目的画像として選択する時の条件である、正解クラスタとの距離の閾値 t を段階的に変化させた時のグラフを図 8.13 に示す。なお、最初の収集実験と同様、クラスタの選択の条件は $m = 3$ とした。このグラフも同様に最終的な目的画像の適合率と再現率を表しており、グラフの左上が t が小さい時に対応し、グラフの右下の方が t が大きいときに対応している。このことから、B 群の選択条件の閾値は、最終収集結果に大きな影響を与えることが分かる。

8.6 考察

本節では図 8.10 および図 8.11 に基づいて実験結果の考察を行う。

A 群画像については「赤ちゃん」以外はクラスタリングによるノイズ画像の除去後の選択画像の適合率がほぼ 9 割以上になっており、いずれの実験例の適合率も選択画像の方がキーワード評価によって収集した収集画像よりも上回っている。再現率はほぼ 9 割近くになっているが、これは A 群選択が基本的にノイズの除去で、A 群収集画像のうちの 8~9 割程度の枚数を選択しているため



図 8.6 A 群から選択された「ライオン」の目的画像の例.

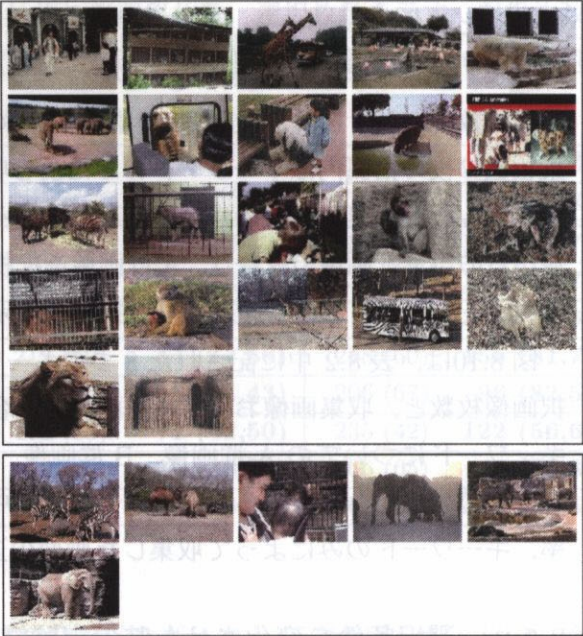


図 8.7 B 群から選択された「ライオン」の目的画像の例.

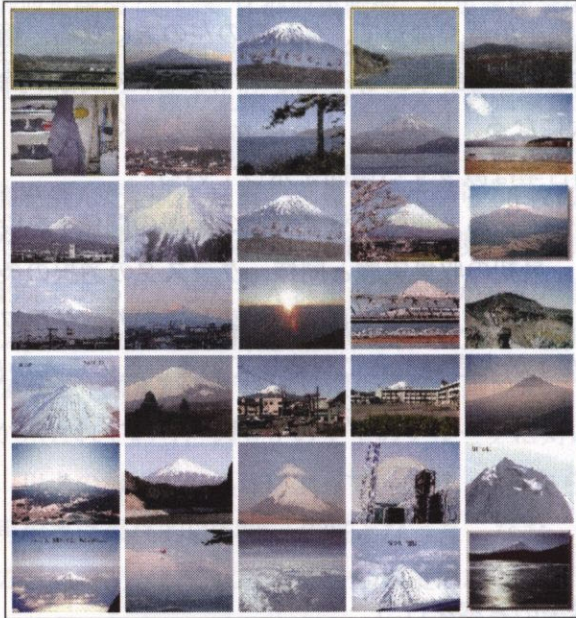


図 8.8 A 群から選択された「富士山」の目的画像の例.

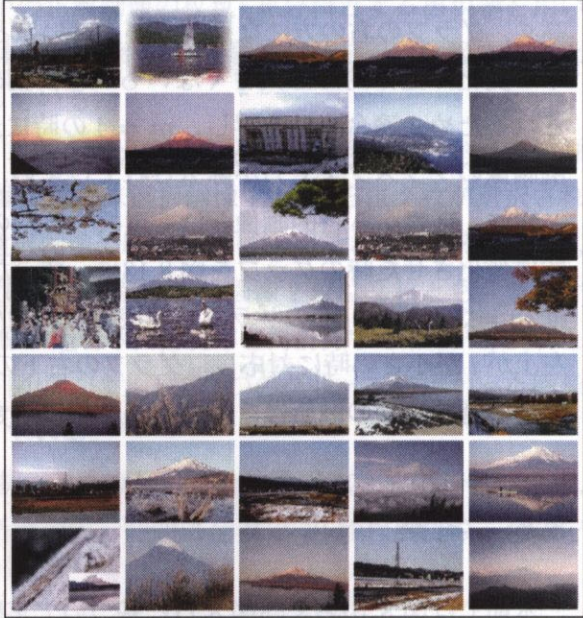


図 8.9 B 群から選択された「富士山」の目的画像の例.

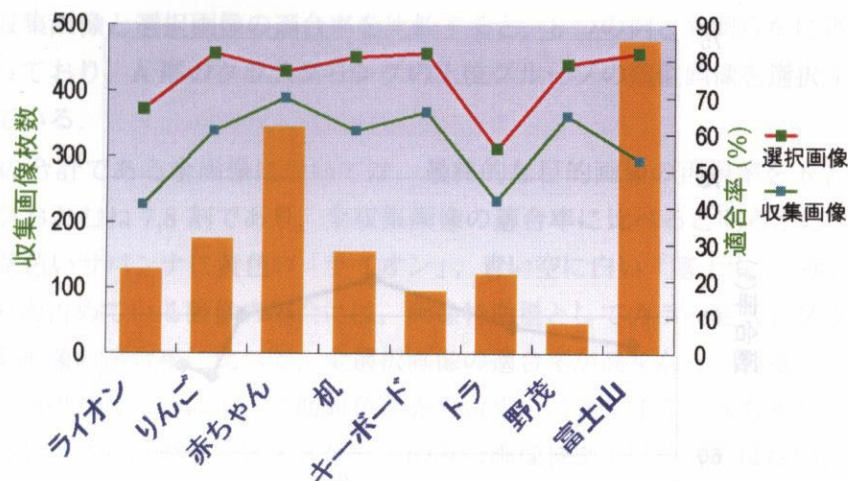


図 8.10 全画像の選択画像枚数，収集画像および選択画像の適合率。

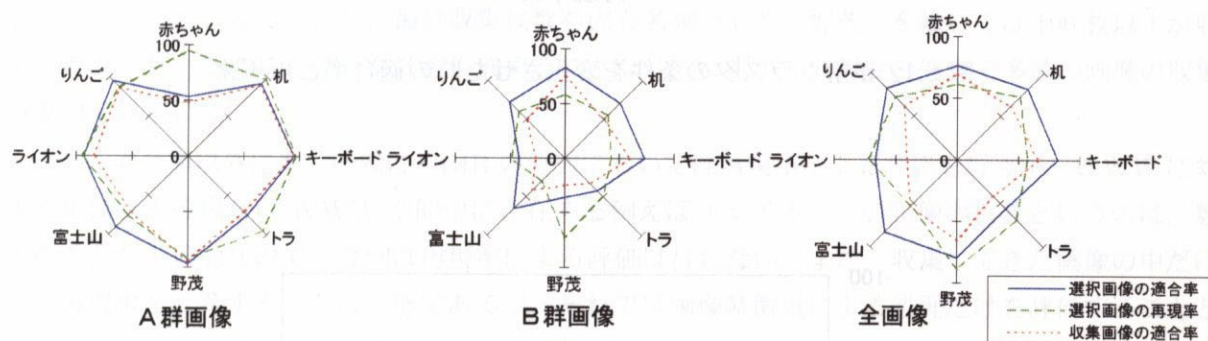


図 8.11 選択画像の適合率，再現率および収集画像の適合率の比較グラフ。

ある。また，選択画像の適合率がほとんどの場合，9割以上にもなっているのは，キーワードによって A 群画像の選ぶ段階でのヒューリスティクスが非常に有効に働いているからである。特にファイル名は有効で，ファイル名が「lion」になっている画像のほとんどすべては「ライオン」の画像である。例外的に「赤ちゃん」の適合率が低いのは，ベビー用品メーカーのカタログに大量にヒットしてしまったからである。また，「富士山」では，富士山だけ 541 枚（うち正解画像 384 枚）も A 群画像が収集され，クラスタの条件の収集枚数の 5% の値が，27 枚以上と大きくなってしまったため，多くのクラスタが除去されてしまい，選択された画像は 201 枚（うち正解画像 185 枚）だけとなってしまう。そのままであると，A 群の再現率は 48% であるが，その後の A 群未選択画像の再選択の処理によって，さらに 116 枚（うち正解画像 102 枚）が追加されて，再現率は 75% に上昇した。

B 群画像については，キーワードの条件を A 群より緩くしているためにキーワードのみによって収集した収集画像の適合率が 3～6 割程度と低くなっている。選択画像の方は，適合率が高いと再現率が低くなり，適合率が低いと再現率が高くなるという傾向が見える。適合率と再現率の balan

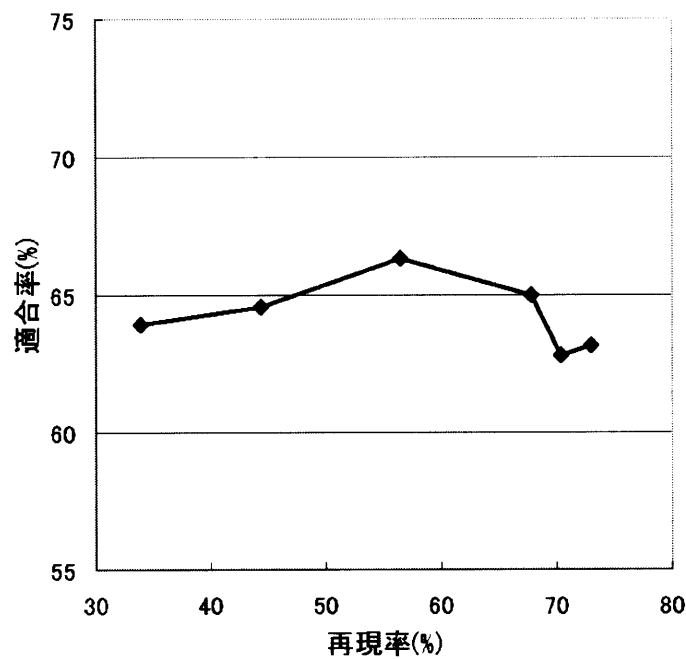


図 8.12 正解クラスタの条件を変化させた時の適合率と再現率.

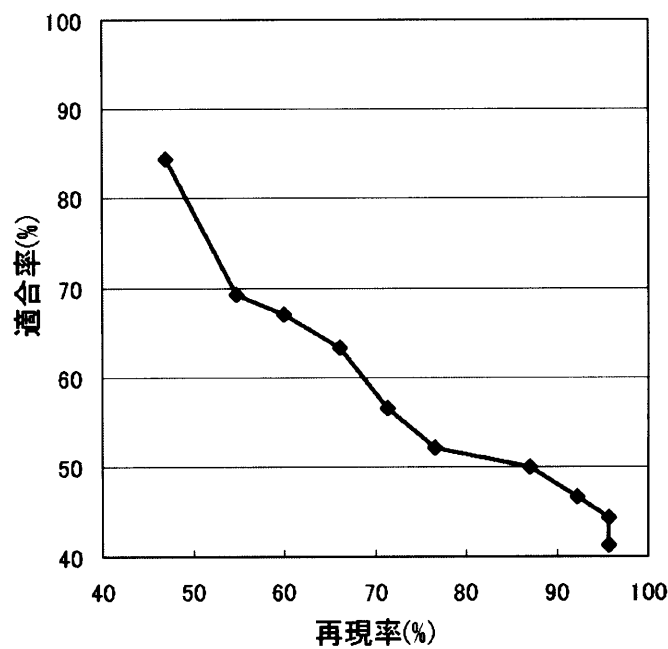


図 8.13 B 群選択の閾値を変化させた時の適合率と再現率.

スは、図 8.13の実験結果で示したように B 群選択条件の閾値を変化させることによって調節可能である。次に、収集画像と選択画像の適合率を比較すると、8つの例とも明らかに選択画像の適合率の方が高くなっており、A 群のクラスタリングの上位グループの類似画像を選択するという手法が有効に作用している。

A 群、B 群の合計である全画像については、最終的な目的画像の再現率を 6,7 割程度に保ちながら、適合率がおおむね 7,8 割であり、全収集画像の適合率に比べると明らかに上回っている (図 8.10)。特に、茶色いサバンナに黄色い「ライオン」、青い空に白い「富士山」、赤い「りんご」の様に色が重要な要素占めている画像の場合には、画像特徴量としてカラーヒストグラムを用いているために、全収集画像の適合率に比べて、全選択画像の適合率が高くなっている。一方、様々な色の洋服を着て、光の回り具合によって顔の色が赤や黄や白に変化する「赤ちゃん」の場合は、適合率の向上の割合が小さく、カラーヒストグラムのみ画像特徴では十分ではない。したがって、今後は、エッジやテクスチャなどの色以外の画像特徴も利用するようにすると、色のみでは十分でなかったキーワードについても選択画像の適合率の向上が望められると思われる。

以上のことから、本章での提案手法である、ユーザーによる処理途中での指示が不要な方法であるクラスタリングの上位グループの選択および上位グループの類似画像の選択という手法の有効性が示せているといえる。また、最終収集枚数も固有名詞である「野茂」を除いては 100 枚以上が収集できており、予めインデックスを作成することなく、わずか 5 分程度の処理で多数の画像の収集が実現できている。

なお、適合率以外にも、一般に CBIR の研究においては再現率による評価を行うが、WWW に対する画像収集の場合は WWW 空間中に存在する例えば「ライオン」の画像の枚数というのは、数えることが不可能なので、今回は再現率による評価は行わない。また、収集してきた画像の中だけで、再現率を計算することは可能であるが、それでは画像特徴量による処理だけの評価することとなってしまう、システム全体の評価としては適当でない。

8.7 まとめ

本章では、キーワードと画像特徴を利用した WWW(World-Wide Web)からの自動画像収集のためのシステム Image Collector の実装、および実験結果について述べた。今回の実験結果より、テキスト検索エンジンを利用して画像を WWW より収集し、さらにユーザによる処理途中での指示が不要な方法であるクラスタリングの上位グループの選択およびその類似画像の選択という手法を用いた提案システムの有効性が示せたといえる。

現段階では、B 群として収集される画像の適合率が高いとはいえないので、B 群の分類基準を検討することが必要である。今後は、A 群と C 群の間に 1 つ設けている B 群をキーワードとの関係の強さに応じて数段階に分けて、選択時の距離の閾値を段階的に変化させることなども検討している。また、画像特徴量についても、現在カラーヒストグラムのみ利用であり、CBIR の分野で提案されている様々な方法を今後取り入れていることが必要である。

本システムでは、クラスタ分析と B 群から画像を選択する際において、距離の閾値が用いられて

いるが、この閾値の決め方についても今後検討が必要である。今後は、こうした様々な課題を検討して、最終的な目的画像の適合率、再現率を向上させていく予定である。

また、現在は収集部と解析部が別々になっており、5 分程度の処理時間が掛っているが、収集、解析を統合して並列化することによって、今後、処理時間の短縮を図る予定である [11]。