

実世界画像自動分類のための Web 画像マイニング

Web image mining for automatic classification of real world images

柳井 啓司
Keiji Yanai

電気通信大学 情報工学科

Department of Computer Science, The University of Electro-Communications

Thanks to the recent rapid spread of digital imaging devices, the demand for generic image recognition of various kinds of scenes becomes greater. It is, however, hard to collect various kinds of training images for recognition of various kinds of scenes so far. Then, we propose a generic image classification system with an automatic knowledge acquisition mechanism from the World Wide Web. We call this knowledge acquisition on Web images "Web image mining". The system gathers a large number of images related to given class keywords from the Web and classifies an unknown image into one of the classes corresponding to the class keywords using gathered images as training ones. By the experiments, we show that utilizing of Web images as training images, which are as diverse as real world images, is effective for generic image classification.

1. はじめに

デジタルカメラなどの画像入力機器の普及や記憶装置の大容量化によって、日常の実世界シーンをデジタル画像として計算機に取り込み、大量に蓄積することが容易になってきている。しかし、その一方で、日常の実世界シーンの画像が極めて多種多様に及ぶものであるために、画像の意味内容を自動的に認識することは現状では難しい。今後は、制約のない日常シーンの画像に対する自動キーワード付与や、画像内容による分類や検索などの画像の意味内容を扱う処理の実現が望まれている。

現在までに、実世界シーンの画像データベース中の画像に対して、自動的にキーワードを付けたり [森 01, Barnard 01, Duygulu 02], 意味的に類似している画像を検索する研究 [Belongie 97] が行われている。これらの研究では、事前に単語と画像の対応付けを学習する必要があるため、学習データとして予めキーワードが付与されている画像が必要である。そのためこれらの研究では、各画像にキーワードが予め付けられている市販の画像コレクションが学習用データとして用いられている。しかしながら、一般に市販の画像コレクションや商用の画像データベースは、プロの写真家が撮影した整った画像のみを集めていて、しかも、同じカテゴリに含まれる画像は同一のカメラマンが似たような構図で撮影した場合が多く、我々の日常シーンなどの“一般の画像”とは掛け離れているといえる。

我々は World Wide Web から画像を自動収集し、それを学習画像として用いて、一般の実世界画像に対する自動分類を行うことを提案する。Web 上の画像 (Web 画像) は、様々な人が様々な目的で撮影した画像であり、類似している画像が多く含まれる商用の画像データベースとは異なり、実世界の一般的な画像の多様性をそのまま反映していると考えられる。また、Web 画像はそれを含んでいる Web ページの HTML 文書を解析することによって、画像に関連するキーワードを抽出することが可能であるという特徴を持つために、目的の画像を自動収集することが可能である。

人工知能研究においては、システムに大量の知識を人手で与えることに限界があるという知識獲得のボトルネックの問題を解決するために、近年、Web からの知識獲得 (Web マイニング) の研究が盛んに行われている。本研究もそれと同様に、実世界画像とその意味内容との対応の知識 (ここでは画像知識と呼ぶ) を Web から自動的に獲得し、その知識を実世界画像分類や自動キーワード付与などに応用することを目的とする。ただし、従来からの Web からの知識獲得の研究では主にテキストデータのみを対象にしていたのに対して、本研究ではテキストデータに加えて画像データを主な対象にしている。我々は、こうした Web からの画像知識の獲得を「Web 画像マイニング」と呼ぶこととする。

と称することとする。

本研究では、Web 画像収集システムと画像分類システムを組合せることによって、Web 画像マイニングによる一般画像の自動分類の実験を行い、その可能性について論ずる。

具体的には、まず最初に、Web からの自動画像収集システムを用いて、数種類の単語に関係する画像を収集する。そして、収集画像を学習画像として、それぞれの画像から抽出した画像特徴量を比較し、テスト画像を学習画像の各クラスのうち最も類似しているクラスに分類し、画像の自動的なクラス分類を試みる。画像特徴は対象に依存しない色、テクスチャなどの一般的なものを利用し、基本的には画像データベースの検索手法である意味内容に基づく画像検索 (Content-Based Image Retrieval, CBIR) と同様の方法で画像分類を行う。実験では、ブロック分割と色特徴を用いた方法と、領域分割と色、テクスチャ、領域形状を用いた 2 通りの方法で分類を行う。画像特徴間の距離 (非類似度) としては、2 つの離散分布間の距離を求めることの出来る Earth Mover's Distance [Rubner 00] を用いた。

本稿では、まず全体の構想について述べ、次に画像分類を中心に実験方法について触れ、最後に実験結果と考察を述べる。

2. 構想

本研究では、Web から画像を収集し、画像特徴量に基づく類似画像検索による未知画像のクラス分類を試みる (図 1)。

Web 上に存在する画像は、それを含んでいる Web ページの HTML 文書を解析することによって、画像に関連するキーワードを抽出することが可能であるという特徴を持つ。そのため、テキスト検索エンジンを用いることによって、キーワードに関連する画像のみを収集することが可能である [柳井 01, Yanai 03]。また、Google Image Search や Ditto などの画像検索エンジンを利用することも可能である。しかしながら、Web からの画像収集の精度は 100%ではなく、必ず不必要な画像が含まれるという特徴がある。これを 100%にするには人手によって、不適切な画像を除去するしかないが、枚数が数万、数十万というように増えるにしたがって人手による作業は困難なものとなる。したがって、従来の画像認識/分類の学習画像は 100%正しいということが前提となっていたが、Web 画像マイニングを用いた画像認識/分類では、学習画像に誤りが含まれていても学習が可能であることが望まれる。

WWW 上には商用画像検索エンジン、例えば Google Image Search, Ditto, Lycos Multimedia Search, AltaVista Image Search などが存在しているが、これは主にキーワード解析のみで画像を収集しており、画像の内容に関する特徴量解析を行っていない。一方、画像特徴量解析を用いる Web 画像検索システム [Framkel 96, Smith 97, Sclaroff 99] も提案されてい

連絡先: 柳井 啓司 電気通信大学情報工学科 〒182-8585
東京都調布市調布ヶ丘 1-5-1 E-mail: yanai@cs.uec.ac.jp

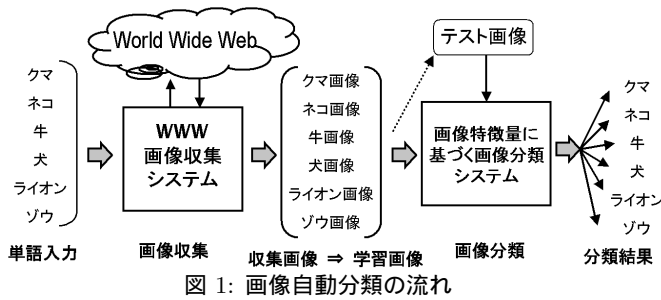


図 1: 画像自動分類の流れ

るが、これらのシステムでは商用検索エンジンと同様に大量の Web データを予め収集して巨大なインデックスを作成する必要があり、実験レベルでの運用しか行われていない。また、画像特徴量に基づく類似画像検索の処理のために、検索途中でユーザが目的とする画像を選択する必要がありインタラクティブな検索となっている。そのため大量な画像の検索には不向きで、ユーザに手間がかかる。他にもリンク構造を考慮した画像検索手法 [Lempel 02] も提案されているが、いずれも少量の画像を「検索」するのに適した方法であり、Web 画像マイニングに必要な大量の画像を「収集」するには不向きである。

本研究においては、我々が先に提案した大量の画像を Web から収集するシステム Image Collector [柳井 01, Yanai 03] を用いて、画像収集を行うこととする。

次に画像分類の方法であるが、これは本研究においては基本的に既存の方法を利用する。まず、学習画像を領域分割（もしくはブロック分割）し、各部分領域について画像特徴量を計算する。次に、未知画像（テスト画像）の各部分領域との類似度を計算する。そして、多くの類似領域が存在したクラスに分類することとする。つまり、これは Nearest Neighbor 法 (NN 法) によって分類するということと等価である。Web から収集された画像は、同一クラスでもあっても様々な画像が含まれており、特徴空間中に広く分布するために、本研究では分類器 (classifier) としては最も単純であるが、どのような分布に対しても平均的な性能を発揮する Nearest Neighbor 法 (NN 法) を用いて分類することとする。

本研究では、「同じクラスに分類される画像は画像全体の見た目が類似している」という仮定に基づいている。つまり、本研究で用いる画像分類の方法は、画像中に写っている対象物体以外の物体も対象物体に密接に関係したものであるため、対象物体を切り出さずにそれも含めて分類の手がかりとして利用しようという方法である。例えば、牛なら草原や牧場の柵なども多くの画像で同時に写っているし、ライオンならサバンナが動物園の風景と一緒に写っている。その場合、対象物体と他の物体の組み合わせは多くなるが、「Web から収集した大量の画像を学習画像とすることにより、実世界のバリエーションの豊富さを補うことが可能である」ということを仮定する。本研究では、以上のような仮定に基づいて、予めクラスが既知である学習画像を大量に収集し、次にクラスが未知である一般画像と類似している学習画像を検索し、その学習画像のクラスにクラス未知の画像を分類することによって、自動画像分類を実現する。

3. 方法

3.1 画像収集の方法

Web から画像収集は、画像収集システム Image Collector [柳井 01] を利用することによって行う。このシステムでは「Web 上の画像はその内容を説明する HTML 文書とともに存在している」！キーワードに関連のある画像のうち、類似している画像が多く存在する一般的な画像の方がキーワードとの関連性がより強い」という 2 つの経験則に基づいて、HTML 文書に対するキーワード解析と、画像に対する特徴量解析を行い、Web から画像を収集する。特徴としては、テキスト検索エンジンを利用するために (1) 事前に大量なデータを蓄積することなく、さらに、クラスタリングを利用した自動画像選択を

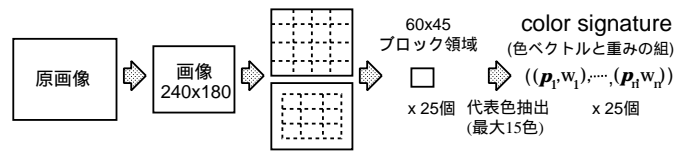


図 2: 分類方法 1: color signature の作り方

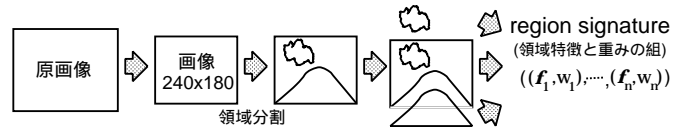


図 3: 分類方法 2: region signature の作り方

行うことによって (2) 処理途中でのユーザのインタラクティブな処理が不要でユーザがキーワードを指定するだけで、そのキーワードを表す画像を数百～千枚単位で Web から自動的に検索し収集することが可能であるということがあげられる。

3.2 自動分類の方法

収集した画像から画像特徴量を抽出し、類似画像検索によりクラス分類を行う。画像特徴は単純な多次元ベクトルで表現するのではなく、多次元ベクトル p_i とその重み w_i の組 (p_i, w_i) の集合の組の集合である signature によって表現する。実験では、signature を構成する特徴として、色情報のみを用いた color signature と、領域分割した後の領域の色、テクスチャ、形状の情報を用いた region signature の 2 通りの signature を画像特徴として用いる。Color signature の場合、 p_i は色を表す代表色ベクトルとなり、 w_i はその色の画素数の割合、region signature の場合、 p_i は各領域の特徴を表す特徴ベクトル、 w_i は画像全体に対する領域の面積比となる。

Signature 同士の間隔を求めるには、Earth Mover's Distance (EMD) [Rubner 00] を用いる。EMD は線形計画問題の一つであるヒッチコック型の輸送問題の解に基づいて計算される。一方の signature を「土 (earth)」の供給地の集合とそれぞれの供給量、もう一方の signature を「穴」を埋めるための「土」の需要地の集合とそれぞれの必要量とみ立てて、供給地と需要地間の輸送コストが与えられた時に、需要地の需要を満たすように供給地から需要地に「土」を輸送する最小輸送コストが EMD による距離となる。

従来のヒストグラムによる色分布表現にはヒストグラムの分割 (bin) が固定であるために表現力が不十分で、さらに距離計算にユークリッド距離を用いると人間の感覚に合わないことがあるという欠点があった。Signature の場合は対象に応じて要素数を変えることができ、柔軟な表現が可能である。また、EMD によって距離計算を行うことによって、signature の要素間の距離を全体の距離に反映させることができ、より人間の感覚に近い「距離」の計算を行うことが出来る。

3.3 分類方法 1: ブロック分割と color signature

方法 1 では、画像を初めにブロック分割し、それぞれのブロックについて color signature を抽出する (図 2)。まず、画像を 240×180 の大きさに正規化し、縦横を 4 分割と半ブロック分ずらして 3 分割し、合計 25 個のブロック画像を生成する。次に、各ブロック画像の色分布を k -means 法によるクラスタリングによって最大 15 色に減色し、各クラスタの代表色 (平均色) を色ベクトル、ブロック画像の全画素数に対するクラスタの要素数の割合を重みとして、color signature を構成する。クラスタ数 k はすべてのクラスタの重み値が 0.005 以上となり、かつ $k \leq 15$ を満たす最大の k とすることとした。なお、色ベクトルは Lu^*v^* 色空間の 3 次元ベクトルで表現した。

クラス分類の方法としては、予め、学習画像、テスト画像共に 1 枚の画像あたり 25 個の color signature を抽出しておいて、次に学習画像のある 1 つのブロックとテスト画像の全ブロックの間の最小距離をクラス毎に求める。これを 25 回繰り返して、学習画像の各クラス毎にテスト画像の全ブロックとの最小距離を合計して、その中で最も小さい値となったクラスに

そのテスト画像を分類することとする。

3.4 分類方法 2: 領域分割と region signature

方法 2 では、画像を領域分割して得られた各領域の画像特徴量と面積比を要素とする region signature を用いて、EMD による画像間の距離計算を行う。

ここでは、 k -means 法によるクラスタリングによって、領域分割を実現する [Wang 01]。具体的には、画像を 4×4 の小ブロックに分割し、それぞれのブロックの特徴量を 3 次元の色ベクトルと、3 次元のテクスチャベクトルの合計 6 次元のベクトルで表現する。3 次元の色ベクトルとしては、 Lu^*v^* 色空間上での 4×4 の小ブロックの平均色ベクトル、3 次元のテクスチャベクトルとしては、 4×4 の小ブロックに対して Daubechies-4 基底を用いたウェーブレット変換を施して得られた 2×2 の HL 成分、LH 成分、HH 成分それぞれの 2 乗平均の値を用いる。そして、6 次元の特徴ベクトルを持つ各小ブロックを k -means 法によるクラスタリングすることによって、領域分割を行う。 k の決め方は方法 1 と同様である。

次に領域分割された各領域について、平均特徴ベクトルを求め、さらに各領域の 1 次から 3 次までの重心周りのモーメントを求め、合わせて領域を表現する 9 次元特徴ベクトルとする。そして、各領域の特徴ベクトルとその領域の画像全体に対する面積比を組として、それを要素とする signature を各画像について構築する。方法 1 では、ブロック分割された各ブロック画像がそれぞれ signature を持っていたが、ここでは画像 1 枚について 1 つの signature で表現する。したがって、1 枚の画像に対する特徴データの量は大幅に削減され、処理の高速化が実現できる。

こうして学習画像、テスト画像方法について signature を求めて、学習画像とテスト画像間の距離を EMD によって計算する。距離が小さい上位 k 個のうち、最も多かったクラスに分類するという k -NN 法によって、テスト画像の分類を実現する。実験においては、 k の値として予備実験より求めた値 5 を用いることとする。

4. 実験結果 及び 考察

表 1 に示す実験 1 から実験 6 までの 6 種類の実験を行った。

表 1: 6 種類の分類実験

実験 no.	クラス数	学習画像 枚数	適合率 (%)	テスト画像	
				枚数	入手元
1	10	4548	68.2	20	Web
2	10	3102	100†	20	Web
3	10	500	100‡	10	Corel
4	20	5694	64.5	50	typical*
5	20	3485	100†	50	typical*
6	50	22725	未調査	100	leave-one-out

†人手で適合画像を選択。 ‡Corel 画像を学習画像として使用。* 各クラス 50 枚の典型画像。

表 2: 実験 1~6 の結果

実験 no.	方法 1			方法 2		
	再現率	適合率	F 値	再現率	適合率	F 値
1	31.8	37.3	34.3	29.4	30.3	29.8
2	36.9	45.5	40.8	35.4	37.2	36.2
3	62.5	67.4	64.9	67.1	69.2	68.1
4	39.8	50.7	44.6	40.1	43.1	41.5
5	38.5	60.4	47.0	42.1	47.9	44.8
6	34.2	49.0	40.3	27.6	28.6	28.1

実験 1 では、初めに「クマ」「ネコ」「イヌ」「ゾウ」「熱帯魚」「ライオン」「ペンギン」「ヒツジ」「トラ」「クジラ」の 10 個の動物に關係した単語についての画像を収集した。収集時には、5 つの日本語テキストサーチエンジンからそれぞれの単語について約 1 万件の URL を収集した。最終的に収集された画像の主観評価による適合率 ((単語の表す対象が含まれる画像の枚数) / (収集した画像の枚数) として計算) は平均で 68.2%、合計枚数は 4548 枚であった (表 2)。例えば「ライオン」の場合、9916 件の URL から、5497 個の JPEG 画像のファイル名を抽出し、HTML ファイルの解析、画像解析によって、最終的に合計 504 枚の画像を収集し、その適合率は 77.0%であった。

さらに実験 1 では、収集した画像のうち各単語について 20 枚ずつをテスト画像として、残りの画像をすべて学習画像として、交差法 (cross validation) による分類実験を行った。結果を表 2 に、再現率 ((そのクラスに正しく分類された画像の枚数) / (テストに用いたそのクラスの画像の枚数))、適合率 ((そのクラスに正しく分類された画像の枚数) / (そのクラスに分類された画像の全枚数))、およびそれら 2 つの値の調和平均である F 値 (F-measure) を示す。なお、値はすべて % 表示である。以下では、主に F 値を用いて実験結果の評価を行うこととする。完全にランダムな場合は、適合率、再現率の平均値、それぞれの平均値の F 値は $100 / (\text{クラス数})$ となるので、10 クラスの場合 10.0 となる。それに対して、得られた平均 F 値は方法 1 のブロック分割+color signature で 34.3、方法 2 の領域分割+region signature で 29.8 であった。なお、実験 6 までのすべての実験において、方法 1 の方が優れた値を示したので、方法 1 による結果はグラフ (図 4, 図 5) にその詳細を示した。分類時間に関しては、Pentium4(2.0GHz) の計算機上で実行した場合、実験 1 の場合、方法 1 が 1 枚当たり約 0.05 秒なのに対して、方法 2 は 100 倍の約 5 秒も掛っており、実行時間の点では方法 2 の方が優れている。

実験 2 では、実験 1 で収集した画像を手でチェックし、適切でない画像を除去した上で、同様の実験を行った。実験 1 に比べて、方法 1 では平均の F 値が 6.5 だけ向上している。

実験 3 は、Web から収集した画像ではなく、市販の画像コレクションである Corel 社の Corel Image Gallery に含まれる画像を各クラス 50 枚ずつ用いて、分類実験を行った。実験に用いた画像のクラスは、実験 1, 2 と同一である。画像枚数が少ないので、テスト画像にはそのうち 10 枚ずつを抜き出して、交差法による実験を行った。Corel Image Gallery は互いに類似している画像が多いので、実験 2 の 40.8 に比べてかなり高く、平均の F 値は 64.9 であった。従来の研究では、Corel 画像によって CBIR や自動分類の性能評価を行っている場合が多く、一般の画像に比べて Corel 画像が互いに類似している画像が多いという事実を評価の際に考慮に入れる必要がある。

実験 4 では、20 種類の単語に対して画像収集を行い実験を行った。全部で 5694 枚の画像を収集し、その主観評価による適合率は 64.5% であった。実験 5 は、実験 4 での収集画像から人手によって不適切な画像を除去してから分類実験を行った。両実験の結果を表 2 および方法 1 による結果の詳細を図 5 に示す。テスト画像としては、一般の実世界画像に対する分類能力の評価として、Web 収集画像と Corel 画像から人手で各クラスの典型的な画像を 50 枚ずつ選び出し用いた。ここで「典型的」とは、類似している画像が多く存在している画像のことで、各クラスともに数パターンの典型的画像パターンが存在する。本来なら標準的な評価用データセットが存在していれば、それを用いて分類性能の評価を行うべきであるが、一般に実世界画像に関してはそのような評価用データというものは存在しない。そこで、クラスによって異なるが、最大 Corel 画像から 15 枚、残り Web 収集画像から人手によって取出して、各クラスについて 50 枚の評価用画像セットを構築した。図 5 に使用した評価画像のうち各クラス 2 枚ずつの画像を示す。なお、Web 画像について分類を行う場合は、そのテスト画像を学習画像から除く leave-one-out 法によって分類を行った。

平均の F 値は、方法 1 で 44.6、47.0 となっており、クラス数が 10 しかない実験 1, 2 よりもよい結果が得られている。これは、実験 1, 2 の 10 クラスがすべて動物で、異なるクラスでも互いに画像が類似していることが多いのに対して、こちらの 20 クラスは種類がばらばらでクラス間で画像特徴が比較的分離していると考えられる。実験 1, 2 や実験 4, 5 の個々のクラスの分類結果は、そのクラスの画像のまとまりの良さを表していると考えられる。つまり、F 値が高いほど、そのクラス内で画像特徴の分布がまとまっているということである。

実験 6 では、50 個のクラスの画像、合計 33950 枚を Web より収集した。50 個のクラスは、10 個のジャンルについてそれぞれ 5 種類ずつクラスキーワードを選んだ。分類実験では、このうち各クラス最大 600 枚までを用いて、合計 22725 枚の

画像を学習画像として利用した。収集画像の中の収集時の評価の高かった各クラス上位 100 枚の画像合計 5000 枚をテスト画像として用いた。評価はテスト画像 1 枚以外をすべて学習画像として、それを 5000 枚分繰り返す leave-one-out 法によって行った。なお、50 クラスの場合は、収集枚数が多過ぎて、第実験 1~5 のように人手で適合画像のみを選びだすことは困難であったので、適合画像のみの実験は行っていない。つまり、ここでの実験は、不適合画像が含まれている収集画像を学習にも評価にも用いていることになる。F 値の平均は、方法 1 による結果の方がよく、40.3 であった。これは、同様の実験である 10 クラスの場合の実験 1 の結果 34.3 よりよく、20 クラスの場合の実験 4 の 44.6 を若干下回る程度であった。このことから、本研究で行った数十クラス程度のカテゴリ分類の場合、クラスの数よりも、分類対象クラスの中でどの程度互いにその画像が視覚的に類似しているクラスがあるかどうかの方が、分類精度に影響するといえる。つまり、多様な 50 クラスを分類するよりも、似たような画像、例えば動物ばかり 10 クラスを分類する方が難しいと考えられる。また、この実験 6 のような多様なクラスの画像を多数集めることは、市販の画像コレクションや商用画像データベースを利用することでは難しく、Web を利用することによって初めて可能になったと言える。

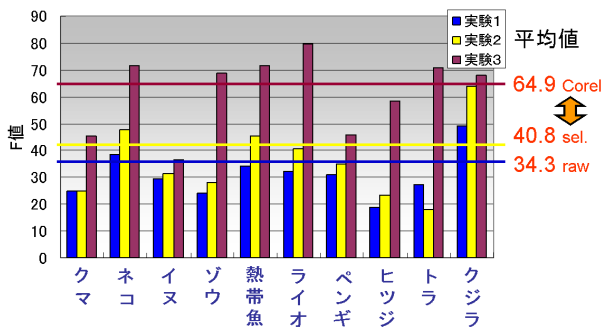


図 4: 実験 1, 2, 3 の結果 (方法 1)。

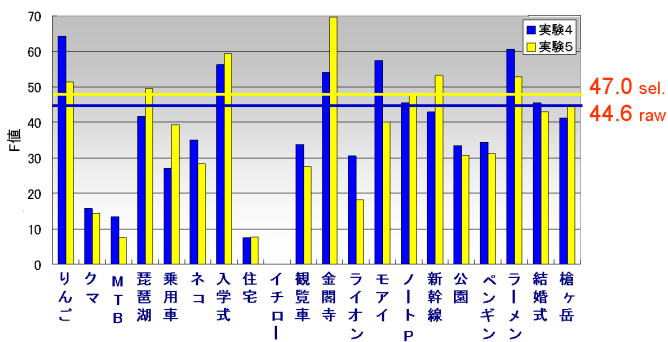


図 5: 実験 4, 5 の結果 (方法 1)。

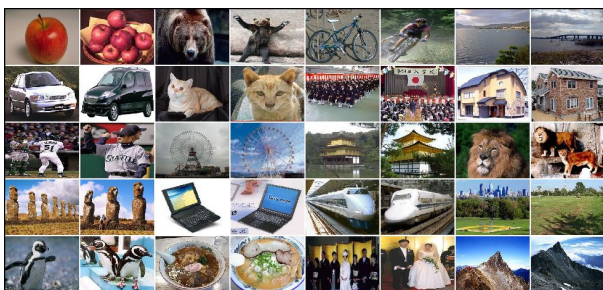


図 6: 実験 4, 5 に用いた評価画像の一部。

5. おわりに

本研究では、従来の学習画像を手で与えるのとは異なり、キーワードを入力するのみで、画像に関する知識やさらには

学習画像自体もユーザが与えることなく、Web から学習画像を自動的に収集して、実世界画像の自動分類を行うシステムを実現した。従来、Web からの知識獲得は、数値データやテキスト情報が中心であったが、本研究では画像情報についての Web からの知識獲得を行ったことになる。実験では、人手で選び出した典型画像各クラス 50 枚に対する分類結果として 44.6 の F 値を得た。このことは、本手法を発展させることによって、一般画像の自動分類がある程度は実現可能であることを示している。

今後の課題としては、さらにクラス数と収集画像を増やして実験を行うことがあげられる。その際は画像分類の高速化の工夫が必要である。

今回は画像全体を 1 つのクラスに分類した。しかし、現実には「ネコ」と「イヌ」が写っている画像も存在する。こうした画像に対処するために、やはり対象の切出しが必要なのか、それとも「ネコ」クラスと「イヌ」クラスの両方に分類できるようにすればよいのかは問題である。この問題に対処するためには、少なくとも画像の部分同士のマッチングを行うことが必要である。

参考文献

- [Barnard 01] Barnard, K. and Forsyth, D. A.: Learning the Semantics of Words and Pictures, in *Proc. of IEEE International Conference on Computer Vision*, Vol. II, pp. 408–415 (2001).
- [Belongie 97] Belongie, S., Carson, C., Greenspan, H., and Malik, J.: Recognition of Images in Large Databases Using a Learning Framework, Technical Report 07-939, UC Berkeley CS Tech Report (1997).
- [Duygulu 02] Duygulu, P., Barnard, K., de Freitas, J. F. G., and Forsyth, D. A.: Object Recognition as Machine Translation: Learning a Lexicons for a Fixed Image Vocabulary, in *Proc. of European Conference on Computer Vision* (2002).
- [Framkel 96] Framkel, C., Swain, M. J., and Athitsos, V.: Web-Seer: An Image Search Engine for the World Wide Web, Technical Report TR-96-14, University of Chicago (1996).
- [Lempel 02] Lempel, R. and Soffer, A.: PicASHOW: Pictorial Authority search by Hyperlinks on the Web, *ACM Transactions on Information Systems*, Vol. 20, No. 1, pp. 1–24 (2002).
- [Rubner 00] Rubner, Y., Tomasi, C., and Guibas, L. J.: The Earth Mover's Distance as a Metric for Image Retrieval, *International Journal of Computer Vision*, Vol. 40, No. 2, pp. 99–121 (2000).
- [Sclaroff 99] Sclaroff, S., LaCascia, M., Sethi, S., and Taycher, L.: Unifying Textual and Visual Cues for Content-Based Image Retrieval on the World Wide Web, *Computer Vision and Image Understanding*, Vol. 75, No. 1/2, pp. 86–98 (1999).
- [Smith 97] Smith, J. R. and Chang, S. F.: Visually Searching the Web for Content, *IEEE Multimedia*, Vol. 4, No. 3, pp. 12–20 (1997).
- [Wang 01] Wang, J. Z., Li, J., and Wiederhold, G.: SIMPLIcity: Semantics-Sensitive Integrated Matching for Picture Libraries, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 9, pp. 947–963 (2001).
- [Yanai 03] Yanai, K.: Image Collector II: A System for Gathering More Than One Thousand Images from the Web for One Keyword, in *Proc. of IEEE International Conference on Multimedia and Expo* (2003), (to appear).
- [森 01] 森, 高橋, 岡: 単語群つき画像の分割クラスタリングによる未知画像からの関連単語推定, *電子情報通信学会論文誌 D-II*, Vol. J84-D-II, No. 4, pp. 649–658 (2001).
- [柳井 01] 柳井啓司: キーワードと画像特徴を利用した WWW からの画像収集システム, *情報処理学会論文誌: データベース*, Vol. 42, No. SIG10 (TOD11), pp. 79–91 (2001).