

第 9 章

WWW からの収集画像を用いた画像分類の提案

本章では、第8章にて提案した画像収集法を用いて様々な実世界画像を自動収集し、それを学習画像として、実世界画像に対する画像分類を行うことを提案する。最初に学習段階では、WWW(World-Wide Web)から認識したい対象、例えば、「ライオン」「りんご」などの画像を各種類(クラス)毎に数百枚から数千枚程度収集し、色、テクスチャなどを画像特徴として抽出し、各クラス毎に画像特徴に関する知識ベースを構築する。次に、分類段階では、分類対象の画像から同様に画像特徴を抽出し、知識ベースと照合を行い、最も可能性の高いクラスに分類する。

本章では、まず WWW からの収集画像を画像分類の学習データとして用いる構想について述べ、分類実験の方法およびその実験結果について述べる。

9.1 はじめに

デジタルカメラやデジタルビデオ、イメージスキャナーなどの画像入力機器の普及により、日常の実世界のシーンをデジタル画像として計算機に取り込むことが容易になって来ている。また、ハードディスクなどの記憶装置の大容量化によって、大量の画像を計算機に蓄えることが可能となっている。しかし、その一方で、日常の実世界のシーンの画像が非常に多種多様に及ぶものであるために、対象を限定しないと適用できないことが多い従来の画像認識の手法では、画像の内容を自動的に認識することは難しく、現在のところは、大量に蓄積された画像をその内容によって整理分類するためには、人手の介入が不可欠である。

例えば、動物園で撮影した「ライオン」の画像を探したい場合、予め何らかのキーワードをつけない場合は、画像を検索する手がかりとしては画像ファイルのタイムスタンプくらいしかなく、結局は画像を表示して、その中から人間が目視によって探し出すしか方法がない。予め画像にキーワードをつけておけば、検索は容易であるが、計算機に取り込んだすべての画像に1枚ずつ手作業でキーワードを付けることは極めて手間の掛る作業である。このように、デジタル画像の作成や蓄積が容易なのに比べて、現状では画像がどういうシーンを表しているかという画像の意味内容に踏

み込んだ処理を自動的に行うことは困難であり、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け、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が可能となることが望まれている。

現在までに、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 [57][56]、意味的に類似している画像を検索する研究 [55] 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ので、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、その評価実験において、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用いられている。一般に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て、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、こうした研究の評価実験における検索精度は現実の日常シーンの画像に対する場合のそれよりもかなり良くなっていると考えられる。例えば、画像データベースの研究で頻繁に用いられている著作権フリーの画像コレクションである Corel 社の Corel Image Gallery に含まれる 6 万枚の画像のうち、「ライオン」の写真は、ほとんどすべてサバンナのライオンの写真であり、動物園で檻の中にいるライオンや、サファリパークで放し飼いになっているライオンの写真はまったく含まれていない。これは、我々がデジタルカメラでスナップ写真として撮影するような日常シーンの画像は、データベース化されておらず、大量に入手することが困難であるからである。また、世の中のあらゆる対象について、商用の画像コレクションが存在しているわけではなく、画像コレクションの中のある特定の対象についての画像の枚数も限られている (例えば、Corel 画像は各カテゴリ 100 枚ずつ) ので、学習画像を増やそうとすると限界がある。

そこで、本章では、WWW(World-Wide Web) から実世界のシーンを撮影した写真画像を自動収集し、それを市販の画像コレクションの代りに、単語と画像の対応付けのための学習画像として利用することを提案する。WWW 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。また、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、類似している画像が多く含まれる市販の画像データベースとは異なり、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る。

具体的には、第 8 章で提案した WWW からの自動画像収集システムを用いて、分類したい画像のクラスを表現するキーワードを入力して、それに関係する画像を収集する。そして、次にそれらを学習画像として、それぞれの画像から画像特徴量を抽出して、分類したい画像 (テスト画像) の画像特徴量とを比較することによって、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による画像認識を試みる。つまり、本研究における「認識」とは、画像が表すシーンの内容に応じて画像まるごと 1 枚をシーンに応じたクラスに分類することであり、ここでは「認識」と「分類」を等価に扱っている。

9.2 方針

本研究は、従来の画像認識システムで行われていた、画像知識ベースを構築するための学習画像を手で与えるのとは異なり、最初にシステムに認識したい対象を表す単語を与えるだけで、対象に関する知識や学習画像をユーザが与えることなく、WWW から学習画像を自動的に収集して、画像認識のための画像知識ベースを自動構築する新しい試みである。

本研究では、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。この方法は、「画像の見た目が十分類似している画像は同じクラスの画像である」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もので、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てしま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か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、「WWW から収集した大量の画像を学習画像とすることにより、実世界の視覚的変化の多様性を補うことが可能である」ということを仮定する。

本研究では対象の切出しをしないので、画像特徴の分布が複雑になることが予想できる。そこで、同一クラス内での画像特徴の分布の統計モデルは仮定しないで、画像特徴空間の十分近傍では、つまり（定義した距離において）良く似ているならば、同じクラスに属するとみなすという、ノンパラメトリックな方法によって分類を行う。対象の切出しをすれば、顔画像のように統計モデルを持つと仮定することもできるが、そのためには学習画像中の対象物体の位置を手で指定してやる必要がある。また、M.Weber らの研究 [61, 62] の様に、多数の学習画像から統計的にクラスに共通の特徴を抽出する方法も存在するが、その場合は学習画像の質が重要であり、かなり似通った画像を学習画像として与える必要がある。しかし、本研究では自動的に画像認識のための知識ベースを構築することが目的であり、処理途中での人手の介入は目的に反するので、人手による対象物の切出しや画像の選別はせずに、Web から自動収集した画像をそのまま学習画像として利用することとする。

入力画像と類似している学習画像のクラスに分類するという方法は、Nearest Neighbor 法 (NN 法) によって分類するということである。WWW 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、特徴空間中での分布の統計モデルを仮定することが困難であるので、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。

方法としては簡単であるが、WWW から収集した様々な画像を用いることによって、人間が様々な意図を持って撮影した写真画像のうち、クラス分けし易いクラスと、分類が難しいクラスが実験を通して明らかになることが期待できる。つまり、画像の見た目の多様性が非常に多いクラスと、見た目の多様性が少ないクラスが明らかになる。従来の画像データベースを用いることでは、画像の統一感を出すために同一クラス内で意図的に互いに類似している画像を選んだり、同一の撮影者

による同一の場所での撮影されたものの画像を多く含んでいたりしたために、クラスがもつ本来の多様性は知ることが出来なかった。それに対して、WWW は様々な撮影者が様々な意図を持って、撮影した画像を大量に含んでいるために、WWW から画像を収集することによって、真に「実世界シーンの一般的な画像」を得ることが可能であると考えられる。

以上述べた分類方法は、大量にクラスが既知の画像を用意し、それと十分似ている画像はその画像と同じクラスに分類するというものであるが、その場合「画像の類似性」の定義が重要である。ここでは、対象とする画像が実世界画像であるということ以外はまったく制約がないので、どのような画像にも適用可能な一般性のある方法で画像の類似性を定義することとする。画像間の類似性を定量的に定義するために、対象を限定しない画像特徴の抽出法と、抽出された特徴間の距離 (もしくは非類似度 (dissimilarity)) の計算方法として、以下の 3 通りの方法を用いた。

- ブロック分割した各ブロックから、色ヒストグラム [15], 高次自己相関特徴量 [190, 191], DCT (離散コサイン変換) 係数を特徴量として抽出し、主成分分析 (Principal Component Analysis) を行い、その結果の寄与率の高い成分のみを類似度計算の特徴として利用する。距離は Euclid 距離によって計算する。

これは 1 つのブロックの特徴量を 1 つ特徴ベクトルとして、ベクトル間の Euclid 距離をブロック間の非類似度とする方法で、従来の画像検索では最も一般的な方法で、かつ古典的な方法である。

- ブロック分割した各ブロックから、15 色程度の代表色を抽出し、Earth Mover's Distance (EMD) [16] によって距離を求める。

EMD による方法では、任意の代表点とその重みの組みの集合で表現された離散分布間の距離を求めることができ、従来の分割の仕方が固定されていたヒストグラムによる色分布の比較 [15] よりも高い精度で色分布の距離を求めることが可能である。

- 画像全体を色とテクスチャ特徴によって領域分割し、Integrated Region Matching (IRM) [17] によって距離を求める。

IRM は、EMD を領域分布同士の距離計算に応用した方法で、距離計算時に領域のテクスチャ特徴と形状特徴を考慮にいれている点が EMD とは異なる。

初めの 2 つの方法では、各ブロックについて最も類似しているブロックを求めて、多くのブロックが最も類似しているクラスに分類することとする。基本的には、領域分割の代りに多重解像度のブロック分割を用いる以外は、S. Belongie らの研究 [55] に類似した方法である。また、3 つ目の IRM による方法では、画像のブロック分割を行わないので、単純に最も類似していた画像のクラスに分類する。なお、最後の方法については、上位 k 個の中で最も多く類似しているクラスに分類する k -NN 法を用いた。

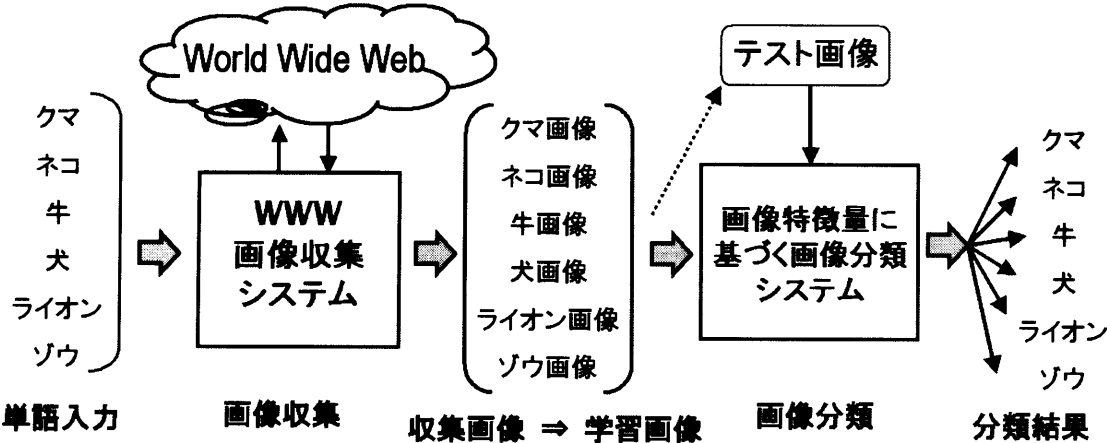


図 9.1 画像分類実験の方法

9.3 方法

9.3.1 画像収集の方法

最初に、第 8 章で提案した WWW 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。画像収集システムは、検索エンジン用キーワード、画像分類用キーワード、ファイル名キーワードの 3 種類のキーワードを入力するだけで、WWW からキーワードに関連した画像を収集することができるため、画像分類を行いたい画像に関するキーワードをユーザは入力するだけで、分類実験を行うことが可能である。図 9.1 では、「クマ」「ネコ」「牛」「犬」「ライオン」「ゾウ」の 6 種類の画像について、画像を収集し、それらを用いて画像分類実験を行う例を示している。本章では、画像収集システムに入力したキーワードと画像分類でのクラスが一对一に対応することとして、以後の話を進める。

9.3.2 自動分類の方法

画像収集システムによって収集された画像をクラス既知の学習画像として利用して、画像分類を試みる。初めに、数種類のキーワードに関する画像を収集し、それらを学習画像として、それぞれの画像から抽出した画像特徴量を分類したい画像の画像特徴量と比較することによって、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する (図 9.1)。分類は基本的には、Nearest Neighbor 法 (NN 法) を分類器として用いることとする。NN 法ではノンパラメトリックな分類方法で、すべての学習画像がそのまま各クラスの代表画像となるので、クラス内の学習画像の特徴量分布がまったく予想できないような今回の試みには向いているといえる。

画像から抽出する特徴量と、特徴量間の距離の計算方法としては、対象に依存しない方法として、色ヒストグラムと高次自己相関特徴量を特徴としユークリッド距離を用いる方法、Earth Mover’s Distance(EMD)[16] による方法、Integrated Region Matching(IRM)[17] による方法の 3 通りの方

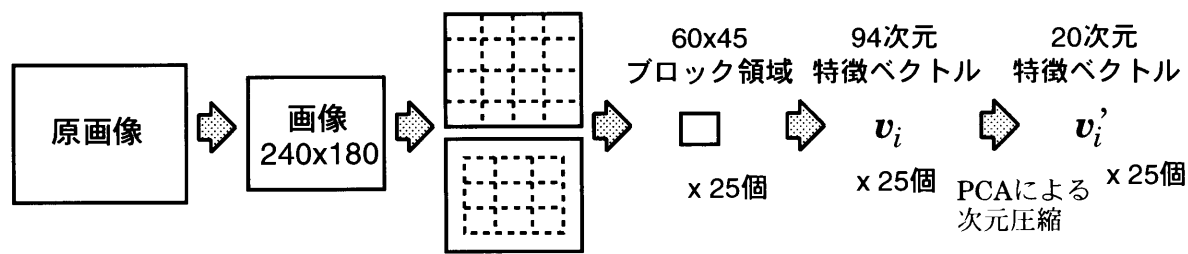


図 9.2 画像特徴ベクトルの作り方 (方法 1)

法を用いた。なお、3 番目の IRM による方法では、*k*-means クラスタリングによる領域分割と、JSEG[192, 193] による領域分割の 2 通りの方法を用いて実験を行った。次に、それぞれの場合の画像特徴抽出方法とクラス分類の方法について述べる。

9.3.3 方法 1：カラーヒストグラムと自己相似特徴量の利用

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から画像特徴量を抽出した (図 9.2)。具体的には、画像を 240×180 の大きさに正規化し、縦横をそれぞれ 1 から 4 分割し、さらに縦横それぞれに半ブロック分ずらして 0 から 3 分割する。実験で行った最大の分割数の 4 の場合、4 分割の 16 個の 60×45 のブロック画像を生成し、さらに半ブロック分ずらして、左右上下の端を除いた縦横それぞれ 3 分割の 9 個の 60×45 のブロック画像を生成する。これら合わせて 25 個のブロックから、それぞれ画像特徴量を抽出する。4 分割以外の場合は、1 分割の場合はブロック分割なし、2 分割の場合は縦横 2 分割とずらした後の 1 ブロックの合計 5 ブロック、3 分割の場合は縦横 3 分割と半ブロックずらした後の 2×2 ブロックの合計 13 ブロックとなる。

画像特徴量としては、均等色空間である Lu^*v^* 空間を各軸 4 分割した $4 \times 4 \times 4$ の 64 次元色ヒストグラム [15]、2 次までの高次自己相関特徴量 [190, 191] 25 次元、2 次までの DCT (離散コサイン変換) 係数 5 次元の合計 94 次元の特徴量を用いた。これらの 94 次元の特徴量を各ブロックから抽出する。特徴ベクトルをすべての学習画像のすべてブロックについて求めて、主成分分析 (Principal Component Analysis, PCA) によって 20 次元に圧縮する (図 9.2)。この時、求めた 94 次元ベクトルから 20 次元ベクトルへの変換行列を用いて、分類時にはテスト画像についても同様に 20 次元の特徴ベクトルを求める。なお、今回の行った実験ではほとんどの場合において、固有値の大きい順に 20 番目までの主成分を用いることで、その累積寄与率が 95%を越えていた。

次に分類の段階では、テスト画像の各ブロックの変換後の 20 次元特徴ベクトルを、すべてのクラスのすべての学習画像のすべてのブロックとの 20 次元特徴ベクトルと、ユークリッド距離による距離 (非類似度) を計算し、各クラス毎に最も近い距離を求める。そして、各クラス毎に全ブロック中の最小距離を合計して、さらに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した。各ブロックについての最近距離を合計する様にしたため、画像のすべてが類似していなくても、ある程度の部分の類似していれば、画像の類似性が高いとして正しいクラスに分類されることが期待できる。

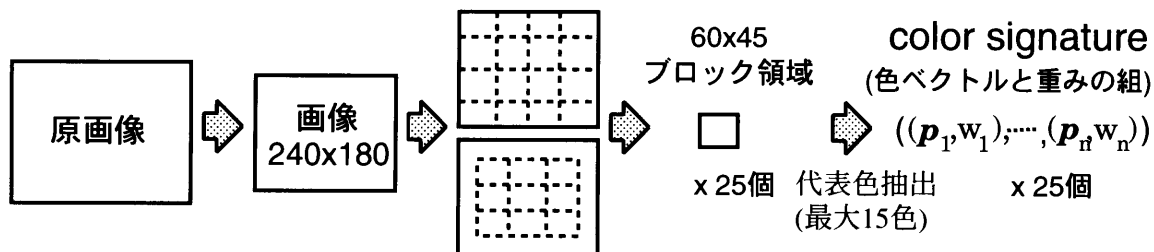


図 9.3 画像特徴 signature の作り方 (方法 2)

なお、クラス毎に個別の主成分分析を行いベクトル圧縮を行う部分空間法による分類や、判別分析法によりベクトルの圧縮を行い分類を行う方法についても予備実験を行ったが、分類結果にあまり大きな改善は見られなかったので、本章では、ベクトル圧縮を用いた分類実験としては、全クラスをまとめて主成分分析を行った方法の分類実験だけについて述べることにする。

9.3.4 方法 2 : Earth Mover's Distance の利用

前節の方法 1 は NN 法による多次元ベクトルの分類に若干手を加えた程度の簡単な分類方法であり、パターン認識の古典的な方法であった。本節で述べる 2 番目の方法では、画像特徴を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 \mathbf{p}_i とその重み w_i の組 (\mathbf{p}_i, w_i) の集合として画像特徴を表現する (図 9.3)。この多次元ベクトルとその重みの集合による特徴表現を signature と呼ぶ。

ここでは、signature を構成する特徴として色特徴のみを用いた。これを color signature と呼ぶ。Color signature の場合、 \mathbf{p}_i は色を表す色ベクトルとなる。例えば、白と黒を半分ずつ含む画像を color signature により表現には、白を表す色ベクトルを $(0, 0, 0)$ 、黒を表す色ベクトルを $(1, 1, 1)$ とすると、 $\{((0, 0, 0), 0.5), ((1, 1, 1), 0.5)\}$ となり、常に固定長で表現される color histogram よりも簡潔に表現できる。また、色数の多い画像の場合は signature の要素である色ベクトルと重みの組を増やすことによって、表現力が向上させることができる。通常は、画像の各画素の色ベクトルを k -means 法などによってクラスタリングして、10~20 色程度に減色する。そして、各クラスタの代表色 (平均色) を色ベクトル、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。本章の実験では、図 9.3 に示したように、方法 1 と同様にブロック分割した各ブロック画像に対して、 k -means 法によるクラスタリングによって、15 色からなる color signature を作成した。

Signature はベクトルと重みの集合であり、離散分布を表現しているために、従来のユークリッド距離のような簡単な計算方法では、2 つの signature 間の距離を求めることは不可能である。そこで、Y. Ruber らによって、signature 間の距離 Earth Mover's Distance (EMD) [16] が提案された。EMD は線形計画問題の一つである輸送問題 (transportation problem) の解に基づいて計算される。一方の signature を供給地の集合とそれぞれの供給量、もう一方の signature を需要地の集合とそれぞれの容量とみたとて、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荷物を輸送する総輸送コストが最小になるような荷物の輸送方法を探る。なお、この輸送問題による signature 間の距離の計算法は小林による絵画の分類についての研究 [182] においても同様に提案されている。

次に、 m 個の要素を持つ signature $P = \{(\mathbf{p}_1, w_{p_1}), \dots, (\mathbf{p}_m, w_{p_m})\}$ と n 個の要素を持つ signature $Q = \{(\mathbf{q}_1, w_{q_1}), \dots, (\mathbf{q}_n, w_{q_n})\}$ の間の EMD について考える。なお、 $1 \leq i \leq m, 1 \leq j \leq n$ の時に $\mathbf{p}_i, \mathbf{q}_j$ は P, Q の特徴ベクトルを表し、 w_{p_i}, w_{q_j} はそれぞれの特徴ベクトルの重みを表すこととする。まず最初に、すべての \mathbf{p}_i と \mathbf{q}_j の間の距離 ground distance $d_{ij} = d(\mathbf{p}_i, \mathbf{q}_j)$ を与える。通常は、これはユークリッド距離が用いられることが多い。次に、 P, Q 間の $\text{EMD}(P, Q)$ を計算するために、signature P を供給地、signature Q を需要地、ground distance d_{ij} を供給地と需要地間の輸送コストとして、 P から Q への最適フロー $\mathbf{F} = f_{ij}$ を以下の輸送問題を解くことによって求める。なお、これは、ヒッチコック型の輸送問題 (transportation problem)[194] であり、多くの効率的な解法が提案されている [195]。

$$\text{minimize WORK}(P, Q) = \sum_{i=1}^m \sum_{j=1}^n f_{ij} d_{ij} \quad (9.1)$$

$$\text{subject to } f_{ij} \geq 0 \quad (1 \leq i \leq m, 1 \leq j \leq n) \quad (9.2)$$

$$\sum_{j=1}^n f_{ij} \leq w_{p_i} \quad (1 \leq i \leq m) \quad (9.3)$$

$$\sum_{i=1}^m f_{ij} \leq w_{q_j} \quad (1 \leq j \leq n) \quad (9.4)$$

$$\sum_{i=1}^m \sum_{j=1}^n f_{ij} = \min\left(\sum_{i=1}^m w_{p_i}, \sum_{j=1}^n w_{q_j}\right) \quad (9.5)$$

式 9.1 は、 $\mathbf{p}_i, \mathbf{q}_j$ 間の輸送量に輸送コストを掛けて和を取ったもので、総輸送コストを表す。制約式 9.2 は、供給から需要への一方向にしか輸送が行われなことを示す。制約式 9.3, 制約式 9.4 は、それぞれ、供給地 i から供給できる量の制約、需要地 j で受け取れる量の制約を示している。最後の制約式 9.5 は、総輸送量が総供給量が総需要量のどちらか少ない量と一致することを表している。

最終的に $\text{EMD}(P, Q)$ は、次の式で表されるように、求めた最小の総輸送コスト $\text{WORK}(P, Q) = \sum_{i=1}^m \sum_{j=1}^n f_{ij} d_{ij}$ を総輸送量で割って正規化した値となる。

$$\text{EMD}(P, Q) = \frac{\sum_{i=1}^m \sum_{j=1}^n f_{ij} d_{ij}}{\sum_{i=1}^m \sum_{j=1}^n f_{ij}} \quad (9.6)$$

方法 2 で用いる color signature の場合は、色ベクトルを Lu^*v^* 色空間の 3 次元ベクトルで表現することとし、ground distance にはユークリッド距離を用いた。 Lu^*v^* 色空間は、任意の 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ように構築された色空間であり、均等色空間と呼ばれている [181]。そのため、通常の RGB 色空間よりも画像間の距離を求める場合には適しているといえる。

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 (ビン (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 を用いて、EMD によって距離計算を行うことによって、より人間の感覚に近い「距離」の計算を行うことが出来る。図 9.4 の例では、 f_1, f_2 の色分布の距離と f_3, f_4 の色分布の距離について、(a) ユークリッド距離で計算した場合と (b) EMD によって計算した場合の違いについて示している。ユークリッド距離の場合は同一の位置のビン間の重みのみが比較されるのに対して、EMD の場合は近い色同士の間で対応付けが行われる。そのため、(a) では $d(f_1, f_2) > d(f_3, f_4)$ となってしまう人間の感覚と合わないが、(b) だと $EMD(f_1, f_2) < EMD(f_3, f_4)$ となり、より人間の感覚に近い距離になっている。EMD は画像特徴に基づく画像検索の研究において数多く提案されている画像間の類似度の計算方法としては平均的に最も優れた性能を示すことが EMD の提案者の Y. Rubner らによって示されている [196]。

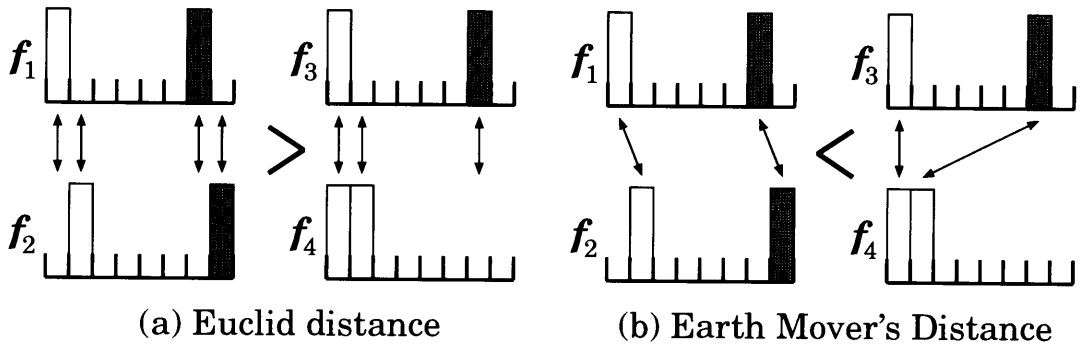


図 9.4 (a) Euclid 距離による距離計算. (b) Earth Mover's Distance (EMD) による距離計算. (a) では $d(f_1, f_2) > d(f_3, f_4)$ となってしまう人間の感覚と合わないが、(b) だと $EMD(f_1, f_2) < EMD(f_3, f_4)$ となり、より人間の感覚に近い距離になっている。

最後に方法 2 での分類の方法であるが、距離計算に EMD を用いる以外は基本的に同じで、学習画像、テスト画像共に、最初に方法 1 と同様に画像をブロック分割して、それぞれのブロック画像について color signature を求める (図 9.3)。次に、テスト画像の各ブロックの color signature と、すべてのクラスのすべての学習画像のすべてのブロックの color signature との間で EMD を計算し、各クラス毎に最小距離を求める。そして、各クラス毎に全ブロック中の最近距離を合計して、さらに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。なお、実際には、EMD の下限値を最初に計算して、現在得られている最小の EMD の値より下限値が下回った時のみ、実際の EMD を計算することとし、すべての組み合わせについて計算コストの掛る EMD を求めることを避けている。文献 [16] によると、2 つの signature の重みの総和が等しい時 ($\sum_{i=1}^m w_{p_i} = \sum_{j=1}^n w_{q_j}$)、EMD の下限値 (lower bound) は以下に示すように、2 つの signature の重心 \bar{P}, \bar{Q} 間のユークリッド距離となる。

$$EMD(P, Q) \geq \|\bar{P} - \bar{Q}\| \quad (9.7)$$

ただし、

$$\bar{P} = \frac{\sum_{i=1}^m w_{p_i} \mathbf{p}_i}{\sum_{i=1}^m w_{p_i}}, \quad \bar{Q} = \frac{\sum_{j=1}^n w_{q_j} \mathbf{q}_j}{\sum_{j=1}^n w_{q_j}}, \quad (9.8)$$

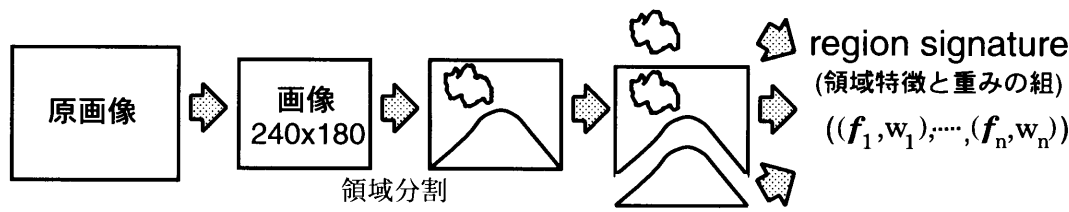


図 9.5 画像特徴 signature の作り方 (方法 3)

とする.

9.3.5 方法 3 , 方法 4 : Integrated Region Matching の利用

Integrated Region Matching (IRM)[17] 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 signature を用いて、EMD による画像間の距離計算を行う. 方法 3 , 4 では画像の領域分割を行うため、方法 1 , 方法 2 で行っていた様な画像のブロック分割した後に各ブロック画像毎に特徴量を抽出することは行わない.

方法 3 では、 k -means 法によるクラスタリング (LBG アルゴリズムによるベクトル量子化と等価) によって、領域分割を実現する. 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 法によってクラスタリングする. k の値は最初 3 としてある一定以下の大きさのクラスタが現れるまで、順次 k を増加させて、繰り返しクラスタリング行う. k の最大値は 15 とし、ある一定以下の大きさのクラスタが現れたらその 1 つ手前のクラスタリング結果を、 $k = 15$ となったらそこでのクラスタリング結果を、各クラスタを分割領域と見なすことによって、最終領域分割結果とする.

この方法は実装が簡単であるために、画素間の連結性は考慮されずに色とテクスチャ特徴のみによって領域が分割される. 一方、通常の領域分割アルゴリズムでは、画素間の連結性が考慮されるのが普通であるが、実装は複雑になる. 領域分割アルゴリズムの研究は古い歴史を持っているが、近年においては画像検索のためのカラー画像に対する領域分割アルゴリズムがいくつも提案されている. 例えば、normalized cuts[136], graph-based segmentation[197], Blobworld[132, 133], JSEG[192, 193] などがある. 方法 4 では、方法 3 の k -means クラスタリングの代りに、JSEG によって領域分割を行うこととする. 異なるのは領域分割の方法のみで、領域分割後の処理は、方法 3 と方法 4 は同一である.

JSEG では、最初に k -means 法によって原色処理を行い、次に指定領内に関して色分布の偏りを表す J 値を求める. JSEG では、この J 値を画像全体で小さくする様に領域成長法によって画像の

領域分割を行う。詳細は文献 [193] に記述がある。JSEG では、予め指定する必要のあるパラメータが 3 種類ある。これらについては、予備実験によって求めた固定値を用いることとする。

図 9.6 に k -means クラスタリングによる分割結果および JSEG による分割結果の例を示す。JSEG による分割結果は領域が連続しているのに対して、 k -means クラスタリングによる分割結果は画素の連結性を考慮していないことが分かる。この結果からは、背景が単純な画像 (主に Corel 画像) に対してはうまく働いているものの、複雑な画像 (主に WWW 収集画像) に対しては領域分割アルゴリズムを用いた JSEG による方法でも、あまりうまく分割出来ていないことが分かる。

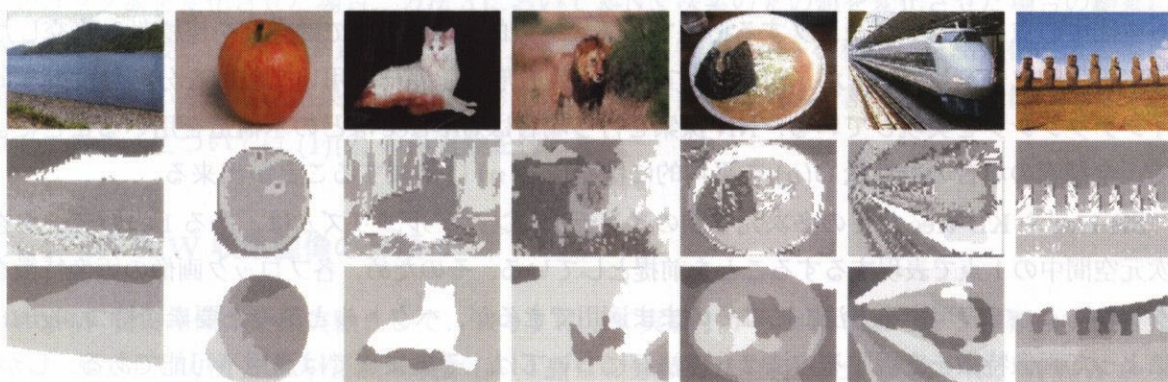


図 9.6 k -means クラスタリングによる分割結果 (中段), および JSEG による分割結果の例 (下段).

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慣性モーメント [198] 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 2 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、類似度が高い上位 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。後述べる実験においては、 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。

IRM と同様に領域分割した結果の各領域の特徴ベクトルとその面積比を signature の要素とし、EMD によって距離 (非類似度) を計算した研究には、谷田川の研究 [199] や H.Greenspan らの研究 [200] がある。これらの研究では、領域分割アルゴリズムとして JSEG [193], Blobworld [132, 133] がそれぞれ使われている。ただし、これらの研究は画像分類ではなく、画像検索システムの研究である。

なお、本章の実験では、基本的にすべてのテスト画像は学習時に用いた画像のクラスのどれかに属することとして実験を行う。ただし、第 9.4.3 節の実験においては、閾値を設けて、合計距離が閾値を超えた場合は「分類なし」とした。

9.3.6 SR-tree の利用による k -NN 検索の高速化

単純な全探索による類似画像検索を行うと、画像枚数に比例しただけの時間が掛ってしまい、学習画像の枚数を増やした場合、検索時間がその分増大してしまうという問題が発生する。EMD を用いる場合は、下限値を用いることによって、コストのかかる EMD の計算の回数を必要最低限におさえることが出来るが、全探索であることには違いない。そこで、探索の高速化の方法として、高速インデクシング手法である SR-tree(Sphere/Rectangle-tree)[201, 202, 18] を先に述べた 3 つの方法に適用する。

SR-tree は、kd-tree などと同様に多次元空間を分割して木構造を作り、多次元ベクトルで表される空間中の点をインデクシングする方法である。一般的な kd-tree では空間を座標軸と直交した平面で再帰的に分割していくが、SR-tree では包囲長方形と包囲球を用いて空間を分割し、効率的なインデクシングを実現している。NN 探索を行う場合は SR-tree などの木構造を用いることによって、全探索の場合の計算量 $O(n)$ を平均的には $O(\log_2 n)$ に削減することが出来る。

SR-tree や KD-tree などの多次元空間のインデクシングアルゴリズムは、ある 1 つのデータを多次元空間中の 1 点で表現することを前提としている。そのため、各ブロック画像の画像特徴をベクトルによって表現する方法 1 にはそのまま適用できるが、ベクトルと重みを要素に持つ signature によって画像特徴を表現する方法 2、方法 3 については、そのままでは適用不可能である。しかし、EMD による距離は 2 つの signature の重心間のユークリッド距離が下限値になるという特性があった。そこで、signature の重心を表すベクトルを SR-tree でインデクシングし、ユークリッド距離での重心同士の NN 探索を行うこととする。NN 探索では 1 つの signature に対して、距離が近い順に上位 n 個程度の signature を取出し、その n 個に対して上位から改めて EMD の計算を行う。その際に、現在得られている最小の EMD の値より下限値が上回った時には、EMD の計算を終了することとする。上位 n 個のうちに真に最小の EMD 値をもつ signature が存在しない可能性も考えられるので、 n の値が小さいと全探索の場合に比べて分類精度が低下する場合もある。一方、 n の値が大きいと、それだけ多くの EMD を計算することとなり、分類時間が増大することとなる。

9.4 分類実験

本節では分類実験の結果を示す。最初に (1)WWW から収集した 10 クラスおよび 20 クラスの画像について、それ自身をテスト画像として用いた場合の分類実験の結果を示す。次に、(2) 市販の画像コレクションをテスト画像として用いた場合、および、各クラスについて典型的であると考えられる画像 50 枚を収集画像から人手によって選び出し、テスト画像とした場合についての分類実験の述べ、続いて、(3) 分類するかどうかについて合計距離に関する閾値を導入し、その値を変化させた場合の再現率と適合率の変化について述べる。そして、さらに、(4) 方法 1, 方法 2 においてブロック数を変化させた場合、方法 3 において k -NN 探索の k の値を変化させた場合の結果について示す。最後には、(5) 高速インデクシング手法である SR-tree を利用した高速化による実行時間の変化についても実験結果を示す。なお、方法 3 と方法 4 は領域分割方法以外は同じ手法であるので、方法 4 については (1)(2) のみ実験を行った。

9.4.1 WWW 収集画像の分類

WWW 収集画像の分類実験として、表 9.1 に示す実験 1 から実験 5 までの 5 種類の実験を先に述べた方法 1, 方法 2, 方法 3, 方法 4 の 4 通りの分類方法で行った。実験 1 は、WWW から自動的に収集した 10 クラスの画像をそのまま分類した実験である。自動収集であるので、すべてが適切な画像であるという訳ではなく、不適切な画像も含まれている。実験 2 は実験 1 で用いた収集画像から人手によって不適合画像を除去した画像セット、つまり適合率 100% の画像セットを用いて同様に分類実験を行った。次に、実験 3 では、WWW 画像と市販の画像コレクションの特性を比較するために、実験 1, 実験 2 と同じクラスの画像を Corel 社の Corel Image Gallery から取出して、それに対して同様の分類実験を行った。また、実験 4 では 20 クラスで同様に実験を行い、実験 5 では実験 4 で用いた画像セットから不適合画像を除去した上で実験を行った。

本節の分類実験は、学習画像、テスト画像共に収集画像を用いる交差法 (cross validation) によって評価した。実験 3 では 1 回につき各クラスにつきテスト画像 10 枚ずつ、それ以外の実験では 1 回につき 20 枚ずつテスト画像を選び出し、それ以外の画像を学習画像として、それぞれ、異なるテスト画像を用いて、表 9.1 の交差法の () 内に示されただけの回数の分類実験を行い、その結果の合計によって評価を行った。なお、実際には完全にすべての画像をテスト画像に利用している訳ではないので、厳密に交差法とは言えないが、ここでは便宜的に交差法と呼ぶ。なお、単純に交差法で分類精度の評価を行ったので、実験 1, 実験 4 ではテスト画像にも不適合画像 (ノイズ画像) も含まれていることを留意する必要がある。

実験 1 の画像収集では、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」の動物に関係した 10 個の単語を画像収集システムに入力し、それぞれのクラスの画像を収集した。収集時には、6 つのテキストサーチエンジン (Goo, Infoseek, Excite, Google, Lycos, Infonavi) から重複を除いて、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 ((クラスに対して適切な画像の枚数) / (収集した画像の枚数) として計算) は 56% から 90% 程度で、347 枚から 592 枚の画像を収集することが

表 9.1 分類実験.

実験 no.	クラス数	学習画像		テスト画像	
		枚数	適合率 (%)	方法	入手元
1	10	4548	68.2	交差法 (10)‡	WWW
2	10	3102	100†	交差法 (7)	WWW
3	10	500	100	交差法 (5)	Corel
4	20	5694	61.2	交差法 (8)‡	WWW
5	20	3485	100†	交差法 (4)	WWW

†人手で適合画像のみを選択. ‡不適合画像も含めて利用.

出来た (表 9.2). 10 個の単語に関する収集画像の平均の適合率は 68.2%, 合計枚数は 4548 枚であった. 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, A 群として 307 枚, B 群として 669 枚の画像を収集した. さらに画像解析によって, 最終的に A 群, B 群から 283 枚と 221 枚の合計 504 枚の画像を選択し, その適合率は 77.0%であった.

表 9.2 実験 1, 実験 2 で用いる画像の収集結果.

クラス名	収集枚数	適合率
クマ	419	56.4
ネコ	354	62.0
イヌ	570	75.7
ゾウ	506	65.5
熱帯魚	275	89.9
ライオン	504	77.0
ペンギン	576	57.0
ヒツジ	347	64.0
トラ	405	68.7
クジラ	592	72.4
合計／平均	4582	68.2

次に実験 1 の画像分類では, 収集した画像のうち各クラスについて 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, テスト画像の分類実験を行った. 方法 1 から 4 までの 4 種類の分類方法について, テスト画像を変えて 10 回試行した結果を表 9.3 および図 9.7 に示す. 表 9.3 には, 各クラスに関する再現率 R ((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 P ((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)($F = \frac{2PR}{(P+R)}$) と, それぞれの平均値を記してある. なお, 値はすべて%表示である. 以下の分類実験では, すべて 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 クラスのこの場合 10.0 となる. それに対して, 得られた平均 F 値は方法 2 の場合に 34.3 でこれを大幅に上回る結果であった. 個々

のクラスについては、「クジラ」だけが飛び抜けてよい結果を示した。これは海面からクジラが少しだけ胴体を出しているという構図の「クジラ」画像(図 9.10)が多く存在したからである。また、3つの分類方法の中では、平均的には方法2が最もよい結果を示した。

表 9.4, 図 9.8は、実験2の結果である。実験2では、実験1で収集した画像を人手でチェックし、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、平均のF値がそれぞれの方法において6~7程度向上している。

実験3では、WWWから収集した画像ではなく、実験1, 実験2と同じクラスの画像をCorel社のCorel Image Galleryから各クラス50枚ずつ取出し、分類実験を行った。実験に用いた画像のクラスは、実験1, 実験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、実験を行った。その結果を表 9.5, 図 9.9に示す。実験3で得られたF値は実験2の最大のF値40.8に比べてかなり高く、特に、方法3においては68.1であった。これは、Corel Image Galleryが互いに類似している画像が多い含んでいることによると考えられる。また、領域分割手法以外は同様の方法である方法4が方法3に比べて結果が大幅に良くなかったことは、意外な結果であった。このことから、IRMによる画像分類においては、画像分割結果の各領域が連続していることは必要なことではないことが言える。

個々のクラスを見ると、「ライオン」が方法2において最もよい結果となっている。「ライオン」は実験1, 2においては、それほどよい結果ではなかったが、Webから収集した「ライオン」画像(図 9.12)とCorel「ライオン」画像(図 9.13)の傾向がだいぶ異なるためである。また、実験1, 2において比較的良い結果であった「クジラ」(図 9.11)も実験3においては、よい結果となっている。

実験4では、人物、場所、植物、動物、人工物など様々な種類の20のクラスの画像について画像収集を行い実験1と同様の分類実験を行った。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61.2%であった。分類実験の結果は表 9.6, 図 9.14に示す。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。実験の結果をそれぞれ表 9.6, 表 9.6に示す。実験3で得られた平均のF値を見ると、まったくランダムな場合は5.0となるので、それに比べるとかなり良い値が得られており、方法2によって、実験4, 実験5でそれぞれ37.2, 43.3のF値が得られている。これは、10クラスの分類実験である実験1, 実験2よりも良い値である。理由としては、実験1, 実験2では分類した10クラスがすべて動物で、異なるクラスでも比較的類似している画像があったのに対して、実験4, 実験5では倍の数の20クラスであっても様々な種類のクラスから成っていて、異なるクラスでの画像の類似が実験1, 実験2ほど多くなかったことが考えられる。また、実験4と実験5を比べた場合は、実験1と実験2の時と同様に不適切な画像を除去した実験5の方がF値が5~6ポイント向上している。

また、個々のクラスの結果に見てみると、クラスによって結果に大きく差があることが分かる。特に、赤い「りんご」、金色の建物が緑色の池の後方にある「金閣寺」(図 9.16)、どんぶりを真上から撮影した画像が多い「ラーメン」(図 9.17)、人の頭が多数写っている「入学式」(図 9.18)などは画像の構成が一定しており画像間の変化が少なく、類似した特徴を持つ画像が多く、実験5においては方法2において40から最大65もの値を得られている。その一方で、「イチロー」「住宅」「自転車」は10前後のF値しか得られなかった。「イチロー」(図 9.19)は顔のアップ、遠くからの写真、

後ろ姿、着ている服装が様々であるなどの様々な写真が含まれおり、「住宅」は様々な色、形のものが含まれていた。また、「自転車」は、自転車自体が細いフレームとタイヤで出来ており、画像特徴としては背景や乗っている人の特徴の方が、今回用いた色やテクスチャに関する画像特徴抽出では主要な特徴として抽出されてしまい、うまく分類することが出来なかった。「イチロー」のような人間を認識する場合は、画像毎に変化が大きく、「イチロー」の写っている画像をいくら大量に集めてもうまく行かない可能性が高い。「イチロー」を認識するためには、認識対象の切出しを行って、人間の顔の識別が出来ないといけないのはもちろんのこと、「野球選手」「マリナーズの 51 番」などの知識も総動員しなければ、精度の高い認識を実現することは不可能であろう。

実験 4, 実験 5 のような多様なクラスの画像を多数集めることは、市販の画像データベースを利用することでは難しく、WWW を利用することによって初めて可能になったと言える。本研究のような一般の実世界画像を認識することを目的とする場合には、市販の画像コレクションを利用するだけでは、その中にある画像しか認識できなくなるという問題がある。それに対して、常に新しい画像が追加されている WWW から画像収集を行うことで、ほとんどどのような対象についてもその学習画像が収集可能である。

最後に、図 9.20, 図 9.21 に平均的に最も結果のよかった方法 2 による実験 1, 2, 3, および実験 4, 5 の結果をそれぞれまとめて示した。図中の横の線はそれぞれの実験結果の平均値を示し、raw は収集した画像をそのまま分類実験に用いた実験 1, 4, sel. は人手によって正解画像のみを選択 (select) した後に分類実験を行った実験 2, 5, Corel は Corel 画像による分類実験を行った実験 3 をそれぞれ表している。これらの図からは、raw に比べると sel. の方が 1~2 割程度 F 値が向上しているが、キーワードによっては逆に低下している場合もあることが分かる。また、実験 3 の Corel については、すべての場合において、raw や sel. よりも結果が良くなっており、キーワードによっては 2 倍程度もの違いがある場合がある。

なお、参考までに付録 A に、50 クラスの場合の分類実験の結果を載せた。50 クラスの場合は、方法 2 によって 40.3 の F 値が得られており、20 クラスの実験 4 の結果と大差ない値となっている。50 クラスは様々なジャンルからクラスが選ばれており、このことから画像のクラスの種類が多様である場合では、20 クラスが 50 クラスに増えたとしても分類精度に大きな変化はないことがいえる。

表 9.3 実験 1 (10 クラス, 不適合画像あり) の分類実験の結果.

クラス名	方法 1			方法 2			方法 3			方法 4		
	再現率	適合率	F 値	再現率	適合率	F 値	再現率	適合率	F 値	再現率	適合率	F 値
クマ	17.8	23.3	20.2	21.0	31.1	25.1	27.5	26.4	27.0	24.0	23.1	23.5
ネコ	20.4	48.2	28.6	28.0	60.9	38.4	30.5	48.8	37.5	24.0	36.9	29.1
イヌ	27.0	17.6	21.3	40.0	23.3	29.4	33.0	22.2	26.6	26.0	19.2	22.1
ゾウ	21.9	17.8	19.6	25.0	23.1	24.0	25.0	22.3	23.6	22.0	19.9	20.9
熱帯魚	16.7	55.6	25.6	22.0	74.6	34.0	29.5	47.6	36.4	17.5	40.2	24.4
ライオン	34.4	21.9	26.8	45.0	25.2	32.3	26.0	21.9	23.8	21.5	17.5	19.3
ペンギン	34.4	20.1	25.4	33.5	29.0	31.1	28.5	27.8	28.1	22.5	18.8	20.5
ヒツジ	10.4	33.3	15.8	13.0	34.2	18.8	11.5	18.1	14.1	12.5	17.0	14.4
トラ	14.1	25.7	18.2	24.0	32.2	27.5	20.5	23.6	21.9	18.5	19.7	19.1
クジラ	61.5	38.2	47.1	66.5	39.0	49.2	61.5	44.1	51.4	54.5	41.4	47.1
平均値	25.9	30.2	27.8	31.8	37.3	34.3	29.4	30.3	29.8	24.3	25.4	24.8

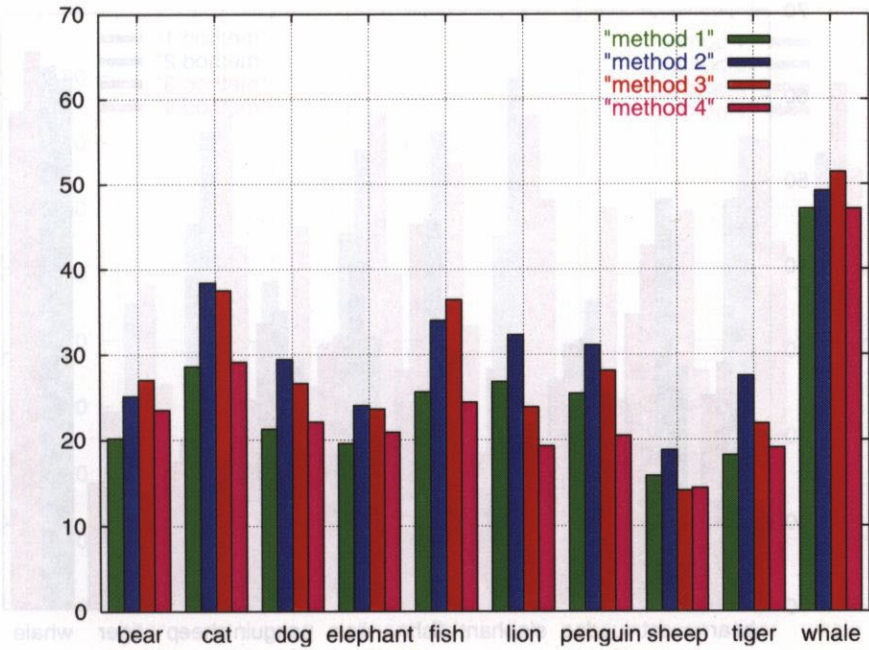


図 9.7 実験 1 (10 クラス, 不適合画像あり) の分類実験の結果.

表 9.4 実験 2 (10 クラス, 不適合画像なし) の分類実験の結果.

クラス名	方法 1			方法 2			方法 3			方法 4		
	再現率	適合率	F 値	再現率	適合率	F 値	再現率	適合率	F 値	再現率	適合率	F 値
クマ	7.1	43.5	12.3	17.1	46.2	25.0	22.9	38.1	28.6	22.1	28.2	24.8
ネコ	17.9	67.6	28.2	34.3	78.7	47.8	35.7	52.1	42.4	30.7	38.7	34.3
イヌ	54.3	16.7	25.5	58.6	21.5	31.4	44.3	23.6	30.8	36.4	20.9	26.6
ゾウ	17.1	28.6	21.4	25.0	32.1	28.1	30.7	25.6	27.9	18.6	17.4	18.0
熱帯魚	22.1	47.7	30.2	35.7	62.5	45.5	36.4	58.6	44.9	22.9	35.6	27.8
ライオン	37.9	27.5	31.8	47.9	35.1	40.5	34.3	29.8	31.9	32.1	29.8	30.9
ペンギン	45.0	21.5	29.1	47.9	27.3	34.8	38.6	29.3	33.3	27.9	21.7	24.4
ヒツジ	11.4	37.2	17.5	17.1	36.4	23.3	16.4	24.7	19.7	15.0	19.8	17.1
トラ	7.9	50.0	13.6	10.7	60.0	18.2	17.9	32.5	23.0	11.4	20.8	14.7
クジラ	72.1	54.6	62.2	75.0	55.6	63.8	76.4	57.2	65.4	67.1	51.6	58.4
平均値	29.3	39.5	33.6	36.9	45.5	40.8	35.4	37.2	36.2	28.4	28.5	28.4

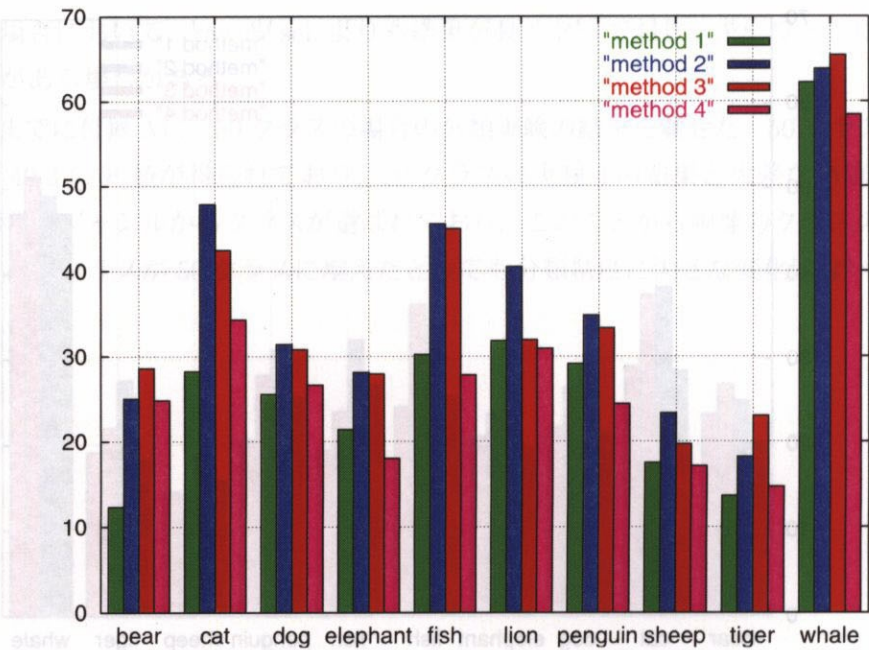


図 9.8 実験 2 (10 クラス, 不適合画像なし) の分類実験の結果.

表 9.5 実験 3 (10 クラス, Corel 画像) の分類実験の結果.

クラス名	方法 1			方法 2			方法 3			方法 4		
	再現率	適合率	F 値	再現率	適合率	F 値	再現率	適合率	F 値	再現率	適合率	F 値
クマ	18.0	32.1	23.1	36.0	62.1	45.6	40.0	60.6	48.2	28.0	41.2	33.3
ネコ	51.0	65.8	57.5	61.2	85.7	71.4	73.5	80.0	76.6	51.0	58.1	54.3
イヌ	44.0	55.0	48.9	24.0	75.0	36.4	44.0	81.5	57.1	26.0	44.8	32.9
ゾウ	50.0	64.1	56.2	68.0	69.4	68.7	72.0	76.6	74.2	48.0	52.2	50.0
熱帯魚	22.0	64.7	32.8	58.0	93.5	71.6	56.0	82.4	66.7	34.0	54.8	42.0
ライオン	64.0	49.2	55.7	82.0	77.4	79.6	86.0	65.2	74.1	66.0	56.9	61.1
ペンギン	48.0	32.9	39.0	50.0	42.4	45.9	70.0	52.2	59.8	34.0	28.3	30.9
ヒツジ	26.0	28.9	27.4	80.0	46.0	58.4	66.0	54.1	59.5	40.0	31.7	35.4
トラ	66.0	56.9	61.1	72.0	69.2	70.6	68.0	72.3	70.1	52.0	57.8	54.7
クジラ	70.0	36.5	47.9	94.0	53.4	68.1	96.0	66.7	78.7	92.0	51.1	65.7
平均値	45.9	48.6	47.2	62.5	67.4	64.9	67.1	69.2	68.1	47.1	47.7	47.4

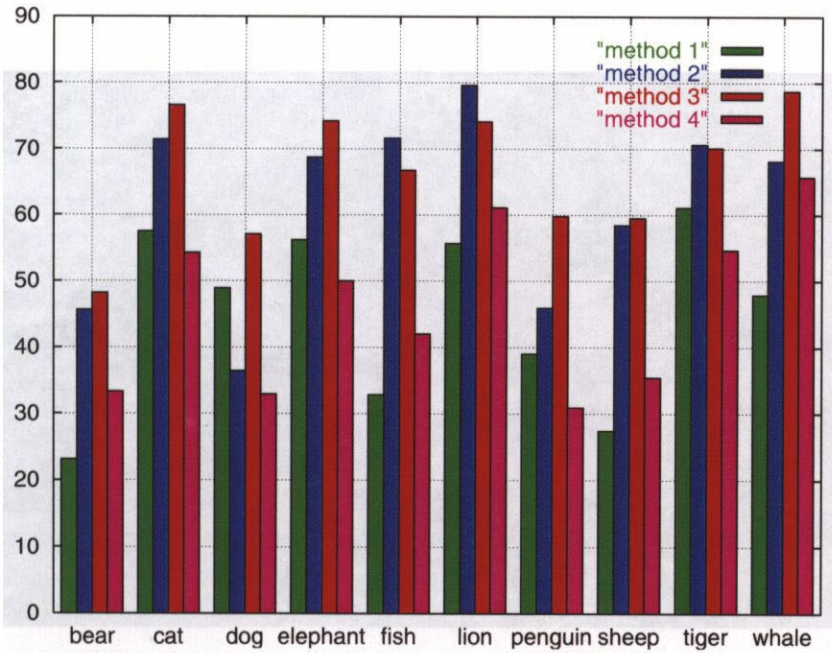


図 9.9 実験 3 (10 クラス, Corel 画像) の分類実験の結果.



図 9.10 「クジラ」 Web 画像.

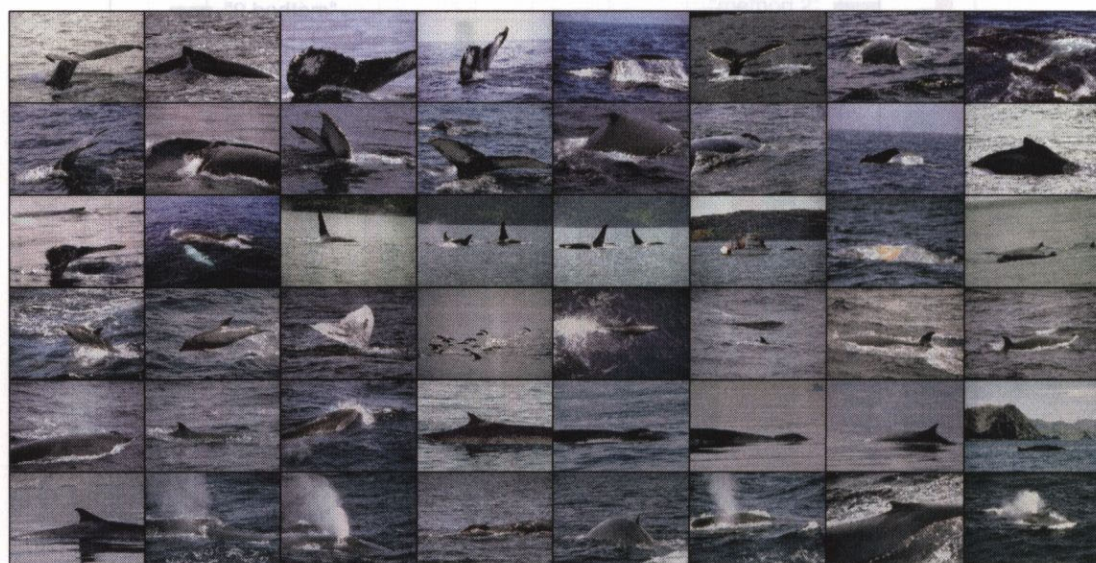


図 9.11 「クジラ」 Corel 画像.



図 9.12 「ライオン」 Web 画像.

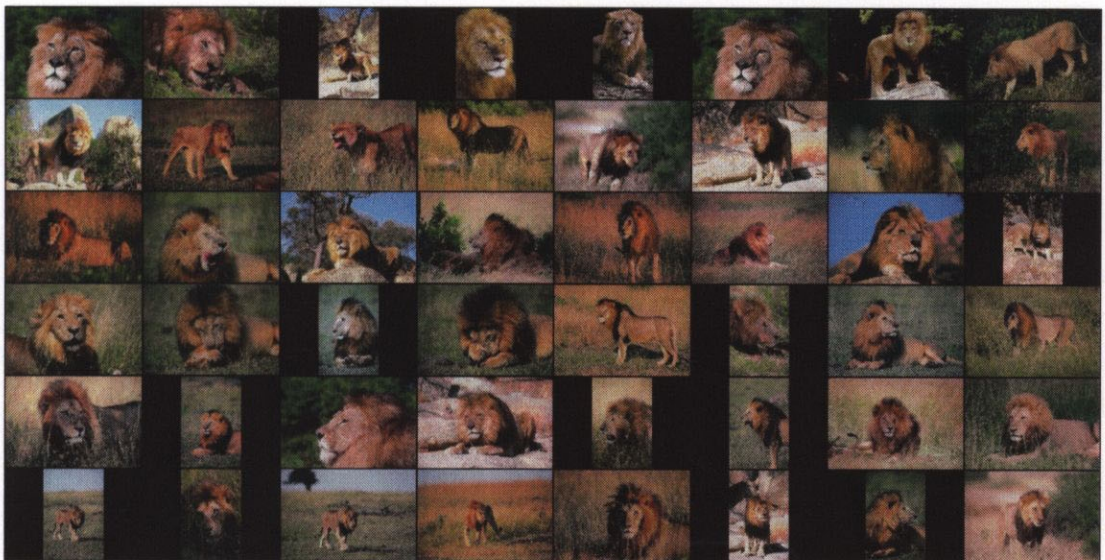


図 9.13 「ライオン」 Corel 画像.

表 9.6 実験 4 (20 クラス, 不適合画像あり) の分類実験の結果.

no.	クラス名	方法 1			方法 2			方法 3			方法 4		
		再現率	適合率	F 値	再現率	適合率	F 値	再現率	適合率	F 値	再現率	適合率	F 値
1	りんご	32.9	39.0	35.7	30.0	56.8	39.3	26.2	46.2	33.5	25.0	37.4	30.0
2	クマ	12.9	16.7	14.5	25.7	30.5	27.9	23.1	22.2	22.6	19.4	17.2	18.2
3	マウンテンバイク	20.0	14.3	16.7	10.0	60.9	17.2	13.8	21.6	16.8	15.6	21.2	18.0
4	琵琶湖	22.1	16.2	18.7	42.9	20.8	28.0	37.5	26.4	31.0	31.9	20.7	25.1
5	乗用車	24.3	50.0	32.7	16.4	79.3	27.2	29.4	55.3	38.4	26.9	45.7	33.9
6	ネコ	18.6	19.8	19.2	52.1	21.7	30.6	45.0	25.7	32.7	31.2	18.2	23.0
7	入学式	30.0	25.5	27.5	53.6	34.6	42.0	55.0	31.7	40.2	37.5	27.3	31.6
8	住宅	12.9	40.9	19.6	15.7	88.0	26.7	37.1	57.3	45.0	38.4	62.2	47.5
9	イチロー	0.7	33.3	1.4	2.1	100.0	4.2	6.2	50.0	11.1	4.4	30.4	7.7
10	観覧車	33.6	23.9	27.9	15.7	62.9	25.1	29.4	40.9	34.2	21.2	27.9	24.1
11	金閣寺	55.0	36.2	43.6	59.3	41.7	49.0	58.8	42.3	49.2	45.0	32.0	37.4
12	ライオン	15.7	43.1	23.0	13.6	86.4	23.5	15.6	34.7	21.6	16.9	26.2	20.5
13	モアイ	22.1	53.4	31.3	19.3	96.4	32.1	29.4	52.2	37.6	21.2	30.1	24.9
14	ノートパソコン	27.9	52.7	36.4	35.7	86.2	50.5	35.0	57.7	43.6	38.1	70.9	49.6
15	新幹線	36.4	15.1	21.3	47.9	23.8	31.8	40.0	23.6	29.7	28.7	21.4	24.5
16	公園	25.0	18.4	21.2	51.4	23.0	31.8	34.4	23.1	27.6	24.4	17.9	20.6
17	ペンギン	15.7	21.8	18.3	36.4	38.1	37.2	21.2	22.4	21.8	23.1	21.6	22.4
18	ラーメン	52.1	52.5	52.3	68.6	53.6	60.2	64.4	47.5	54.6	51.9	39.7	45.0
19	結婚式	40.0	22.4	28.7	42.1	35.8	38.7	34.4	28.9	31.4	28.7	24.7	26.6
20	槍ヶ岳	38.6	32.7	35.4	60.0	31.0	40.9	49.4	43.4	46.2	36.9	31.1	33.7
	平均値	26.8	31.4	28.9	34.9	53.6	42.3	34.3	37.7	35.9	28.3	31.2	29.7

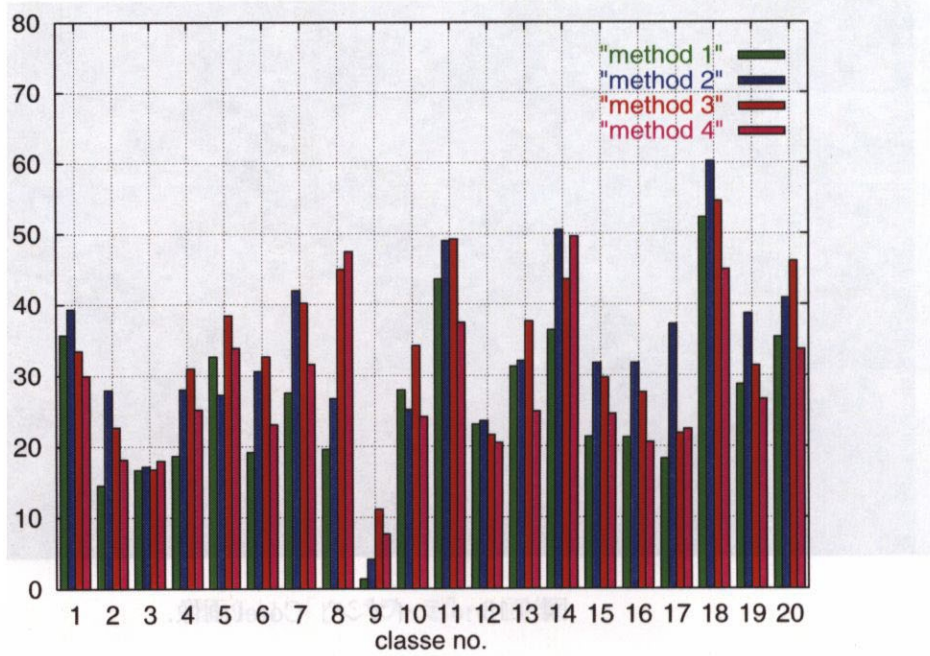


図 9.14 実験 4 (20 クラス, 不適合画像あり) の分類実験の結果.

表 9.7 実験 5 (20 クラス, 不適合画像なし) の分類実験の結果.

no.	クラス名	方法 1			方法 2			方法 3			方法 4		
		再現率	適合率	F 値	再現率	適合率	F 値	再現率	適合率	F 値	再現率	適合率	F 値
1	りんご	37.5	60.0	46.2	30.0	96.0	45.7	30.0	66.7	41.4	32.5	60.5	42.3
2	クマ	17.5	20.6	18.9	18.8	93.8	31.2	18.8	29.4	22.9	23.8	30.2	26.6
3	マウンテンバイク	15.0	14.3	14.6	6.2	100.0	11.8	7.5	17.6	10.5	7.5	13.3	9.6
4	琵琶湖	43.8	26.7	33.2	43.8	44.9	44.3	37.5	29.4	33.0	46.2	33.3	38.7
5	乗用車	27.5	37.3	31.7	27.5	78.6	40.7	36.2	48.3	41.4	37.5	52.6	43.8
6	ネコ	36.2	28.2	31.7	68.8	19.4	30.2	57.5	25.3	35.1	41.2	21.6	28.3
7	入学式	35.0	25.5	29.5	61.3	39.2	47.8	63.7	34.7	44.9	55.0	35.2	42.9
8	住宅	6.2	55.6	11.2	1.3	100.0	2.6	38.2	67.4	48.7	28.9	57.9	38.6
9	イチロー	2.5	66.7	4.8	6.4	100.0	12.0	7.8	75.0	14.1	3.9	33.3	7.0
10	観覧車	20.0	88.9	32.7	16.2	100.0	28.0	33.8	84.4	48.2	26.2	48.8	34.1
11	金閣寺	66.2	37.9	48.2	48.8	86.7	62.4	56.2	62.5	59.2	46.2	44.0	45.1
12	ライオン	17.9	73.7	28.9	23.0	100.0	37.3	19.7	52.2	28.6	26.2	48.5	34.0
13	モアイ	15.0	44.4	22.4	11.2	100.0	20.2	22.5	51.4	31.3	13.8	28.2	18.5
14	ノートパソコン	26.2	50.0	34.4	26.2	87.5	40.4	32.5	76.5	45.6	37.5	68.2	48.4
15	新幹線	43.8	16.0	23.4	42.5	34.3	38.0	33.8	26.7	29.8	40.0	31.4	35.2
16	公園	35.0	19.6	25.1	75.0	17.0	27.8	52.5	22.1	31.1	47.5	19.6	27.7
17	ペンギン	11.2	13.4	12.2	30.0	52.2	38.1	25.0	26.7	25.8	22.5	28.6	25.2
18	ラーメン	42.5	50.7	46.3	70.0	40.6	51.4	67.5	38.8	49.3	56.2	36.3	44.1
19	結婚式	42.5	22.5	29.4	47.5	37.3	41.8	40.0	30.8	34.8	33.8	24.1	28.1
20	槍ヶ岳	38.8	35.2	36.9	58.8	28.3	38.2	58.8	44.3	50.5	42.5	37.0	39.5
	平均値	29.0	39.4	33.4	35.7	67.8	46.7	37.0	45.5	40.8	33.5	37.6	35.4

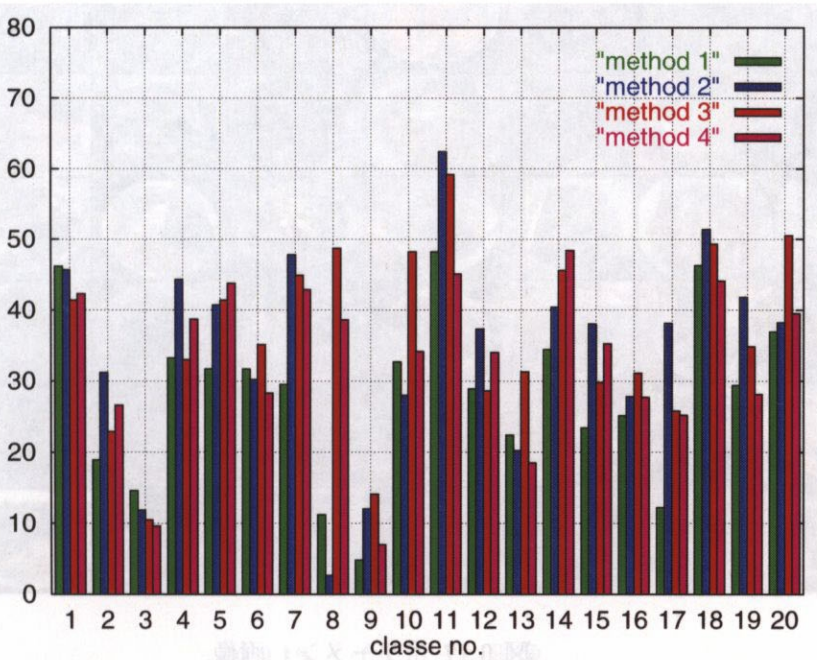


図 9.15 実験 5 (20 クラス, 不適合画像なし) の分類実験の結果.