

意味を表現する空間上に、 検索結果の画像群を表示する可視化インタフェース

武者 義則, 広池 敦

新情報処理開発機構 情報ベース機能日立研究室

〒185-8601 東京都国分寺市東恋ヶ窪一丁目280 (株)日立製作所 中央研究所内

あらまし 検索結果を“意味軸”上の散布図として表現する類似画像検索システムを開発した。この“意味軸”とは、あるキーワードについて画像特徴量から推定される適合性を、その画像の座標とする軸である。あらかじめ画像に添付されたキーワードとその画像特徴量との間を重回帰分析により学習しておき、検索結果の画像間の違いを最も表現するキーワードを自動選択して、ユーザへ提示する。ユーザが提示されたキーワードから2つ選ぶと、検索された画像を、2つのキーワードの推定値を座標として可視化空間上に配置し、散布図として表現する。本稿では、類似画像検索の結果の画像群に対する未学習データのキーワード推定を行い、その性能を評価したこと、および開発した2次元可視化インタフェースについて報告する。

Visualization interface displaying the retrieved images in a 2-D semantic space

Yoshinori MUSHA, Atsushi HIROIKE

Information-Base Functions Hitachi Laboratory, RWCP

{sha, he}@crl.hitachi.co.jp

Abstract We are developing a similarity-based image retrieval system which represents retrieved images as a scatter diagram in a *semantic space*. An axis of the space shows suitabilities of a keyword assigned to the images. The suitabilities are estimated by linear transformation of image features of retrieved data, and the coefficients of the transformation are learned by multiple regression analysis. The system performs in the followings. First, a user makes image retrieval. Then the system presents the user some keywords that give large variances of the estimated suitabilities among retrieved results. Next, the user can choose two axes from the given keywords. Lastly, the system displays the retrieved images in the *semantic space* spanned by the axes. We examined our method quantitatively by using a large number of samples, each of which is a pair of an image and assigned keywords.

1 はじめに

近年、画像コンテンツビジネスやマルチメディア情報を扱うデータベース業界での需要を背景に、画像自体の類似性に基づく画像検索に関する研究が盛んに行なわれている [3, 4]。特に、検索システムに人間の判断を反映させるために、人間の感性や好みに基づく画像検索の研究 [7, 13] などが行なわれている。しかし、検索システムの出力結果において、ユーザにとって望みの画像であると判断するのは、そのユーザであるべきで検索システム側であっては

ならない。また、そもそも画像を検索しようと試みているユーザの判断自体が曖昧であることも大きな問題となっている。つまり、類似画像検索システムにとって、ユーザの感性や判断基準に則した適切な類似性判断ができるだけでなく、ユーザの曖昧な検索意図を明確化するために、ユーザの試行錯誤を支援する機能が重要である。

以上の観点から、我々は類似画像検索の結果を可視化表示するインタフェースの開発に力を入れてきた。我々の可視化インタフェースは、類似画像の検

索結果の画像群において違いを最も表現する可視化軸を構成し、2次元空間上の縮小画像の散布図として [10]、あるいは3次元空間の時空間パターンとして [6]、画像群を表現する。できるだけ多くの画像や検索された画像群の多様な類似性をユーザに提供し、ユーザを支援することを目的としている。

近年、我々の表現手法と同様の2次元可視化を行う研究 [11] がある。検索により画像群が決まると、多次元尺度構成法によって一意に決定される画像配置を、ユーザへ提供するインタフェースとなっている。大量の画像を可視化すると計算量が急増する問題点や、多様であるはずの画像間の類似性判断の1つのみをシステムがユーザへ提示するインタフェースである点が、我々の目的に沿わないが、可視化に人間の色判別能力が反映されるという特徴がある。

一方、我々は、システムが判断の主体ではなく、ユーザの認識能力を最大限に引出すことによって、ユーザの判断を支援することが一つの目標である。以前、検索結果の画像群の特徴量を主成分分析し、その主軸を可視化軸とする2次元可視化システムを開発した [10]。しかし、その主軸をユーザが選択するというインタフェースにおいては、ユーザへ主軸を選ぶ動機付けを与えにくかった。だからこそ、ユーザが積極的に可視化軸を選択できる“意味軸”の導入に大きな意義がある。また、画像間の多様な類似性を提示するという我々の目的からも、意味的な指標による可視化には意義がある。

以上の観点に従って、我々は、画像特徴量からその画像に関連するキーワードの適合性を推定し、その推定値に応じて画像を配置する2次元可視化システム [9] を開発した。キーワードは画像特徴量から推定されるため、配置される個々の画像データはテキスト情報を持つ必要がない。

これに関連して、画像情報とテキスト情報を検索に統合的に使用する研究 [1, 2] を挙げることができる。インタラクティブに画像検索とテキスト検索を繰り返すことにより効果的な絞り込みを行うが、被検索データはテキスト情報と画像情報の両方を持っている必要がある。そして、テキスト検索時にはテキスト特徴量のみを、画像検索時には画像特徴量のみを用いるため、テキスト情報が少ない、若しくは存在しない画像データベースに、これらの手法を適用するのは不適切である。

本稿では、画像特徴量からのキーワード推定の性能を評価したこと、および、推定されたキーワード

の適合性が画像の座標となる2次元可視化システムについて報告する。

2 画像特徴量

本稿では、以下に示す特徴量を類似画像検索とキーワード推定に用いた。

2.1 RGB色特徴量

画像の色分布を反映する色特徴量を構成した。RGB色空間の各軸R,G,B軸をそれぞれ N_c 等分して、 N_c^3 個のボクセルを構成する。画像の各画素の色に従って、これらボクセルに投票することで各ボクセルの画素頻度を算出し、色ヒストグラムを構成する。これを全画素数で正規化することで特徴抽出を完了する。さらに、画像の全領域を縦と横それぞれ N_s 分割してできる N_s^2 領域のそれぞれについて上記の特徴抽出を行う。ここで、 $N_c = 4$ 、 $N_s = 4$ を採用すると、1,024次元のRGB色特徴量が構成される。

2.2 微分方向特徴量

画像の形やテクスチャを反映する特徴量として微分方向特徴量を構成した。まず180度を N_d 等分しておき、方向毎の N_d 個のピンを用意しておく。カラー画像を輝度画像に変換し、各画素につきx軸、y軸の微分演算を行い勾配ベクトルを計算する。その勾配ベクトルの成す角度に対応するピンに、その勾配ベクトルの2乗ノルムを加算する。これを全画素について行うことで各ピン中の2乗ノルムの総量を算出する。そして全画素数で正規化し平方根をとることで特徴抽出を完了する。さらに、解像度が $1/2^0, 1/2^1, \dots, 1/2^{N_r-1}$ の N_r 階層の輝度画像を用意し、それぞれの全領域に対して、更に縦と横それぞれ N_s 分割して N_s^2 領域を作成する。この解像度と領域分割で用意された $N_r N_s^2$ 画像に対してそれぞれ上記の特徴抽出を行う。ここで、 $N_d = 8$ 、 $N_r = 4$ 、 $N_s = 4$ を採用すると、512次元の微分方向特徴量が構成される。

3 キーワード推定

検索結果を可視化する際、検索結果毎にダイナミックに処理を行うために、高速処理が可能な重回帰分析を採用した。しかし、このような線形処理では、排他的論理和のような入出力関係のものは推定できないように、本質的に推定できないキーワードも存在すると思われる。そこで、画像特徴量からの

キーワード推定に対してどれほどの性能をもたらすかを評価した。本システムは、(I) あるキーワードが添付されている画像 (正解画像) の推定、(II) 任意の画像群全体を表現するキーワード (代表キーワード) の推定、(III) 任意の画像群内の違いを表現するキーワード (可視化キーワード) の推定、の3機能を持つ。(I) は画像の可視化に使用され、(II)(III) はユーザへ画像群を特徴付けるキーワードを提供するのに使用される。特に (III) は可視化軸に設定するキーワードをユーザへ提供するための機能である。

3.1 重回帰分析によるキーワード推定

画像特徴量にて検索した画像群をキーワード的な軸で可視化表示するために、重回帰分析でのキーワード推定を検討した。重回帰分析の特徴は、線形計算であるため、画像群のキーワード推定値の平均と分散を求める場合に、その画像群の特徴量の平均と分散共分散行列を求めてから、全てのキーワードに対して適用することができる点である。これで計算量を削減することができる。また、キーワード毎に独立に計算できる点も重回帰分析の特徴と言える。

キーワードを学習する際の教師値を (1) 式とおく。これは画像 i に対するキーワード k に関する教師値である。

$$y_i(k) = \begin{cases} Y_e(k) & \text{キーワード } k \text{ が添付} \\ Y_n(k) & \text{それ以外} \end{cases} \quad (1)$$

教師値 $y_i(k)$ の与え方を2通り用意する。一つは、 $\{Y_e, Y_n\} = \{1 - P(k), -P(k)\}$ とした Model-A である。ここで、 $P(k)$ は学習時の画像群におけるキーワード k の出現頻度であり、このとき、 $y_i(k)$ の平均 $\bar{y}(k)$ は0となる。また、これは2群の判別分析に帰着することが分っている。他方は、 $\{Y_e, Y_n\} = \{1, 0\}$ とした Model-B である。

重回帰分析によって、(2) 式のように偏回帰係数が求められる。

$$b_0(k) = \bar{y}(k) - \bar{\mathbf{x}}^T \mathbf{b}(k), \quad \mathbf{b}(k) = \mathbf{C}_{xx}^{-1} \mathbf{C}_{xy}(k) \quad (2)$$

ここで、 \mathbf{x}_i を画像 i の特徴量ベクトルとすると、 \mathbf{C}_{xx} は $\{\mathbf{x}_i\}$ の分散共分散行列、 $\mathbf{C}_{xy}(k)$ は、 $\{\mathbf{x}_i\}$ と $\{y_i(k)\}$ との共分散ベクトルである。 $\bar{\mathbf{x}}$ と $\bar{y}(k)$ は、それぞれ $\{\mathbf{x}_i\}$ 、 $\{y_i(k)\}$ の平均である。さらに、各画像のキーワード推定値 $\hat{y}_i(k)$ は、(3) 式によって求められる。これは、前述の (I) に使用される。

$$\hat{y}_i(k) = b_0(k) + \mathbf{x}_i^T \mathbf{b}(k) \quad (3)$$

また、代表キーワード (II) や可視化キーワード (III) を算出する際に用いる式がそれぞれ (4) 式、(5) 式である。各キーワード毎にキーワード推定値の平均を (4) 式で計算し、その値の大きいものが検索結果画像群を代表するキーワードであると期待される。同様に、各キーワード毎にキーワード推定値の分散を (5) 式で計算すると、検索結果の画像群内の違いを表現する可視化に適したキーワードが得られることが期待される。

$$\bar{y}(k) = b_0(k) + \bar{\mathbf{x}}^T \mathbf{b}(k) \quad (4)$$

$$\hat{\sigma}_y^2(k) = (\mathbf{b}(k))^T \tilde{\mathbf{C}} \mathbf{b}(k) \quad (5)$$

ここで、 $\bar{\mathbf{x}}$ と $\tilde{\mathbf{C}}$ は画像群における特徴量の平均ベクトルと分散共分散行列である。 $\bar{y}(k)$ の値の大きなキーワードを代表キーワード、 $\hat{\sigma}_y^2(k)$ の値の大きなキーワードを可視化キーワードと定義する。

3.2 実験方法

キーワードが付属している画像コンテンツとして、PhotoDisc 社の 24,728 枚の画像を用い、画像データベースを作成した。各画像から RGB 色特徴量 (1,024 次元) および微分方向特徴量 (512 次元) を抽出し、主成分分析によりそれぞれ $N_F/2$ 次元に圧縮して、合計 N_F 次元の特徴量データベースを作成した。更に、画像データベースを奇数番目と偶数番目に2分し、奇数番目をキーワード学習に用い、偶数番目を評価に用いた。

この画像データベースには、1 画像あたり約 20 個あまりのフリーキーワードが添付されている。その総数は 21,637 個であるが、学習データである画像に付いているキーワードは 16,170 個である。その内 1 個もしくは 2 個の画像にしか添付されていないキーワードは統計処理に向かないものとして、あらかじめ除外し、8,914 個のキーワードを学習に用いることとした。

まず代表キーワードの推定性能の評価方法を説明する。Model-A、Model-B の手法に対して、ランダムに選んだ画像を検索キーとして実際に未学習の画像データを対象とした類似画像検索を行う。(4) 式から検索結果の画像群の代表キーワードの推定を全キーワードに対して行い、その推定値の大きい順にソートしておく。そして、検索結果の画像群に含まれる複数のキーワードが、ソートされた全キーワードの 1 位 ~ N 位 ($N = 2, 5, 10, 20, \dots, 10000$) の中に含まれる個数を計測する。同じキーワードが含ま

れていればその個数を得点として、その中に実際に含まれていたキーワード群の得点をその中で可能な最大得点に対する比率で評価する。これを1,000回の類似画像検索について平均し、グラフ化した(図1)。この得点は、検索結果の複数の画像に共通に多く含まれるキーワードほど高くなり、つまり代表キーワードとして相応しいことを反映する。

次に、あるキーワードにおける正解画像の推定性能の評価方法を説明する。全画像について、1つのキーワードの推定値を求めてソートしておき、そのキーワードが添付された複数の画像が、1位~N位($N = 2, 5, 10, 20, \dots, 10000$)の中に含まれる個数を計測する。そして、そのキーワードが添付された画像数に対する、N位におけるその個数の割合(再現率)を全キーワードに対して平均し、グラフ化した(図2)。また、1位~N位の中に実際に含まれていた正解画像数を、その中で可能な最大個数に対する比率で評価する。これを全キーワードに対して平均し、グラフ化した(図3)。

本システムでは、画像群の可視化を行うためのキーワードを、(5)式の値が大きな可視化キーワードからユーザに選ばせるようになっている。よって、特に、画像群におけるキーワード推定値の分散が大きなキーワードについて同様の評価を行った。前述同様に、1つのキーワードの推定値を全画像について求めソートしておく。そして、そのキーワードが添付された複数の画像が、1位~100位の中に含まれる個数を計測し、その中で可能な最大個数に対する比率で評価する。これをキーワード推定値の分散が大きい順に、K個($K = 2, 5, 10, 20, \dots, 10000$)のキーワード群に対して平均し、グラフ化した(図4)。

3.3 実験結果と考察

図1は $N_F = 100$ の場合の代表キーワードの推定性能を示している。○と△に注目すると、△“Model-B”が10位以内において97%にも達しているのに対し、○“Model-A”では63%ほどしかない。

ところで、キーワード別に出現頻度を測定してみると、未学習画像において、例えば25%以上の出現頻度のものが9個存在することが分かった(表1)。このような出現頻度の高いものは、多くの画像で上位に推定される傾向があり、ユーザにとって情報量の少ないキーワードであると思われる。そこで、これら9個のキーワードの影響を確かめるために、それ以外のキーワードの順位を保ったままその9個のキ

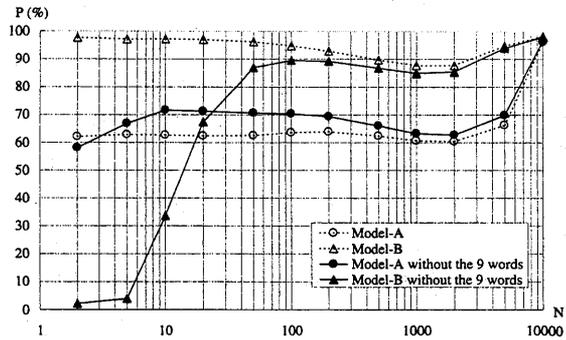


図1: ある順位までの得点比率 (%)

表1: 1/4 以上の出現頻度のキーワード

出現頻度	キーワード
69.6 %	color
66.6 %	center
53.8 %	horizontal
52.6 %	vertical
44.8 %	white
40.0 %	interior
31.2 %	exterior
28.0 %	blue
27.3 %	black

ワードを取り除いた評価を試みた。その結果が図1中の●“Model-A without the 9 words”と▲“Model-B without the 9 words”のグラフである。これによると、5位~50位の順位において、Model-Bが著しく正解率が低下しているのに対し、Model-Aは4~9ポイント上昇しているのが分かる。5位以内を比較すると、Model-Bが97%から4%へ急落しているのに対し、Model-Aでは、63%から67%へ上昇している。これにより●Model-Aと▲Model-Bでは、5位以内の点で約16倍あまりの性能差が見られる。△“Model-B”が前述の9個の高頻度キーワードが上位に入ること、高正解率を出していたのに対し、○“Model-A”は期待通り低頻度のキーワードの正解率も上位に維持していたことが分かる。

また、図1の9キーワードを除いたグラフにおいて、●Model-Aと▲Model-Bが20位~30位で反転しているが、上位のキーワードが重要であること、実際のインタフェースにて提示する代表キーワードの数が20~30個程度であることを考える

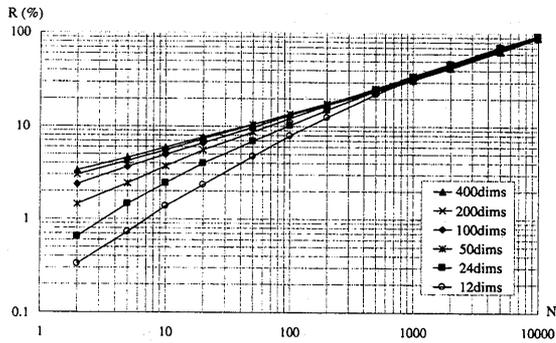


図 2: 順位 N 位以内における平均再現率 $R(\%)$

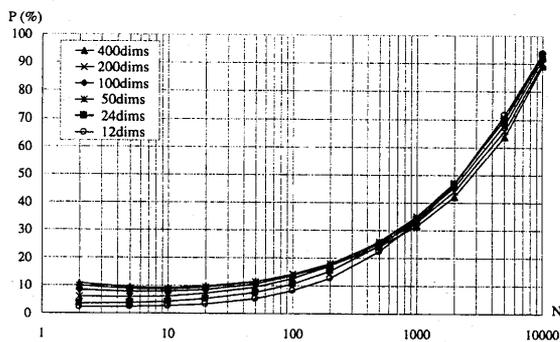


図 3: N 位までに正解画像が含まれる率 (%)

と、Model-A の方が適切であると考えられる。

図 2 は、Model-A の各 N_F における平均再現率のグラフである。次元数を多くすると再現率も向上するが、既にほとんど飽和している様子が分かる。上位 100 位以内を正解とした場合の平均再現率は特徴量 400 次元でも 13% 程度である。

図 3 は、Model-A の各 N_F における精度のグラフである。この精度は、実際の可視化において配置される画像群のキーワード推定の正確さを反映している。従って、この精度が高いほど、可視化効果が向上すると考えられる。上位 100 位以内を見ると次元数が多くなるに従い精度も向上するが、やはりほとんど飽和しているようである。100 位における特徴量 400 次元の精度でも 14% 程度である。

ところで、本システムにおいて、可視化の軸に用いるキーワードは、可視化した画像群が可視化空間上で広く散らばるような可視化キーワードからユーザが選ぶようになっている。そこで、キーワード推定値の分散が大きいキーワードから順に評価に加え

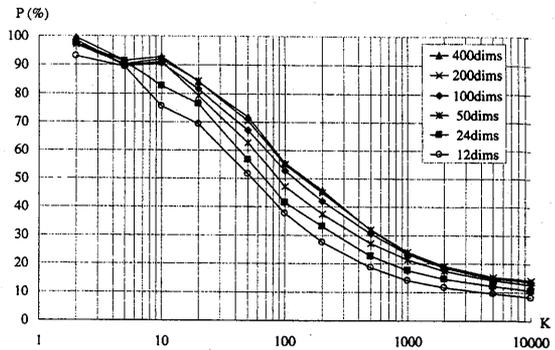


図 4: N 位までに正解画像が含まれる率 $P(\%)$: 各キーワードにおいて、100 位以内に正解画像が含まれる割合をそのキーワード推定値の分散が多い方から K 個のキーワードについて平均した。

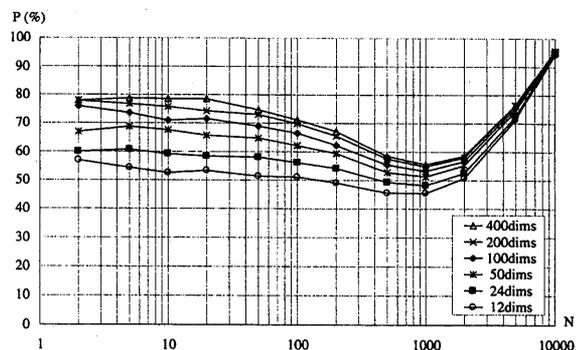


図 5: N 位までに正解画像が含まれる率 (%) : キーワード推定値の画像群における分散が大きい 50 個のキーワードのみを平均した。

ていき、精度を評価した (図 4)。これによると、分散が大きなキーワードのみで評価した精度の方が高い数値を出す傾向がある。全キーワードで平均した精度は、400 次元でも 14% 程度であるが、上位 10 個のキーワードでは約 92% になっている。ただし、この結果は全画像に対する分散であり、類似画像検索の結果の画像群に対する分散とは必ずしも一致しないが、1つの目安を与えるものと考えられる。

上位 50 個のキーワードについて、図 2 と同様に再現率を評価すると、上位の順位についてわずかに低下する傾向が見られたが (図示せず)、図 3 と同様に精度を評価すると、格段に向上する傾向が見られた (図 5)。20 位以内において、特徴量 400 次元で精度が約 78% となっている。

4 可視化システム

“意味軸”を可視化空間の座標軸として採用した、類似画像検索のための可視化システムを開発した。このシステムは、サーバとクライアントによって構成され、サーバでは画像検索および統計処理などの数値計算を行う。クライアントは JAVATM で記述され、主にユーザインタフェースとして画像やキーワードの提示とユーザの操作をサーバに伝える機能を持つ。本システムに特徴的な機能は主に3つあり、それぞれ第3章の機能 (I)(II)(III) に対応している。

4.1 可視化インタフェースの特徴

“意味軸”を利用した我々の2次元可視化インタフェース(図6)には、以下の(a)~(j)の機能がある。

図6(a)は、可視化空間であり、図のように検索結果の縮小画像を、キーワード推定値に対応した座標へ配置する機能を持つ(I)。可視化空間の座標軸の性質を画像自身によって示すことでより直感的な理解を促すと期待される。マウスカーソルで次の検索のキー画像を指定し、直に検索を実行することができる。また、矩形領域を指定することで、絞り込んだ画像群を可視化表示する機能も持つ。他に、マウスカーソルを任意の画像に合せてしばらく待つと、その高解像度な画像が表示される機能や、検索された画像群が見易くなるよう背景色を自由に設定する機能がある。また、原点を中心としてx軸とy軸が目盛りのついたスケールが表示される。これはキーワード推定値を測る目安となる。

図6(b)と(c)の文字列表示部は、検索結果の画像群の違いを表現するキーワード(可視化キーワード)を提示する機能である(III)。一般に、ユーザが任意にキーワードを選ぶ機能があっても、実際に適切なキーワードを選ぶことは、キーワードが多いほど困難である。そこで、可視化に効果的なキーワードの候補を画像群毎に用意するのが本機能の目的である。あらかじめ登録されているキーワードの中から、キーワード推定値のその画像群における分散が高いものが選ばれ、そのいくつかがユーザに提示される。各軸、1度に5個のキーワードが表示される。(b)の上下、(c)の左右の矢印ボタンを押すことで、5個単位でキーワードを入れ替えることができ、キーワード推定値の分散が大きい方から合計100個のキーワードを提示することができる。図6(b)と(c)の文字列表示部はボタンでもあり、提示された100個のキーワードから、可視化軸として2つのキーワ

ードを選ぶ機能を提供する。ここで選択したキーワードに関して、画像群の可視化表示が行なわれる。

図6(d)は、スクロールバーの連続的な操作で、可視化空間の拡大や縮小をする機能である。表示画像の大きさを一定にしたまま空間だけが伸縮するため、画像の密度が変更される。拡大表示すると、例えば、キーワード推定値が近接しているため重なっている画像を分離することができる。

図6(e)は、可視化空間を移動する機能である。(d)の拡大縮小機能と組み合わせると表示画像群の任意の領域に焦点を当てて詳細構造を見ることができる。

図6(f)は、データベース中の画像データ名を指定することで、検索キーに使う例示画像を呼び出す機能である。画像データ名は、メニュー表示したディレクトリの中から選ぶことができる。呼び出した例示画像群も検索画像群と同様にキーワードによって可視化表示される。

図6(g),(h)は、画像検索の類似度と表示画像のサイズとの対応関係を設定する機能である。これにより画像の大きさで、その類似性を表現することができる。図6(g)の横軸は検索結果の類似度、縦軸は変換された類似度を示す。(g)に示されているヒストグラムは、検索結果の類似度の分布を示している。また、四角いボタンは、横軸から縦軸への射影を指定する連結直線の連結部を示す。この連結部の数や位置は任意に指定できる。検索数を限った場合、一般に検索結果の類似度分布が偏る傾向がある。ヒストグラムを見てその範囲内に直線の傾斜が収まるように指定すると、類似度の変化を敏感に画像サイズに反映させることができる。また、図6(h)は、上記で変換された類似度を使って画像サイズを実際に決定するための、2つのパラメータを指定するパネルである。横軸はフォーカス“Focus”(F)、縦軸は密度“Density”(D)に対応している。変換された類似度(R)と画像サイズ(S)との関係をこれらのパラメータで記述すると(6)式のようになる。

$$S(R) = D \cdot R^F \quad (6)$$

$$\text{ただし, } 0.0 \leq D \leq 3.0, \quad 0.0 \leq F \leq 4.0$$

Dが大きくなると、全ての検索結果の画像サイズが大きくなるため、画像群が混み合って表示される。また、Fが大きくなると、類似度に応じてその違いが画像サイズに敏感に反映され、検索キーに近いものに焦点をあててサイズを大きくしたように表示される。逆に、Fが小さくなると類似度によって画像

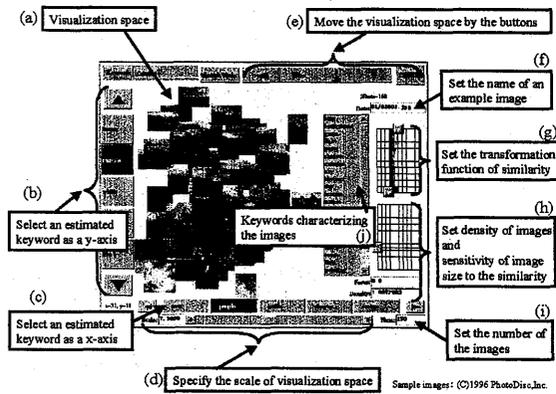


図 6: “意味軸” を利用した可視化インタフェース

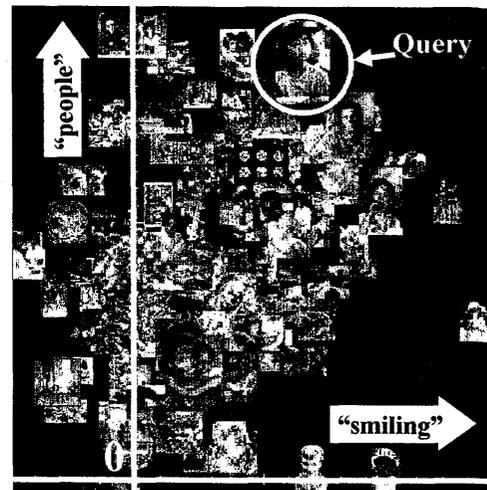
サイズが変わりにくくなる。 D や F の上限値は十分に画像サイズが調整できるように設定した。

図 6 (i) は、表示する画像数の上限を設定する機能である。例示画像を表示する場合や、類似画像検索の結果を表示する場合の表示数を制限する。代表キーワードや可視化キーワードを推定する場合の画像数も、この数が用いられている。

図 6 (j) は、検索結果の画像群を代表するキーワード (代表キーワード) をユーザへ提示する機能である (II)。あらかじめ登録されているキーワードの中から、与えられた画像群の平均特徴量に対する推定値の高いものが選ばれ、その上位 25 個がユーザに提示される。(j) のようにリスト表示されるが、このキーワードの一つをマウスでダブルクリックすると、前述の可視化キーワードに同じキーワードが含まれている場合に、x 軸 (c) にそのキーワードが現われ、かつ選択されてそのキーワードによる可視化が行われる。

4.2 可視化インタフェースの表示例

図 7 が、キーワード軸による可視化の例である。白丸で示した検索キーによる類似画像検索の結果 (150 画像) を可視化している。特徴量次元数 N_F は 100 次元を使い、Model-A によるキーワード推定で、x 軸に “smiling”, y 軸に “people” を選んだ場合を示している。図中に示した座標軸の原点は、2つの推定値が 0 となるポイントであり、第一象限の右上になるほど、2つのキーワードを備えた画像が配置されると期待される。実際には、第一象限にも “smiling” と “people” のどちらにも属さないと思われる画像



Sample images: ©1996 PhotoDisc, Inc.

図 7: “意味軸” による可視化の例

もノイズとして含まれているが、概して “smiling” と “people” の傾向を見てとることができる。ちなみに、図 3 の評価法を 1 つのキーワードに対する精度で行うと、 “smiling” と “people” のキーワード推定精度は共に約 40% であった。また、両者のキーワード推定に相関があることは、図中の画像群が右上あがりの楕円の分布を持っていることから分かる。

4.3 本システムにおけるデータフロー

本インタフェースを実現するシステム内のデータフローを図 8 に示す。まず、登録するキーワード毎に (2) 式の処理により、あらかじめキーワード偏回帰係数データベースを作成しておく。

ユーザからの検索要求 <1> に従って類似画像検索 <2> を行うと、画像特徴量データベースをスキャン <3> して検索結果 <4> を出力する。その検索結果に対応する画像特徴量 <6> から平均値と分散共分散行列が計算される <5>。キーワード偏回帰係数データベースからキーワード 1 個分の偏回帰係数を取り出し <9>、画像特徴量の平均値と分散共分散行列 <7> を使って (4) 式と (5) 式の演算を行い、キーワード推定値の平均値と分散値を出力する <10>。これを全てのキーワードについて繰り返し、最大値から上位指定個数だけキーワードをそれぞれ選択する <11>。平均値の大きなキーワードは代表キーワード、分散値の大きなキーワードは可視化キーワードである <12>。それぞれをユーザに提示することで、意味情報をユーザに提供する <13,14>。

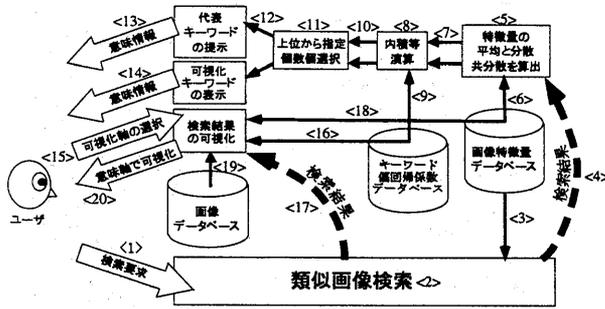


図 8: 本システムにおけるデータフロー

可視化キーワード <14> からユーザが x 軸, y 軸として, それぞれキーワードを選択すると <15>, そのキーワードの偏回帰係数 <16> と検索結果 <17> の画像特徴量 <18> から (3) 式が計算され, その値を座標値として画像 <19> を配置する。こうして“意味軸”によって可視化された画像群がユーザへ提供される <20>。

ところで, 本システムは類似画像検索の結果の画像群を対象として可視化することを想定していた。しかし, “意味軸”による可視化システムの特徴としては, 検索結果とは限らず任意の画像群に対して可視化できることに気づく。実際, 本システムは検索のキーワードを選ぶために, 初期の例示画像を表示する機能があり, その例示画像群を“意味軸”上で可視化している。

つまり, 本可視化機能は本質的に, 画像群を出力する任意の検索システム (キーワードによる画像検索, 文章による画像検索など) 上に適用することが可能であり汎用性が高い。本来, 画像の検索機能とは独立な可視化機能であると言える。

5 おわりに

類似画像検索のためのユーザインタフェースとして, 可視化の軸にキーワードを使うシステムの開発を行った。このシステムでは, 検索の結果得られた画像に対して, ダイナミックに可視化軸および代表キーワードとして相応しいキーワードを提示し, ユーザが選んだ可視化軸のキーワードに従って, 2次元の散布図として検索結果を提示する。キーワード推定に線形で高速処理が可能な重回帰分析を用い, 代表キーワードの推定性能および可視化におけるキーワード推定性能を評価した。その結果, 代表キーワードの推定には判別分析に帰着する Model-A

が相応しいこと, およびキーワード推定値の分散が大きいキーワードに限って精度を評価すると高い数値が得られることが分った。また, 本インタフェースは, 画像群を出力する任意の検索システムに適用可能であり, 汎用性が高いと考えられる。

今回, 処理の高速化のために重回帰分析を使用した, 更なるキーワード推定性能の向上に重点を置いて, 他の分析手法や特徴量を用いて性能を比較する必要がある。例えば, 画像理解の研究において, 画像の分割ベクトル量子化により画像特徴量からキーワードを推定する手法 [8] が提案されており, 本インタフェースの可視化性能を向上させる可能性がある。

参考文献

- [1] F. Chen, U. Gargi, L. Niles and H. Schutze: Multi-Modal Browsing of Images in Web Documents, *SPIE Document Recognition and Retrieval VI*, 3651, pp. 122-133, 1999.
- [2] 遠藤, 指田, 増本, 長田, 棚橋: 画像情報とテキスト情報を統合的に利用したインタラクティブな Web 検索システム, 第 5 回知能情報メディアシンポジウム, pp. 163-170, 1999.
- [3] D. S. Faulus and R. T. Ng: An Expressive Language and Interface for Image Querying, *Machine Vision and Applications*, 10, pp. 74-85, 1997.
- [4] M. Flickner, H. Sawhney, et al.: Query by Image and Video Content: The QBIC System, *IEEE Computer*, 28, 9, pp. 23-32, 1995.
- [5] A. Gupta and R. Jain: Visual Information Retrieval, *Communications of the ACM*, 40, 5, pp. 71-79, 1997.
- [6] A. Hiroike, Y. Musha, A. Sugimoto, and Y. Mori: Visualization of information spaces to retrieve and browse image data, *Third International Conference on Visual Information Systems*, pp. 155-162, 1999.
- [7] 藤山, 斎藤: ニューラルネットワークを用いた感性の学習機能を持つ画像検索システム, 第 3 回知能情報メディアシンポジウム, pp. 253-258, 1997.
- [8] 森, 高橋, 岡: 単語群付き画像の分割ベクトル量子化に基づく画像理解, 信学技報, PRMU99-47, pp. 73-80, 1999.
- [9] 武者, 広池: 類似画像検索における検索結果の可視化インタフェース—可視化軸として意味軸を用いる方法—, 信学技報, PRMU99-57, pp. 59-64, 1999.
- [10] 武者, 広池, 杉本: 類似画像検索における特徴量空間の可視化インタフェース, 信学論, J82-D-II, 10, pp. 1626-1633, 1999.
- [11] Y. Rubner, C. Tomasi, and L.J. Gauibas: Adaptive Color-Image Embeddings for Database Navigation, *IEEE ACCV98*, pp. 104-111, 1998.
- [12] Y. Takane, F.W. Young, and J. Leeuw: Nonmetric individual differences multidimensional scaling: an alternating least squares method with optimal scaling features, *Psychometrika*, 42:7-67, 1977.
- [13] 田辺, 大谷, 石井: 多次元心理空間を用いる類似画像検索法, 信学論, J75-D-II, 11, pp. 1856-1865, 1992.