

# 階層的分類を用いた視覚感性のモデル化と類似画像検索への応用

多田 昌裕<sup>†</sup> 加藤 俊一<sup>††</sup>

本稿ではデータベース内画像データの階層的分類と判別分析を組み合わせた視覚感性のモデル化手法と類似画像検索への応用について論じる。我々は視覚の生理的レベルの特徴抽出機構をまねた3点間コントラストを定義し、これと人間の直感的な画像の分類過程とマッチした階層的なデータベースの分類法を組み合わせた類似画像検索手法を開発した。階層的に分類されたデータベース内のどの画像群が例示画像と最も類似しているかを判定し、その群のみを検索対象として類似画像検索を行うことで、高い検索精度と計算時間の低減を実現した。

## Similarity Image Retrieval System Using Step-by-Step Hierarchical Classification

MASAHIRO TADA<sup>†</sup> and TOSHIKAZU KATO<sup>††</sup>

In this paper, we proposed a method to model user's subjective criterion for similarity by using multivariate analysis and hierarchical classification of database. We defined a tri-contrast feature on color which approximates lateral inhibition in human's early vision process. Using this graphical feature, we developed a hierarchical classification algorithm and a similarity image retrieval algorithm which are adjustable to subjective and intuitive criteria in similarity evaluation. We have adopted these algorithms for large image database including various kinds of contents such as textures, landscapes, paintings and city images. By selecting the most similar group in hierarchically classified database to the given visual image key step by step as the index space, we achieved both better precision and less computation at the same time.

### 1. はじめに

現在の検索ポータルサイトで提供されている画像検索システムの多くは、画像に付加された索引語とユーザが入力したキーワードとのマッチングによるテキスト型検索 (Text-based Image Retrieval) を採用している。しかしテキスト型検索には次のような問題点がある。

- A) テキスト型検索では、同一の索引語を持つ類似画像の候補グループを検索することはできるが、画像内容の類似度に基づき検索結果を順位付けることができない。
- B) グラフィックシンボルやテクスチャ、デザインパターンといった画像を単純な索引語で的確に表現することは非常に困難である。

- C) 画像データは本質的に多義性を有するため、画像から受ける印象やその解釈は、各ユーザの経験・知識により異なる。そのため、第三者により画像に付加された索引語とユーザの主観的解釈によるキーワードとがつねに合致するとは限らない。また一度作成された索引は、時間の経過にともなうユーザの興味の変遷には追従できない。

一方、画像類似検索における別のアプローチとして、キーワードによらず、画像特徴量 (GF) を用いて画像の内容に基づいた索引生成を行い、例示画像を検索キー (Visual Key) として類似画像の検索を試みる内容型検索システム (Content-based Image Retrieval System: 図1) が研究されている<sup>1)~3)</sup>。画像特徴量 (GF) とは画像の特徴を数量化し、多次元ベクトルとして表現したものである。

ところで、写真などの画像を主観的に分類したり、主観的な類似度に基づいて並べたりした場合、ユーザによって結果の異なる場合がある。これは主観的な判断基準がユーザごとに異なるためと考えられる。そのため我々は各ユーザの視覚感性をモデル化し、ユーザ

<sup>†</sup> 中央大学大学院理工学研究科  
Graduate School of Science and Engineering, Chuo University

<sup>††</sup> 中央大学理工学部  
Faculty of Science and Engineering, Chuo University

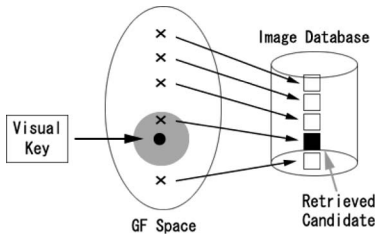


図 1 内容型検索システム  
Fig. 1 Content-based image retrieval system.

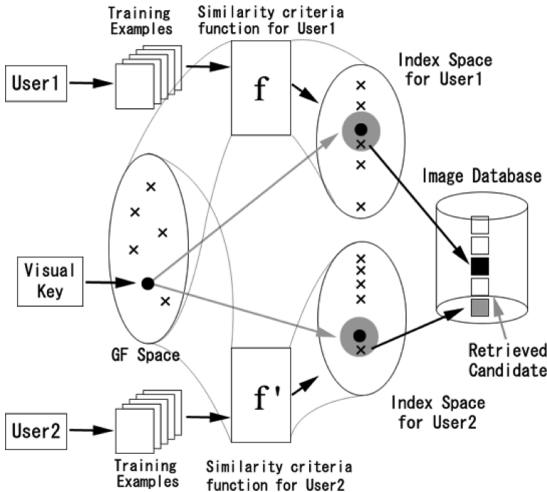


図 2 視覚感性を反映した内容型検索システム  
Fig. 2 Framework of similarity image retrieval based on subjective criteria of each user.

の主観的な画像類似度の尺度に基づいて類似画像検索を行うシステムを開発してきた(図 2)<sup>4)</sup>。本稿で我々は、データベース内画像データの階層的分類と判別分析を組み合わせた視覚感性のモデル化手法を提案する。我々の手法では、まず検索キーとなる例示画像 (Visual key) がユーザの視覚感性に基づいて階層的に分類されたデータベース内のどの群に属するのかを判別分析を用いて段階的に判定し (粗検索)、最終的に決定された群のみを検索対象として詳細検索を行う。検索対象とした群に対してのみ、例示画像に最適な検索空間を動的に生成することで、検索精度の向上と検索時間の短縮を実現した。

## 2. システム概要

### 2.1 視覚感性

視覚心理の分野では、視覚の知覚過程をモデル化する手法として Template theories, Feature theories, Prototype theories, Gestalt theory などが提案されている<sup>14)</sup>。これらの考えに基づき、さまざまな実験

が行われているが、個人によって異なる感性に着目した実験はそれほど報告されていない<sup>14)</sup>。

そこで本研究では、各人各様の視覚感性をモデル化する手法を提案し、実験によりそのモデルの有効性を評価する。

視覚の知覚過程には、特徴抽出機構により対象から抽出した特徴を手がかりに、対象を分類し認識する過程があるといわれている<sup>7),14)</sup>。我々は対象から特徴抽出を行う過程を生理的レベルの感性と考え、また、抽出された特徴をとりまとめて対象を分類する過程を心理的レベルの感性と考えている<sup>10)</sup>。

生理的レベルでの視覚感性をモデル化するためには、視覚の知覚過程が持つ種々の特徴抽出の神経回路を抽象化・単純化し、その振舞いをシミュレートできるようにする必要がある。我々は生理的レベルの感性の工学的なモデル化として、映像から明暗や色彩の平均、エッジ、コントラストなど、種々の特徴を抽出し、視覚情報をパラメータ化している<sup>12)</sup>。このような画像特徴量 (GF) を用いた生理的レベルの視覚感性モデル化手法については 3 章で詳述する。

一方、人間は同一の目の構造を持ち、特徴抽出機構の動きも同様であるにもかかわらず、主観的な画像の類似度判定基準は各人各様である。この違いは、人によって画像中の注目部位や、重視する特徴が異なる (抽出された特徴をとりまとめて画像を分類する過程が異なる)、つまり心理的レベルの視覚感性が異なることに起因すると考えられる。このような心理的レベルの視覚感性をモデル化するためには、ユーザが示す主観的な判断事例を通して特徴パラメータへの重み付けを推定する必要がある<sup>13)</sup>。データベースの階層分類と判別分析を組み合わせた心理的レベルの視覚感性モデル化手法については 4 章で詳述する。

### 2.2 システムの概要

本研究における類似画像検索システムの概略を図 3 に示す。本システムは大別すると視覚感性モデルの教示部 (Training Section) と類似画像検索部からなる。

感性モデル教示部では教示用データ (教師データ: 学習用サンプル画像に対する分類結果) を用いて判別分析を行い、心理的レベルの感性モデルを構築する。次に構築された心理的レベルの感性モデルを画像データベース全体に適用し、自動的にデータベースの階層的分類を行う。

類似画像検索部では検索キーとなる例示画像 (Visual Key) が提示されると、まず例示画像から画像の特徴を数値化する画像特徴量を抽出する (3 章で詳述)。

次に粗検索として、階層的に分類された画像データ

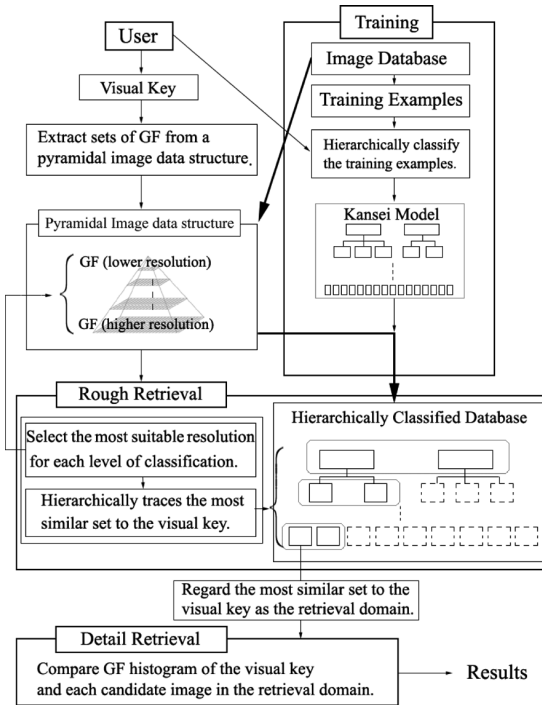


図3 システムの概略

Fig. 3 System architecture.

ベース内のどの画像群が例示画像と最も類似しているのかを判別分析により判定する。そして例示画像と最も類似していると判定された画像群のみを検索対象として詳細検索を行い、最終的な検索結果を返す(4章で詳述)。

### 3. 画像特徴量

#### 3.1 局所の特徴量の設計

画像の特徴を数量化する手法としては、カラーヒストグラムがよく知られている。カラーヒストグラムは画像全域にわたる色彩の分布をヒストグラム化したもので、画像の全域的な特徴を表現することができる。

しかし現実には、視覚の知覚過程(生理的レベルの感性)には網膜に映る映像から局所的、全域的な明暗や色彩の特徴を抽出する神経回路が存在することが知られている<sup>7),14)</sup>。この神経回路によって抽出された特徴を統合することで人間はテクスチャや形状を知覚している(心理的レベルの感性)。したがって視覚の知覚過程をシミュレートするためには、画像から全域の特徴量のみでなく、近傍画素間の局所的な関係を表現する局所の特徴量も設計・抽出する必要がある。

画像平面  $P$  上の参照点  $r$  と参照点周りの  $N$  個の変位  $(a_1, \dots, a_N \in R^2)$  との関係を測定する方法は無数

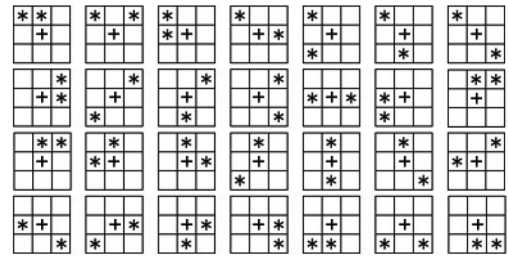


図4 コントラストを測定するパターン

Fig. 4 Local mask patterns for graphical features (GF).

にあるが、画像データでは一般に近接画素間の関係の方が重要であると考えられる。したがって本研究では局所の特徴量を測定する範囲を参照点  $r$  周りの局所的な  $3 \times 3$  画素の領域に限定する。また、画像データが持つ多様な情報を的確に表現するためには画像の直線的な変化の方向性だけでなく曲率も表現できなければならない。そのため我々は変位を  $N = 2 (a_1, a_2 \in R^2)$  とすることで画像の曲線の変化にも対応できるようにした。図4に本研究で採用したコントラストを測定する28種類のパターンを示す。図中の“+”は参照点  $r$ ，“\*”は変位  $a_1, a_2 \in R^2$  を示す。

同一の被写体を扱った画像であっても、位置ずれなどにより構図は画像ごとに異なる。特にテクスチャ画像を扱う場合、構図の変化が画像特徴量に影響を与えるのは好ましくない。そのため、画像特徴量は位置不変であることが望ましい。

位置不変な特徴量としては自己相関関数がよく知られており、大津ら<sup>11)</sup>は参照点  $r \in P$  周りの変位  $(a_1, \dots, a_N \in R^2)$  に対して、

$$y^N(a_1, \dots, a_N) = \int_P f(r)f(r+a_1)\dots f(r+a_N)dr, \quad (1)$$

で定義される高次自己相関関数を用いて顔画像判別<sup>5)</sup>やテクスチャ画像判別<sup>6)</sup>を試みている。 $f(r)$ は参照点  $r$ の輝度値である。しかし、高次自己相関関数は対象点  $r$ の輝度値の積を画像平面全体に対して足し合わせるためダイナミックレンジが広く、ノイズの影響を受けやすい。そこで我々は、「人が検知できる最小の刺激量の差は刺激強度に比例する」というウェーバーの法則を考慮し、3点間コントラストを次式で定義した。

$$C^{(3)}(a_1, a_2, r) = \frac{\{g(r+a_1) - g(r)\} + \{g(r+a_2) - g(r)\}}{|g(r+a_1)| + |g(r+a_2)| + 2|g(r)|}. \quad (2)$$

$g(r)$ は参照点  $r$ の色彩であり、式(2)の分母は視神経への刺激の強度、分子は刺激の差分である。 $S$ を刺

刺激強度とすると、式 (2) は  $\Delta S/S$  の計算と同義である。3 点間コントラストは刺激強度で正規化しているため、ノイズに強いという特性を持つ。

3 点間コントラストを測定する色空間の軸には、明暗と色の独立性を考慮して明度 I および色差 R-G, Y-B を採用した。

### 3.2 本研究で用いる画像特徴量

本研究では色空間を構成する I, R-G, Y-B の 3 軸上で 3 点間コントラストを測定、各々について次式で定義する平均  $m$  と分散  $v$  を算出し、粗検索における局所の特徴量とする。

$$m(\mathbf{a}_1, \mathbf{a}_2) = \frac{1}{N} \int_P C^{(3)}(\mathbf{a}_1, \mathbf{a}_2, \mathbf{r}) d\mathbf{r}, \quad (3)$$

$$v(\mathbf{a}_1, \mathbf{a}_2) = \frac{1}{N-1} \int_P \{C^{(3)}(\mathbf{a}_1, \mathbf{a}_2, \mathbf{r}) - m(\mathbf{a}_1, \mathbf{a}_2)\}^2 d\mathbf{r}. \quad (4)$$

ここで  $N$  は画像平面  $P$  上の全画素数である。式 (3) は

$$K \int \frac{1}{S} ds = K \times \log S + C, \quad (5)$$

の計算と同義となり、ウェーバー・フェヒナーの法則

$$I = K \times \log S + C, \quad (6)$$

に近似的に従う。I, K, S, C はそれぞれ感覚強度、定数、刺激強度、定数である。

粗検索における全域的特徴量としては、同じく色空間を構成する 3 軸上で画像の平均色および分散を計測した。

詳細検索においては、画像全域にわたる 3 点間コントラストの分布をヒストグラム化して (GF ヒストグラム) 局所の特徴量とする。また、全域的特徴量としてはカラーヒストグラムを用いる。

### 3.3 画像ピラミッド

局所的な特徴量を高解像度画像から抽出した場合、利用者が認識し得ないほど細かな部位の情報しか含まない可能性がある。逆に解像度を落としすぎると、利用者が主観的に類似度を判定する際に重視する部位の情報が欠落してしまい、実用に耐えない可能性もある。また、最適解像度は一意に決まるものではなく対象によっても変化する。つまり、ある解像度から抽出した画像特徴量は A 群と B 群を判別する際には有用でも、A 群と C 群の判別には何ら寄与しない可能性がある。

図 5 に示す画像ピラミッドは、高解像度の画像から低解像度の画像までの複数の異なる解像度の画像の集合である<sup>8)</sup>。画像ピラミッドの各画像から画像特徴量の組を抽出し、判別対象群に応じて使い分けることで、各々の判別に最適な解像度の画像から画像特徴量

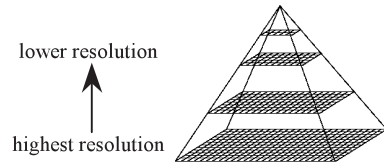


図 5 画像ピラミッド

Fig. 5 Pyramidal image data structure.

を抽出することができる。本研究では 4 種類の解像度 (1/1, 1/2, 1/4, 1/16) の画像で構成される画像ピラミッドから画像特徴量の組を抽出する。

## 4. 階層的分類

### 4.1 階層的分類を行う利点

2.1 節で定義した心理的レベルの感性 (各ユーザが画像の類似度を主観的に判定する過程) をモデル化するためには、次の要件を満たす必要がある。

- A) データ数：対象とする画像数が、統計的に意味のある解析が行えるくらいに、十分に多くなければならない。
- B) 主観的基準：対象画像に対し、各ユーザの主観的な類似度判定基準をルールとして与えなければならぬ。
- C) 感性の経時的な変化への対応：人間は経験し習熟することで、主観的な画像類似度判定基準が変化する。感性モデルをこのような経時変化に対応させるためには、長期間にわたり多量の教示用データ (教師データ) を与える必要がある。

対象画像にユーザの主観的な類似度判定規準を与える 1 つの方法としては、類似度行列が考えられる。類似度行列は 2 枚の画像を一对として、すべてのデータ対に対してユーザが自らの主観的基準に基づき、その類似度を 0 (まったく似ていない) から 1 (非常によく似ている) までの連続量で評価したものである。しかしユーザが自らの主観的基準を与えるために、対象とする大量の画像データを観察し、長期にわたってデータ対に類似度を与える心理的・身体的な負担は非常に大きい。したがって、主観的な画像類似度判定基準を類似度行列の形で教示するのは現実的ではない。

そこで我々は、ユーザの主観的基準にのっとって画像を段階的、階層的にグループ分けすることで教示を行い、判別分析を用いてモデル化する階層的分類法を提案する。この手法を採用する利点としては次の 3 点があげられる。

- A') 段階的に類似度の規準を順次与えることができるので、総合すると類似度行列を与えたような効果が期待できる。また、粗い特徴が類似している

群から細かい特徴が類似している群へと段階的に判定を行うため、ユーザが主観的に判断する際に評価する主要な特徴の変化をたどりやすく、モデル化しやすい。

- B') 複数解像度(画像ピラミッド)から抽出した画像特徴量の組を判別対象群に応じて使い分けることができ、検索精度の向上が期待できる。
- C') 階層的分類法では教示用データに対し、一度にすべての判定基準を与えるのではなく、ユーザが重要視する判定基準から順次、階層的に与えていく。そのため、ユーザの興味の変遷にともなって判定基準の重要度が変化しても、その重要度に合わせて必要最小限の分類の教示や再学習が可能である。

以上の理由により、本研究では階層的分類と判別分析による群判別を組み合わせたモデル化手法を採用する。

4.2 判別分析

今、教示用画像の画像特徴量ベクトル集合  $X = \{x_1, \dots, x_N\}$  は、あらかじめユーザの主観的判断基準に基づいて  $K$  個の群  $X^{(i)} = \{x\} (i = 1, \dots, K)$ ,  $\bigcup_{i=1}^K X^{(i)} = X$  に分類されており、各画像特徴量ベクトルには、どの群に属しているかの情報が与えられているものとする。このとき、画像特徴空間上の  $X^{(i)}$  と例示画像との間のマハラノビス汎距離  $D_i^2$  は、

$$D_i^2 = (x_0 - \bar{x}_i)^T \Sigma_i^{-1} (x_0 - \bar{x}_i), \quad (7)$$

で定義される<sup>9)</sup>。  $x_0$ ,  $\bar{x}_i$ ,  $\Sigma_i^{-1}$  はそれぞれ、例示画像の画像特徴量ベクトル,  $X^{(i)}$  の重心ベクトル,  $X^{(i)}$  の群内分散・共分散行列の逆行列である。

マハラノビス汎距離による判別では、画像特徴空間上で例示画像と各群までのマハラノビス汎距離  $D_i^2 (i = 1, \dots, N)$  を計算し、  $\min\{D_i^2 \mid i = 1, \dots, N\} = D_k^2$  となる群  $k$  に例示画像が属すると判定する。

マハラノビス汎距離を用いた判別法には、よく知られている線形判別法に比して逐次学習が容易であるという特徴がある<sup>9)</sup>。判別結果に自分の主観的判断基準が十分に反映されていないと感じた際の追加学習や、教示用データの入れ替えによる再学習が容易であるため、階層的分類法の特徴である教示用データの追加・入れ替えの容易性とあいまってユーザの視覚感性の経時的変化に対応した感性モデルの再構築を少ない労力、コストで実現することができる。

4.3 データベースの階層的分類アルゴリズム

データベースの階層的分類アルゴリズムは、以下のとおりである。

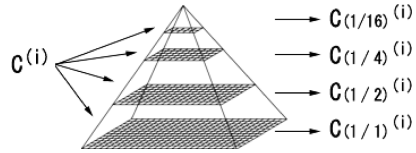


図 6 複数解像度の画像生成

Fig. 6 Construction of a set of pyramidal image data structure.

- (1) 教示用画像データの全体集合  $U = \{g_1, \dots, g_N\}$  を分類対象集合  $C$  とする。  
 $N$ :  $C$  に含まれる教示用画像の枚数
- (2)  $C$  をユーザの主観的な画像類似度判定基準に基づき、任意の  $K$  グループ  $C^{(1)}, \dots, C^{(K)}$ ,  $\bigcup_{i=1}^K C^{(i)} = C$  に分割する。
- (3)  $C^{(i)} (i = 1, \dots, K)$  に画像ピラミッドを適用し、  $C_\alpha^{(i)} (i = 1, \dots, K; \alpha = 1/1, 1/2, 1/4, 1/16)$  を生成する(図6)。
- (4)  $C_\alpha^{(i)}$  に属する各画像から画像特徴量(GF)ベクトル(要素数  $d$ )を抽出し、  $C_\alpha^{(i)}$  に対応した GF ベクトル集合  $X_\alpha^{(i)}$  を生成。  $X_\alpha^{(i)}$  の重心および群内分散・共分散行列  $\Sigma_\alpha^{(i)}$  を計算する。
- (5) (2)~(4) で構築した感性モデルを  $X_\alpha = \bigcup_{i=1}^K X_\alpha^{(i)} (\alpha = 1/1, 1/2, 1/4, 1/16)$  に適用してマハラノビス汎距離を用いた判別分析を行う。教示に用いた  $X_\alpha$  の各画像がどの程度教示どおりに判別されたかで、感性モデルのあてはまりの良さを評価し、  $C^{(i)} (i = 1, \dots, K)$  の判別にはどの解像度による感性モデルが最適であるかを判定する。
- (6)  $C^{(i)} (i = 1, \dots, K)$  を各々分類対象集合  $C$  と見なす。また、これにあわせて  $N$ :  $C^{(i)}$  に含まれる教示用画像の枚数とする。
- (7)  $N \gg d$  の条件下で (2)~(6) を繰り返す。
- (8) (1)~(7) で構築された感性モデルを画像データベース全体に適用し、自動的にデータベースの階層的分類を行う。

階層的分類アルゴリズムでは、分類されたデータベースの階層の深さに制限はない。仮に分類途中でユーザが疲労を感じた場合には、それまでに分類した教示画像から一時的な感性モデルを構築してデータベースの階層化を行い、類似画像検索を試みることができる。また、後日、中断した個所から教示画像の分類を再開し、より細かい自己の主観的な画像類似度基準を与えて感性モデルを再構築することもできる。た

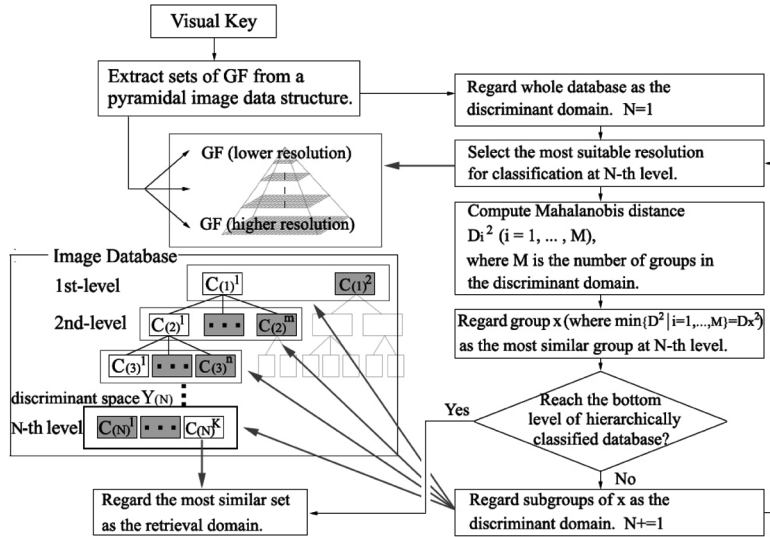


図 7 粗検索のアルゴリズム

Fig. 7 Algorithm of rough retrieval.

だし、統計的に信頼できる解析を行うために、分類された各グループに属する教示画像枚数は、説明変数の数 (GF ベクトルの要素数) よりも多い必要 ( $N \gg d$ ) がある。

#### 4.4 類似画像検索アルゴリズム

##### 4.4.1 粗検索

例示画像  $a_0$  が与えられると、システムは、まず粗検索として  $a_0$  が階層的に分類されたデータベース内のどの群に最も類似しているのかを段階的に判定する。この段階的判別はデータベースの最下層 (図 7 では  $N$  に相当) に到達するまで繰り返され、最下層において  $a_0 \in C_N^{(K)}$  と判定された群  $C_N^{(K)}$  のみを詳細検索の検索対象とする。

4.3 節の階層的分類アルゴリズムを事前に適用することで、画像データベースはユーザにとってよく目につく特徴が類似している画像群から細かな特徴が類似している画像へと階層的に分類されている。そのため粗検索を経て決定された検索対象グループ内 (図 7 では  $C_N^{(K)}$  に相当) の各画像と例示画像は、ユーザから見るとかなり類似したものであると期待できる。

##### 4.4.2 詳細検索

詳細検索では、より精密な画像特徴量を用いて例示画像と検索対象グループ内の画像との類似度を定量化し、類似度による検索結果の順位付けを行う。

詳細検索のアルゴリズムでは、画像特徴空間から検索対象グループを特徴付けている変数を選択し、詳細検索を行うための検索空間を生成する。変数選択の手順は次のとおりである。

まず  $p$  個の変数を用いて検索対象グループ  $C_N^{(K)}$  の重心と、 $C_N^{(K)}$  の判別空間  $Y_{(N)}$  (図 7 の discriminant space  $Y_{(N)}$  に相当) 内の  $C_N^{(K)}$  以外の全画像  $C_N$  の重心間のマハラノビス距離  $D_p^2$  を測定する。 $D_p^2$  は  $p$  個の変数による判別効率と呼ばれる<sup>9)</sup>。次に  $D_p^2$  と  $(p+r)$  個の変数による判別効率  $D_{p+r}^2$  との間に有意差があるのかを  $F$  値により検定する。

この手順を変数増減法を用いて繰り返し、選択された変数 (画像特徴量) のヒストグラム (3 章で論じた詳細検索用の画像特徴量) を比較することで例示画像に最も類似した画像を検索する。

類似検索を行う際の計算回数面でいえば、階層的に分類されたデータベースから例示画像に最も類似した群を特定 (粗検索) するまでに  $O(\log N)$  の計算が必要となる。 $N$  はデータベース内の全画像枚数である。また特定された群を検索対象として詳細検索を行う際には  $O(N')$  回の計算が必要である。 $N'$  は検索対象内の画像枚数であり、 $N \gg N'$  である。

## 5. 実験結果

我々は、テキスト画像、風景写真、都市景観写真、絵画といった 25,000 枚からなる画像データベースを用いて、階層的分類法を用いた画像類似検索システムの有効性を検証した。

### 5.1 データベースの階層分類

はじめに画像データベースから無作為に抽出した 3,000 枚の教示用画像を 4.3 節で論じたデータベースの階層的分類アルゴリズムにより階層分類し、ユーザ

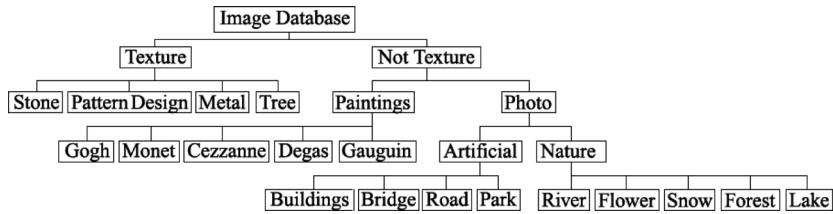


図 8 階層的に分類されたデータベースの一例  
 Fig. 8 An example of hierarchically classified database.

表 1 粗検索における群判別の成功率  
 Table 1 Recognition ratios of each classification.

Parameters	Methods	A	B	C	D	E	F	G	H	I
Our GF	HC	87.9	92.8	89.8	95.6	99.4	89.4	99.1	90.1	91.0
	Non-HC	72.1	54.9	34.7	54.9	61.1	66.9	72.1	59.3	58.1
Color Histogram	HC	51.8	47.3	44.0	43.8	42.7	45.7	42.9	54.9	49.3
	Non-HC	44.8	40.6	35.1	50.1	24.9	39.4	38.0	46.2	44.5

Parameters	Methods	J	K	L	M	N	O	P	Q	R	Average (%)
Our GF	HC	95.1	89.9	87.2	93.8	92.1	88.2	94.0	87.5	91.2	91.8
	Non-HC	74.6	62.9	71.4	47.6	61.7	64.4	72.7	61.0	66.5	62.2
Color Histogram	HC	75.0	76.8	69.1	71.8	79.3	67.0	78.6	63.4	72.0	59.7
	Non-HC	65.3	57.6	56.2	52.0	55.1	47.1	64.7	50.8	59.7	48.5

表中の A-R は、A：石材，B：パターンデザイン，C：メタル，D：木材，E：Gogh，F：Monet，G：Cezzanne，H：Degas，I：Gauguin，J：ビル，K：橋，L：道路，M：公園，N：河川，O：花，P：雪景色，Q：森林，R：湖畔をそれぞれ表す。

の視覚感性モデルを構築する。

今回の実験では、ユーザはまず 3,000 枚の教示用画像をテキストとそれ以外に分類した(第 1 階層)。第 2 階層ではユーザはテキストをメタル、木目、パターンデザイン、石材に、それ以外を絵画と写真に分類した。第 3 階層では絵画を Gauguin, Monet, Gogh, Degas, Cezzanne に、写真を自然と人工物に分類した。第 4 階層では自然を花、川、湖畔、雪景色、森林に、人工物を橋、道路、ビル、公園に分類した。以上の手順を踏んで階層的に分類された教示画像から感性モデルを構築した後、感性モデルを全データベースに適用し、データベースを自動的に階層分類した(図 8)。図 8 に示した例はあくまでも一例であり、ユーザによって分類の結果は異なる。

5.2 検索精度の評価

次に画像データベースから無作為に 1,000 枚の画像を 10 セット抽出し、抽出した評価用画像を用いて類似画像検索を行い、その検索精度を評価する。本稿では 3 章で定義した画像特徴量 (GF) および 4 章で提案した階層的分類法の有効性を評価するため、階層的分類法(表 1, 図 9 中の HC)を用いるか否か(同 Non-HC), 我々が定義した画像特徴量を用いるか(同

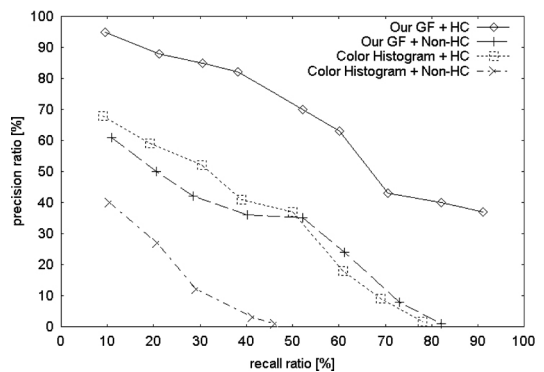


図 9 詳細検索における適合率と再現率  
 Fig. 9 Precision ratio and recall ratio of detail retrieval.

Our GF), カラーヒストグラムのみを画像特徴量とするか(同 Color histogram), 以上の各組合せについて個別に検索精度を評価した。

5.2.1 粗検索の精度

本項では粗検索における類似画像群判定の精度を評価する。まずあらかじめ目視により、各評価用画像が階層分類されたデータベース内のどの画像群に属すべきかを判定する。次に各評価用画像が、階層的に分類されたデータベース中のどの画像群に最も類似して

いるのかをマハラノビス汎距離による判別分析を用いて判定し、判別分析の結果と事前の目視判定の結果が一致した割合（群判別の成功率）で粗検索の精度を評価する（表1）。

我々の定義した局所的な3点間のコントラストを主とした画像特徴量と階層的分類法を組み合わせた場合、群判別の成功率は平均で91.8%であった。しかし、全域的な画像特徴量であるカラーヒストグラムと階層的分類法を組み合わせた場合の群判別の成功率は59.7%にとどまった。この結果から、生理的レベルの視覚感性は、全域的な画像特徴量だけではなく、局所的な画像特徴量も併用しなければ十分にモデル化しえないことが分かる。

一方、我々の定義した画像特徴量を用いたとしても、階層的分類法を併用しなかった場合、群判別の成功率は62.2%にとどまった。以上の結果から、生理的レベルの感性のモデル（我々の定義した画像特徴量）と心理的レベルの感性のモデル（階層的分類法）は独立に使用するのではなく、組み合わせることで高い精度の群判別を実現できることが分かる。

### 5.2.2 詳細検索の精度

次に、4.4節で論じた類似画像検索アルゴリズムの検索精度を評価する。我々は粗検索の評価に用いたのと同じ1,000枚からなる評価用画像群（計10セット）を用いて類似検索を試み、その検索精度を再現率と適合率で評価した（図9）。

図9に示すように、我々の定義した画像特徴量と階層的分類法を組み合わせた場合の再現率および適合率の値は、いずれもそれ以外の組合せによるものよりも高い水準にある。これは、我々の類似検索アルゴリズムでは、粗検索として階層的に分類されたデータベースから例示画像に最も類似した群を特定した後、その群のみを検索対象として詳細検索を試みるため、粗検索の精度が低い場合には、適合率、再現率ともに高い値を期待できないからだと考えられる。

また、検索に要する時間に関していえば、階層分類法を用いた場合、検索時間は平均2.1秒で検索できたのに対し、階層分類法を用いなかった場合には、検索に平均11.8秒を要した（Pentium III 800 MHzのマシン上で計測）。

## 6. ま と め

本稿で我々は、視覚の知覚過程が持つ特徴抽出機構をまねた3点間コントラストを定義し、この3点間コントラストとデータベースの階層的分類法を組み合わせた視覚感性のモデル化手法について論じた。我々

の手法では例示画像が提示されると、まず階層的に分類されたデータベース内のどの画像群が例示画像と最も類似しているかを判定する。そして、例示画像に最も類似しているとされた群のみを検索対象として類似画像検索を行うことで、高い検索精度と計算時間の83%低減を実現した。

## 参 考 文 献

- 1) Kato, T.: Database architecture for content-based image retrieval, *The International Society for Optimal Engineering*, Vol.1662, pp.111-123 (1992).
- 2) Niblack, W., Equitz, R., et al.: The QBIC Project: Querying Images By Content Using Color, Texture and Shape, *SPIE Storage and Retrieval for Image and Video Database*, San Jose, pp.173-187 (1993).
- 3) Flickner, M., Niblack, W., et al.: Query by Image and Video Content: The QBIC System, *IEEE Computer*, Vol.28, No.9, pp.23-32 (1995).
- 4) Tada, M. and Kato, T.: Similarity Image Retrieval System Using Hierarchical Classification, *DEXA 2002*, LNCS2453, pp.779-788 (2002).
- 5) Kurita, T., Otsu, N. and Sato, T.: A Face Recognition Method Using Higher Order Local Autocorrelation and Multivariate Analysis, *Proc. 11th International Conf. on Pattern Recognition*, Vol.2, pp.213-216 (1992).
- 6) Kurita, T. and Otsu, N.: Texture Classification by Higher Order Local Autocorrelation Features, *Proc. Asian Conf. on Computer Vision*, pp.175-178 (1993).
- 7) Spillmann, L. and Werner, J.S.: *Visual Perception*, Academic Press (1990).
- 8) Ballard, D.H. and Brown, C.M.: *Computer Vision*, Prentice-Hall (1982).
- 9) 奥野忠一, 久米 均, 芳賀敏郎, 吉澤 正: 多変量解析法, 日科技連出版社 (1981).
- 10) 西尾章治郎, 加藤俊一ほか: 岩波講座マルチメディア情報学 8 情報の構造化と検索, 岩波書店 (2000).
- 11) 大津展之, 栗田多喜夫, 関田 巖: 行動計量学シリーズ 12 パターン認識—理論と応用 (1996).
- 12) 加藤俊一ほか: 画像対話型商標・意匠データベース TRADEMARK, 電子情報通信学会論文誌 D-II, Vol.J72-D-II, No.4, pp.535-544 (1989).
- 13) 栗田多喜夫, 加藤俊一ほか: 主観的類似度に対応した画像検索, 情報処理学会論文誌, Vol.31, No.2, pp.227-237 (1990).
- 14) Boff, K.R., Kaufman, L. and Thomas, J.P.: *HANDBOOK OF PERCEPTION AND HU-*



*MAN PERFORMANCE*, Vol.2, Chap. 28  
(1986).

(平成 14 年 10 月 7 日受付)

(平成 15 年 1 月 15 日採録)

(担当編集委員 清木 康, 市川 哲彦, 佐藤 聡,  
原 隆浩, 細川 宜秀)



加藤 俊一(正会員)

中央大学工学部経営システム工  
学科教授。京都大学大学院博士課程  
修了,工学博士。ヒューマンメディ  
ア工学の研究に従事。電子情報通信  
学会,日本感性工学会の各会員。



多田 昌裕(学生会員)

2001 年中央大学工学部経営シ  
ステム工学科卒業。2003 年中央大  
学大学院理工学研究科経営システム  
工学専攻博士前期課程修了。同年中  
央大学大学院理工学研究科経営シス

テム工学専攻博士後期課程入学,現在に至る。

---