

## 領域分割と複数解像度を用いた画像の特徴解析と視覚感性のモデル化

多田 昌裕<sup>†</sup>, 加藤 俊一<sup>†</sup>

<sup>†</sup> 中央大学理工学部

本論文ではユーザ毎に異なる視覚感性のモデル化手法について論じる。我々は視覚の知覚過程における特徴抽出機構を模した画像特徴量を定義し、画像から抽出した。また我々は、人間の直感的な画像の分類過程を模したデータベースの階層的分類と MDL を用いた構図推定及び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。本研究では複数の解像度の画像から抽出した画像特徴量集合から、それぞれ視覚感性のモデルを独立に構築し、ユーザが画像の類似度を判定する際に画像のどの領域のどの大きさのオブジェクトや背景に注目したのかを推定して、その注目度に応じてこれらのモデルを統合することで各人の画像中での注目部位を考慮した視覚感性のモデル化を実現した。

### Visual KANSEI Modeling based on Focal Area Analysis on Multiple Resolution and Hierarchical Classification

Masahiro TADA<sup>†</sup>, Toshikazu KATO<sup>†</sup>

<sup>†</sup>Faculty of Science and Engineering, Chuo University

In this paper, we proposed a method to model each user's visual KANSEI process. We developed a tri-contrast parameter as an approximation of lateral inhibition in human's early vision process. We also proposed a method to model each user's subjective criterion for similarity. In our method, based on the parameter, we segment image planes at multiple resolution into focal areas and others by MDL and construct optimal parameter spaces for the areas. By these ideas, we can model each user's subjective criterion for similarity. We have applied them to a similarity image retrieval system including various kinds of contents.

## 1 まえがき

多種多様な画像を含むデータベースから、ユーザが提示した画像と特徴が類似したものを検索するためには、人間が感じる各画像の特徴を数量化し、機械にもその特徴を理解できるようにモデル化しなければならない。視覚感性をモデル化するためには、目が刺激を受けてから、その刺激を解析し対象を認識するまでのプロセス（視覚の知覚過程）を何らかの形で機械的に再現できるようにする必要がある。

我々は視覚感性を個人間の差異があまりない生理的レベルの感性と、各人の知識、経験によって個人差が生ずる心理的レベルの感性の2つのレベルに分けて考えている [1]。本研究では、各レベル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。

## 2 感性のモデル化

### 2.1 生理的レベルの感性

視覚の知覚過程には、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（視

覚受容器）が存在することが知られている [2]。我々はこの視覚受容器が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ている [1]。生理的レベルの感性をモデル化するためには、視覚受容器を抽象化・単純化し、その振る舞いをシミュレートできるようにする必要がある。本研究で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、視覚情報のパラメータ化を行った。

### 2.2 心理的レベルの感性

人間は同一の目の構造を持ち、視覚受容器の働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断基準は各人各様である。これは、各ユーザの経験、知識によって画像中の注目領域や、重視する特徴が異なるからだと考えられる。

我々は視覚受容器により抽出された多数の特徴を統合、取捨選択し対象を分類・認識する過程を心理的レベルの感性と考えている [1]。心理的レベルの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。本研究では、データベースの階層分類とクラスタリング手法及び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を提案する。

### 3 画像特徴量の設計

2.1 節で論じたように、生理的レベルの感性をモデル化するためには、対象の特徴を数量化し多次元ベクトルとして表現する画像特徴量 (GF) を設計して、特徴抽出機構の働きをシミュレートする必要がある。

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あり、全域的な特徴を表現できるため、画像検索などでよく利用される。しかし、局所的な明暗や色彩の特徴を欠いているため、人間とは異なった類似度の評価をする場合も多い。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、画像から全域的特徴量のみではなく、近傍画素間の局所的な関係を表現する局所の特徴量も同様に設計・抽出する必要がある。

#### 3.1 高次自己相関関数

局所の特徴量としては、自己相関関数を高次に拡張した高次自己相関関数がよく知られている。高次自己相関関数は画像平面  $P$  上の参照点  $\mathbf{r} \in \mathbf{R}^2$  周りの  $N$  個の変位  $(\mathbf{a}_1, \dots, \mathbf{a}_N \in \mathbf{R}^2)$  に対して、

$$y^N(\mathbf{a}_1, \dots, \mathbf{a}_N) = \int_P f(\mathbf{r})f(\mathbf{r}+\mathbf{a}_1) \cdots f(\mathbf{r}+\mathbf{a}_N)dr, (1)$$

で定義される。ここで  $f(\mathbf{r})$  は参照点  $\mathbf{r}$  の輝度値である。大津、栗田等は高次自己相関関数の次数  $N$  を 2 までとし、変位を参照点  $\mathbf{r}$  周りの局所的な  $3 \times 3$  画素の領域に限定した高次自己相関特徴を提案している [3]。しかし、高次自己相関特徴は対象点  $\mathbf{r}$  周りの輝度値の積に着目した特徴量であるため、ダイナミックレンジが広く、またノイズや画像撮影時の環境光の影響を受けやすい。

画像中のノイズや撮影時の環境光の違いによって画像特徴量の値が大きく変動するようでは、様々な画質の画像を含んだ画像検索に利用することは難しい。そのため、画像特徴量はノイズや環境光の変化に強い設計であることが望ましい。

#### 3.2 3点間コントラスト

外界景色では周囲の照度レベルが変化しても、背景と対象のコントラストは変化しないことが知られている [4]。そこで本研究では背景と対象のコントラ

ストに着目し、視覚受容器の数理モデルとして3点間コントラストを次式で定義した。

$$\text{Cont}^{(i)}(\mathbf{a}_1^{(i)}, \mathbf{a}_2^{(i)}, \mathbf{r}) = \frac{\{f(\mathbf{r}+\mathbf{a}_1^{(i)})-f(\mathbf{r})\}+\{f(\mathbf{r}+\mathbf{a}_2^{(i)})-f(\mathbf{r})\}}{|f(\mathbf{r}+\mathbf{a}_1^{(i)})|+|f(\mathbf{r}+\mathbf{a}_2^{(i)})|+2|f(\mathbf{r})|}. (2)$$

ここで、 $\mathbf{r}, (\mathbf{a}_1^{(i)}, \mathbf{a}_2^{(i)})$ 、 $f(\mathbf{r})$  はそれぞれ参照点、変位、参照点  $\mathbf{r}$  の色彩であり、(2) 式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。3点間コントラストは刺激強度で正規化している為、刺激強度に対してスケール不変であり、またノイズに強いという特性をもつ。

画像平面  $P$  上の参照点  $\mathbf{r}$  と参照点周りの  $N$  個の変位  $(\mathbf{a}_1, \dots, \mathbf{a}_N \in \mathbf{R}^2)$  との関係を測定する方法は無数にあるが、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の測定範囲を参照点  $\mathbf{r}$  周りの局所的な  $3 \times 3$  画素の領域に限定する。また、変位を  $N=2$  ( $\mathbf{a}_1, \mathbf{a}_2 \in \mathbf{R}^2$ ) とすることで画像の直線的な変化の方向性だけでなく曲線の変化にも対応できるようにした。図1に本研究で採用したコントラストを測定するパターン(全28種)を示す。図中の“+”は参照点  $\mathbf{r}$ 、“\*”は変位  $\mathbf{a}_1^{(i)}, \mathbf{a}_2^{(i)} \in \mathbf{R}^2$  ( $i=1, \dots, 28$ )を示す。

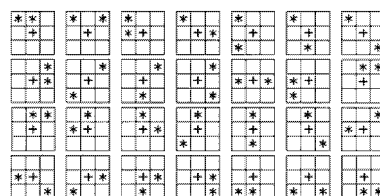


図 1: 3点間コントラストを測定するパターン

ところで、視覚の知覚過程には、ある点が刺激を受けて興奮作用を起こすと、その周辺の点が逆に抑制作用を起こす機構(側抑制と呼ぶ)があることが知られている [2]。側抑制は様々な明るさの背景の下で、注目点の近傍のコントラストを局所並列的に強調するメカニズムであると考えられる。我々は、側抑制機構が視覚受容器から抽出した多数の特徴の取捨選択や強調に有効であると考え、より高次の情報の統合(より低レベルの知覚過程から入力された情報を取捨選択し、統合する過程)に側抑制機構を導入することを考えた。本研究ではこの考えに基づき

(3) 式を定義した .

$$\gamma(i, \mathbf{r}) = \frac{\text{Cont}^{(i)}(\mathbf{a}_1^{(i)}, \mathbf{a}_2^{(i)}, \mathbf{r})}{\sum_{i=1}^{n(I)} |\text{Cont}^{(i)}(\mathbf{a}_1^{(i)}, \mathbf{a}_2^{(i)}, \mathbf{r})|} . \quad (3)$$

ここで ,  $n(I)$  はより低レベルの知覚過程から高レベルの知覚過程へと入力された信号の数 (本研究では  $n(I) = 28$ ) である . さらに我々は  $\gamma(i, \mathbf{r})$  を拡張し , (4) 式で  $\gamma'(i, l, \mathbf{r})$  を , (6) 式で  $g(i, l, P_k)$  を定義した .

$$\gamma'(i, l, \mathbf{r}) = \delta_{i,l,r} \gamma(i, \mathbf{r}), \quad (4)$$

$$\delta_{i,l,r} = \begin{cases} 1 & \text{if } \gamma(i, \mathbf{r}) \in \lambda_l \\ 0 & \text{otherwise} \end{cases}, \quad (5)$$

$$g(i, l, P_k) = w_l \frac{\int_{P_k} \gamma'(i, l, \mathbf{r}) d\mathbf{r}}{\sum_{i=1}^{n(I)} \int_{P_k} |\gamma'(i, l, \mathbf{r})| d\mathbf{r}} . \quad (6)$$

ここで ,  $P_k$  は画像平面  $P$  上の画像領域であり ,  $\bigcup_{k=1}^K P_k = P$  である . (4) 式では , 入力信号をその強度に応じて  $L$  レベル ( $\lambda_1, \dots, \lambda_L$ ) に分類する (本論文では  $L=3$  とした) . (6) 式では (4) 式で分類された信号強度レベルごとに , ある画像領域中でどの信号が相対的に強く反応しているのかを強調し , 入力された信号を取捨選択する .  $w_l$  は信号強度レベルが  $\lambda_l$  である信号が全信号に占める割合であり ,

$$w_l = \frac{\sum_{i=1}^{n(I)} \int_{P_k} \delta_{i,l,r} d\mathbf{r}}{\sum_{l=1}^L \sum_{i=1}^{n(I)} \int_{P_k} \delta_{i,l,r} d\mathbf{r}}, \quad (7)$$

で定義する .

本稿では , 明暗と色の独立性を考慮して明度  $I$  , 及び色差  $R-G, Y-B$  を色空間の軸とする . 色空間を構成する 3 軸上で  $g(i, l, P_k)$  を測定し , 局所的特徴量とする .

また , 画像平面  $P$  上の画像領域  $P_k$  における平均色と色の分散を測定し , 全域的特徴量とする . この全域的特徴量と局所的特徴量を併せて画像特徴量ベクトル  $\mathbf{x}$  とする .

### 3.3 複数解像度からの画像特徴量の抽出

ユーザが画像の類似度判定の際に着目する特徴は対象によって異なる . また , ユーザが同一画像上で異なる大きさのオブジェクトや背景に着目していた場合 , 何らかの方法を用いて特徴量を抽出するための解像度を一意に決めたとしても , 単一解像度の画像から抽出した特徴量のみではユーザの視覚感性を十分に表現できない可能性が高い .

そこで本研究では , 複数の解像度からそれぞれ独立に画像特徴量を抽出し , 解像度ごとにユーザの視

覚感性を数理的にモデル化する . 解像度ごとに独立に構築したモデルを組み合わせ , 画像の部位に応じてモデルを使い分けることで , ユーザの視覚感性をより正確にモデル化しえると期待できる .

## 4 心理的レベルの感性のモデル化手法

### 4.1 データベースの階層的分類

2.2 節で論じたように , 心理的レベルの感性をモデル化するためには , 教示されたユーザの主観的な類似度判断基準から , ユーザが注目する画像領域や画像の特徴を推定する必要がある .

本研究では , ユーザの主観的基準ののっって画像を段階的 , 階層的にグループ分けすることで教示を行う . 階層分類された各画像群に情報量基準を用いたクラスタリング手法を適用することで , ユーザが画像を分類する過程で着目したであろう各画像群の構図を推定し , 構図を構成する画像領域ごとに独立に判別分析を適用することで , ユーザごとに異なる視覚感性の数理モデルを構築する .

ただ , 視覚感性の教示のためにデータベース内の全ての画像を階層分類するとすると , ユーザに課せられる負担は相当なものとなる . そこで , 本研究ではデータベースから無作為に抽出した教示用画像群を階層的に分類することで視覚感性の教示を行い , ユーザの視覚感性の数理モデルを構築する . そして構築した数理モデルを階層的判別分析によりデータベース全体に適用することで , データベースを自動的に階層分類する .

### 4.2 心理的レベルの感性のモデル化概略

多くの画像は画像平面全域が同じ模様で構成されているわけではないため , 画像部位によって画像特徴量の分布は異なる . また , ユーザが重視する画像特徴やその細かさも画像部位によって異なると考えられる . そこで本研究では , まず階層分類された画像群ごとに , 画像平面中で画像特徴量の値の分布が同じであると考えられる領域 (注目画像領域) を情報量基準を用いたクラスタリング手法により抽出する . そして , 抽出された各群の注目画像領域ごとにユーザが重視する特徴を推定し , 各ユーザの主観特徴空間を生成する . 我々は画像特徴量の値の分布が各々異なる領域が集まって構図を形作っていると考えられる . 本研究における注目領域の抽出は , 構図の推定に相当すると考えられる . こうして得られた各群の注目

画像領域及び各領域に対応する主観特徴空間を総称して注目領域モデルと呼ぶ。

ところで、3.3節で論じたように、ユーザが同一画像上で異なる大きさのオブジェクトや背景に着目していた場合、単一解像度から抽出した画像特徴量のみでは、ユーザの視覚感性を十分にモデル化できない可能性が高い。そこで、本研究では低解像度及び高解像度の画像特徴量集合からそれぞれ注目領域モデルを独立に構築し、ユーザが画像の類似度を判定する際に画像のどの領域のどの大きさ（解像度）のオブジェクトや背景に注目したのか（注目度）を推定する。推定した注目度を用いて複数の解像度から構築した注目領域モデルを統合し、統合したモデルを用いて階層的に判別分析を行うことで、ユーザが対象画像を分類し認識する過程、すなわち心理的レベルの感性をシミュレートする。

### 4.3 注目画像領域の抽出

#### 4.3.1 本研究における注目画像領域抽出の目的

画像の領域分割は、一枚の画像から複数のオブジェクトを抽出する事を目的として、画像平面を局所特徴が類似したいいくつかの領域に分割する手法である。領域分割手法としては、微細に分割された画像領域をクラスタリング手法を用いていくつかの領域に集約する手法が広く用いられている。

それに対し、本研究における注目画像領域抽出は、ユーザの主観によって類似していると判断された複数の画像に共通する類似画像領域（注目画像領域）を抽出し、ユーザが類似度判定の際に注目した構図を推定することを目的とする。そのため、本研究では画像平面  $P$  を微細に区切らず、 $M \times M$  の粗いメッシュ状の画像領域  $P_k$  ( $k = 1, \dots, M^2$ ) に区切り、初期クラスタとする。これらの初期クラスタに対し、各画像領域から抽出した画像特徴量ベクトル集合をクラスタ間の類似度判定の指標としてクラスタリング手法を適用し、各画像領域を順次統合してゆく。これにより、教示された画像群に属する各画像に共通する類似画像領域（注目画像領域）を抽出する。

#### 4.3.2 クラスタリング手法

本研究では分析対象データをそれぞれ別のクラスタと見なし、距離の近いクラスタを逐次統合することで最終的な分割結果を得る階層的クラスタリングを採用する。しかし、階層的クラスタリングではクラスタの統合をどの時点で終了するかは分析者の判

断に委ねられるため、そのままの形で本研究に適用することは難しい。そこで本研究では、階層的クラスタリングのクラスタ統合基準として情報量基準の一種である MDL 基準を導入し、領域数を自動的に決定できるようにした。

#### 4.3.3 画像領域統合モデルの MDL

MDL 基準は Rissanen により、符号化における記述長最小化 (Minimal Description Length) 原理として導出されたものであり、モデルのパラメータの記述長とモデルを用いてデータを記述したときの記述長の和が最小になるモデルを最良とみなす [5]。

いま、画像領域  $P_k$  及び  $P_l$  を統合し、画像特徴量ベクトルの集合  $X_{kl} = \{\mathbf{x}_{i,kl} \mid \mathbf{x}_{i,kl} \in X_k \cup X_l\}$  を生成することを考える。 $\mathbf{x}_{i,kl}$  にパラメータ  $\theta_{kl} = (\mu_{kl}, \Sigma_{kl})$  の  $n$  変量正規分布を仮定すると、領域統合モデルの MDL は、

$$\text{MDL}_{(uni)} = -\log L(\hat{\theta}_{kl}) + \frac{J_{(uni)}}{2} \log \frac{2N}{2\pi} + \log \int \sqrt{|I(\theta_{kl})|} d\theta_{kl}, \quad (8)$$

で算出することができる。ここで、 $J_{(uni)}$  は領域統合モデルの自由度である。

#### 4.3.4 画像領域分割モデルの MDL

一方、画像領域を統合しない場合、 $\mathbf{x}_{i,k} \in X_k$ ,  $\mathbf{x}_{i,l} \in X_l$  にパラメータ  $\theta = (\theta_k, \theta_l)$  を持つ確率分布

$$p_{(d)}(\mathbf{x}|\theta) = \frac{p(\mathbf{x}|\theta_k)^{\delta_k} p(\mathbf{x}|\theta_l)^{\delta_l}}{\int p(\mathbf{x}|\theta_k)^{\delta_k} p(\mathbf{x}|\theta_l)^{\delta_l} d\mathbf{x}}, \quad (9)$$

$$\delta_i = \begin{cases} 1 & \text{if } \mathbf{x} \in X_i \\ 0 & \text{otherwise} \end{cases}, \quad (10)$$

を仮定する。ここで、 $p(\mathbf{x}|\theta_k)$ ,  $p(\mathbf{x}|\theta_l)$  はそれぞれパラメータ  $\theta_k$ ,  $\theta_l$  を持つ  $n$  変量正規分布の確率密度関数である。

このとき、領域分割モデルの MDL は、

$$\text{MDL}_{(div)} = -\log L_{(d)}(\hat{\theta}) + \frac{J_{(div)}}{2} \log \frac{2N}{2\pi} + \log \int \sqrt{|I(\theta)|} d\theta, \quad (11)$$

で算出することができる。ここで、 $\hat{\theta} = (\hat{\theta}_k, \hat{\theta}_l)$  は  $\theta = (\theta_k, \theta_l)$  の最ゆう推定量、領域分割モデルのパラメータの自由度  $J_{(div)} = 2J_{(uni)}$  である。また、 $L_{(d)}(\cdot)$  はゆう度関数であり、 $L_{(d)}(\cdot) = \prod p_{(d)}(\cdot)$  である。

#### 4.3.5 注目画像領域抽出アルゴリズム

画像群  $C$  から注目画像領域を抽出する手順は以下のとおりである。

1. 画像群  $C = \{c_1, \dots, c_N\}$  の解像度  $\rho$  の画像平面  $P_\rho$  を  $M \times M$  のメッシュに分割し、各々を領域  $P_{k,\rho} \subset P_\rho$  ( $k=1, \dots, M^2$ ) とする。
2.  $P_{k,\rho} \subset P_\rho$  ( $k=1, \dots, M^2$ ) から画像特徴量ベクトル集合  $X_{k,\rho} = \{\mathbf{x}_{i,k,\rho} \mid i=1, \dots, N\}$  を抽出する。
3. 次元圧縮のため、 $X_\rho = \bigcup_{k=1}^{M^2} X_{k,\rho}$  に主成分分析を適用し、 $Y_\rho = \bigcup_{k=1}^{M^2} Y_{k,\rho}$ ,  $Y_{k,\rho} = \{\mathbf{y}_{i,k,\rho} \mid i=1, \dots, N\}$  を得る。
4. 画像特徴量空間  $Y_\rho$  における各画像領域の重心間のマハラノビス汎距離

$$D_{k,l}^2 = (\bar{\mathbf{y}}_{k,\rho} - \bar{\mathbf{y}}_{l,\rho})^T \Sigma^{-1} (\bar{\mathbf{y}}_{k,\rho} - \bar{\mathbf{y}}_{l,\rho}), \quad (12)$$

を測定し、 $D_{t_1,t_2}^2 = \min\{D_{k,l}^2 \mid k,l=1, \dots, M^2\}$  となる領域  $P_{t_1,\rho}, P_{t_2,\rho}$  を統合対象とする。ここで、 $\bar{\mathbf{y}}_{k,\rho}, \bar{\mathbf{y}}_{l,\rho}, \Sigma^{-1}$  はそれぞれ、 $Y_{k,\rho}, Y_{l,\rho}$  の重心ベクトル、共分散行列の逆行列である。

5. 領域統合モデルと領域分割モデルの MDL をそれぞれ算出し、 $\text{MDL}_{(uni)} \leq \text{MDL}_{(div)}$  ならば、領域  $P_{t_1,\rho}, P_{t_2,\rho}$  を統合する。そうでなければ、処理を終了する。
6. (4)–(5) の処理を繰り返す。

こうして得られた注目画像領域は、ユーザが画像群  $C$  内の画像を知覚する際に着目した構図に相当すると考えられる。

本論文では、注目画像領域を高解像度、低解像度それぞれの画像から独立に抽出し、オブジェクトの形状に重点を置いた場合の構図、テクスチャに重点を置いた場合の構図をそれぞれ推定する。

#### 4.4 主観特徴空間の生成

##### 4.4.1 画像分類用の主観特徴空間

4.2 節で論じたように、ユーザが重視する画像特徴やその細かさは画像平面の領域によって異なると考えられる。そこで本研究では、高解像度及び低解像度の画像から各々抽出した注目画像領域ごとに独立に変数選択を行うことで、ユーザが画像分類の際に当該領域で着目した特徴を推定し、各領域に対応した画像分類用の主観特徴空間とする。

いま、教示用画像群  $C = \{c_1, \dots, c_N\}$  の解像度  $\rho$  の画像平面  $P_\rho$  は、4.3.5 節の注目画像領域抽出アル

ゴリズムによって  $K_\rho$  個の注目画像領域  $P_{k,\rho}$  ( $k=1, \dots, K_\rho$ ) に分割されているとする。また  $C$  は、あらかじめユーザの主観的判断基準に基づいて任意の  $\alpha$  個の群  $C^{(j)} = \{c_1, \dots, c_{N_j}\}$ ,  $\bigcup_{j=1}^{\alpha} C^{(j)} = C$  に分類されているとする。

$q$  個の変数を用いて  $\alpha$  個の群がどの程度判別できるかを示す一つの指標として、次式で定義される Wilks の  $\Lambda$  がある。

$$\Lambda = |S_W|/|S_T|. \quad (13)$$

ここで、 $S_W, S_T$  はそれぞれ群内及び全体の平方和積和行列であり、 $|\cdot|$  は行列式を表す。

いま、 $q$  個の要素からなる特徴量ベクトル  $\mathbf{x}_q$  に、特徴量  $x_r$  を加えることを考える。このとき、 $\alpha$  個の群を判別する能力の増分は、 $\Lambda$  を用いて次式で測定することができる。

$$\Lambda(x_r|\mathbf{x}_q) = \Lambda(\mathbf{x}_q + x_r)/\Lambda(\mathbf{x}_q). \quad (14)$$

ここで、 $\Lambda(\mathbf{x}_q + x_r), \Lambda(\mathbf{x}_q)$  はそれぞれ、 $\mathbf{x}_q$  に  $x_r$  を加えたときの  $\Lambda$ 、 $\mathbf{x}_q$  のみ用いたときの  $\Lambda$  である。本研究では、 $\Lambda(x_r|\mathbf{x}_q)$  を用いて特徴量を取捨選択し、領域  $P_{k,\rho}$  における画像分類用の主観特徴空間  $\Theta_{k,\rho}$  とする。

##### 4.4.2 画像認識用の主観特徴空間

前節で生成した画像分類用の主観特徴空間を用いて判別分析を階層的に適用した結果、様々な画像は最終的に、階層分類された各画像群のうち最下層に位置する画像群のいずれかに分類される。これら最下層に位置する画像群は、ユーザが自らの判断基準を階層的に教示した結果、互いに関わり類似した画像で構成されていると考えられる。

本研究では、ユーザが最下層に位置する画像群内の各画像にどのような類似性を見出したのかを推定するため、解像度  $\rho$  の画像平面  $P_{k,\rho}$  から抽出した画像特徴量ベクトル集合  $X_{k,\rho} = \{\mathbf{x}_{i,k,\rho} \mid i=1, \dots, N\}$  に主成分分析を適用し、画像認識用の主観特徴空間  $\Omega_{k,\rho}$  とする。

本論文で考える心理的レベルの感性のモデルでは、判別分析を用いてユーザが画像をカテゴリ分類する過程をモデル化するには画像分類用の主観特徴空間を用いる。また、画像が分類された後、特定の画像群内でより詳細にユーザの心理的レベルの感性をモデル化するには画像認識用の主観特徴空間を用いる。

#### 4.5 画像領域の注目度

画像の類似度を主観的に判断する際、ユーザは画像中の各特徴を同列に評価せず、画像中の一部の（興味を引かれた）オブジェクトや背景を重点的に評価している。そこで本研究では、各解像度の画像平面から抽出された注目画像領域ごとに、ユーザがどの程度その領域に注目しているのかを推定する注目度を算出し、注目度に応じた重み付けを行う。

我々は領域内における画像特徴量の分布の分散を利用すれば、ユーザが当該領域にどの程度一貫した判断基準を与えているかを測定できると考えた。領域内での画像特徴量の分布の分散が大きければ、その領域内ではユーザの判断基準が一貫していない、すなわちユーザはさほどその領域に注目していないと考えられる。逆に領域内での画像特徴量の分布の分散が小さければ、その領域内ではユーザの判断基準が一貫している、すなわちユーザはその領域に注目していると考えられる。

本論文ではこの考えに基づき、解像度  $\rho$  の領域  $P_{k,\rho}$  ( $k = 1, \dots, K_\rho$ ) における注目度  $z_{k,\rho}$  を次式で定義した。

$$\zeta_{k,\rho} = \frac{\sum_\rho \sum_k \log(|\Sigma_{k,\rho}|)}{\log(|\Sigma_{k,\rho}|)}, \quad (15)$$

$$z_{k,\rho} = \frac{\zeta_{k,\rho}}{\sum_\rho \sum_k \zeta_{k,\rho}}. \quad (16)$$

ここで、 $\Sigma_{k,\rho}$  は領域  $P_{k,\rho}$  の主観特徴空間における画像特徴量ベクトル集合の共分散行列、 $|\cdot|$  は行列式を表す。また、 $\sum_\rho \sum_k z_{k,\rho} = 1$  である。 $z_{k,\rho}$  の値が大きいほど、当該領域内におけるユーザの判断基準は一貫している、すなわちユーザはその領域に着目していると考えられる。

この注目度を用い、各解像度の画像特徴量ベクトル集合から独立に構築した注目領域モデルを統合することで、画像を認識する際、ユーザが画像中のどの部位の、どの大きさ（解像度）のオブジェクトや背景に着目しているのかを推定できる。本論文では、注目度を用いて統合されたこれらの注目領域モデルを心理的レベルの感性のモデルと考える。

#### 4.6 判別分析

いま、教示用画像群  $C = \{c_1, \dots, c_N\}$  はあらかじめユーザの主観的判断基準に基づいて任意の  $\alpha$  個の群  $C^{(j)}$ 、 $\bigcup_{j=1}^\alpha C^{(j)} = C$  に分類され、 $C$  の解像度  $\rho$  の画像平面  $P_\rho$  は  $K$  個の注目画像領域  $P_{k,\rho}$  ( $k = 1, \dots, K_\rho$ ) に分割されているとする。また、 $C^{(j)}$  の

画像領域  $P_{k,\rho}$  から抽出した画像特徴量ベクトル集合を  $X_{k,\rho}^{(j)} = \{\mathbf{x}_{i,k,\rho}^{(j)}\}$ 、 $X_{k,\rho}^{(j)}$  の  $\Theta_{k,\rho}$  における表現を  $X_{k,\rho}^{(j)'} = \{\mathbf{x}_{i,k,\rho}^{(j)'}\}$  とする。ここで、 $\Theta_{k,\rho}$  は領域  $P_{k,\rho}$  の画像分類用の主観特徴空間である。

この時、画像群  $C^{(j)}$  と対象画像  $c_0$  間の距離を次式で定義する。

$$\mathcal{D}_{(j)}^2 = \sum_\rho \sum_{k=1}^{K_\rho} z_{k,\rho} \left\{ \frac{1}{\dim \Theta_{k,\rho}} \times (\mathbf{x}'_{0,k,\rho} - \bar{\mathbf{x}}_{k,\rho}^{(j)'})^T \Sigma_{(j,k,\rho)}^{-1} (\mathbf{x}'_{0,k,\rho} - \bar{\mathbf{x}}_{k,\rho}^{(j)'}) \right\}.$$

$\mathbf{x}'_{0,k,\rho}$ 、 $\bar{\mathbf{x}}_{k,\rho}^{(j)'}$ 、 $\Sigma_{(j,k,\rho)}^{-1}$ 、 $\dim \Theta_{k,\rho}$ 、 $z_{k,\rho}$  はそれぞれ、 $c_0$  の画像特徴量ベクトルの  $\Theta_{k,\rho}$  における表現、 $X_{k,\rho}^{(j)'}$  の重心ベクトル、 $X_{k,\rho}^{(j)'}$  の共分散行列の逆行列、 $\Theta_{k,\rho}$  の次元数、領域  $P_{k,\rho}$  の注目度である。本研究では、対象画像との距離が最小となる群  $\beta$  に対象画像が属すると判定する。

#### 4.7 心理的レベルの感性のモデル化アルゴリズム

心理的レベルの感性のモデル化の手順をまとめると以下ようになる。

1. 画像データベースから、教示用画像データを無作為に抽出する。
2. 教示用画像データの全体集合  $U = \{c_1, \dots, c_N\}$  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
3.  $C$  をユーザの主観的な画像類似度判断基準に基づき、任意の  $\alpha$  個のグループ  $C^{(i)}$ 、 $\bigcup_{i=1}^\alpha C^{(i)} = C$  に分割する。
4. 4.3.5 節の注目画像領域抽出アルゴリズムを用い、 $C^{(i)}$  の解像度  $\rho$  の画像平面  $P_\rho$  を領域  $P_{k,\rho}^{(i)}$  ( $k = 1, \dots, K$ ) に分割する。
5. 4.4 節の手順を踏んで  $C^{(i)}$  の領域  $P_{k,\rho}^{(i)}$  における主観特徴空間を生成する。
6.  $C^{(i)}$  の各画像領域の注目度を算出する。
7.  $C^{(i)}$  を各々分類対象集合  $C$  とみなす。また、これにあわせて  $N$  の値を次のように更新する。  
 $N$  :  $C^{(i)}$  に含まれる教示用画像の枚数
8. 統計的に信頼できる解析を行うために、  
 $N \gg$  画像特徴量ベクトルの要素数の条件下で (2)–(7) を繰り返す。
9. (2)–(8) で構築された心理的レベルの感性のモデル（注目度によって統合された注目領域モデル）を判別分析により画像データベース全体に適用し、自動的にデータベースを階層分類する。

#### 4.8 類似画像検索への応用

我々は提案した視覚感性のモデル化手法を類似画像検索システムに応用し、個人の視覚感性を反映した類似画像検索システムを開発した。

我々のシステムでは、例示画像  $c_0$  が与えられると、まず粗検索として  $c_0$  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、注目度により統合された注目領域モデルに基づき段階的に判別する。この段階的判別はデータベースの最下層に到達するまで繰り返され、最下層において  $c_0 \in C^{(\beta)}$  と判定された群  $C^{(\beta)}$  のみを検索対象とする。

次に詳細検索として、 $C^{(\beta)}$  の解像度  $\rho$  の注目画像領域  $P_{k,\rho}^{(\beta)}$  ( $k=1, \dots, K_\rho^{(\beta)}$ ) と対応する画像認識用の主観特徴空間  $\Omega_{k,\rho}^{(\beta)}$  を用いて、例示画像と  $C^{(\beta)}$  内の各画像との類似度

$$\text{Sim}_{(i)} = \sum_{\rho} \sum_{k=1}^{K_\rho^{(\beta)}} z_{k,\rho}^{(\beta)} \left\{ \frac{1}{\dim \Omega_{k,\rho}^{(\beta)}} \times (\mathbf{s}_{0,k,\rho} - \mathbf{s}_{i,k,\rho}^{(\beta)})^T \Sigma_{(\beta,k,\rho)}^{-1} (\mathbf{s}_{0,k,\rho} - \mathbf{s}_{i,k,\rho}^{(\beta)}) \right\},$$

を測定し、 $\min\{\text{Sim}_{(i)} \mid i=1, \dots, N_\beta\} = \text{Sim}_{(t)}$  となる画像  $c_t^{(\beta)}$  が  $c_0$  に最も類似していると判定する。ここで  $z_{k,\rho}^{(\beta)}$ ,  $S_{k,\rho}^{(\beta)} = \{\mathbf{s}_{i,k,\rho}^{(\beta)} \mid i=1, \dots, N_\beta\}$ ,  $\Sigma_{(\beta,k,\rho)}^{-1}$ ,  $K_\rho^{(\beta)}$ ,  $N_\beta$  はそれぞれ、領域  $P_{k,\rho}^{(\beta)}$  の注目度、領域  $P_{k,\rho}^{(\beta)}$  から抽出した画像特徴量集合  $X_{k,\rho}^{(\beta)} = \{\mathbf{x}_{i,k,\rho}^{(\beta)} \mid i=1, \dots, N_\beta\}$  の  $\Omega_{k,\rho}^{(\beta)}$  における表現、 $S_{k,\rho}^{(\beta)}$  の共分散行列の逆行列、 $C^{(\beta)}$  の解像度  $\rho$  の画像平面での注目画像領域数、 $C^{(\beta)}$  に属する画像枚数である。

### 5 実験結果とその評価

#### 5.1 データベースの階層分類

従来の画像検索システムの多くは、同一のカメラおよび環境下で撮影するなど、検索対象の画質が揃っていることを前提としていた。しかし本稿では、市販の写真素材集、インターネットで公開されている著作権フリーの写真、著者らがデジタルカメラで撮影した写真など、画質（ノイズやJPEGの圧縮率、撮影条件）が各々異なる画像 14,000 枚を用意し、提案した類似画像検索システムの有効性を検証した。

まず、我々は教示用画像群として 3,000 枚の画像をデータベースから無作為に抽出し、これらを階層的に分類することでユーザの主観的な判断基準を示した。図 2 にユーザが示した階層分類の一例を示す。ただし、図 2 に示した教示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。

次に、階層的に分類された各画像群に情報量基準を用いたクラスタリング手法を適用し、各群の注目画像領域を抽出した。本稿では、注目画像領域を高解像度 (1/1)、低解像度 (1/16)、2 種類の解像度の画像群から独立に抽出し、オブジェクトの形状に重点を置いた場合の構図、テクスチャに重点を置いた場合の構図及び各領域の注目度を推定した。こうして得られた画像領域ごとに独立に主観特徴空間を生成し、注目領域モデルを構築した。

我々は注目度により統合した注目領域モデルを階層的判別分析によりデータベース全体に適用し、データベース全体を自動的に階層分類した。

#### 5.2 検索精度の評価

我々は教示用画像と重ならないようにデータベースから 1,000 枚の画像を無作為に抽出し、評価用画像とした。これらを検索キーとして類似画像検索を行い、類似度上位 30 枚検索時の精度を適合率と再現率で評価した（表 1）。また、本稿では比較のため、画像特徴量として次数を 2 までとした高次自己相関特徴及びカラーヒストグラムを用いた場合の適合率と再現率も併せて算出した。

表 1: 非教示データに対する適合率と再現率

	適合率	再現率
本稿で提案した手法	74.1%	68.5%
高次自己相関特徴使用	40.3%	37.3%
カラーヒストグラム使用	34.2%	33.8%

表 1 に示したように、本稿で提案した手法を用いた場合、適合率、再現率ともに高い水準の値を示した。

一方、画像特徴量に高次自己相関特徴やカラーヒストグラムを用いた場合、適合率、再現率ともに提案手法に比して低い水準にとどまっている。高次自己相関特徴は刺激の積に着目した特徴量であるため、画像撮影時の環境光の違いや画像にのったノイズなどの要因により、特徴量の値が影響されやすい。今回の実験で用いた画像データベースには画質が異なる画像が多数含まれていたが、上述の理由により、高次自己相関特徴ではこういった画像の類似性をうまく表現できなかったと考えられる。それに対し、提案特徴量は 3.2 節で論じたように、ノイズや環境光の変化に強い設計となっている。そのため、様々な画質の画像が混在していても、各画像の特徴をうまく表現できたと考えられる。

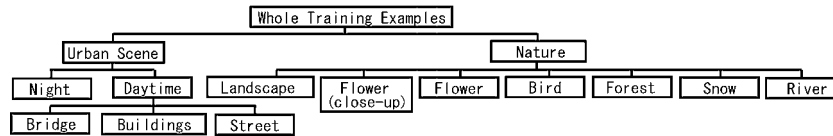


図 2: 階層分類による教示の一例

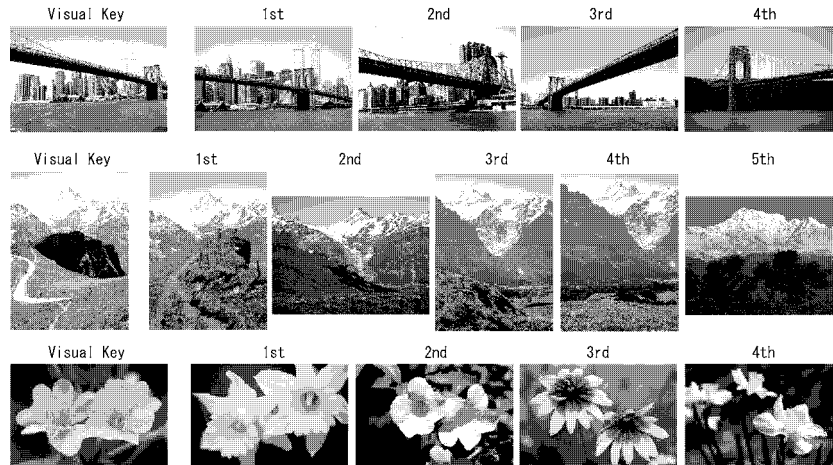


図 3: 検索結果の一例

また、カラーヒストグラムに関しては、局所的な明暗や色彩の特徴を欠いているがゆえに、視覚感性を十分にモデル化できなかったと考えられる。

## 6 むすび

本論文で我々は、視覚感性のモデル化手法を提案し、類似画像検索に応用した。

生理的レベルの感性のモデル化としては、視覚の知覚過程における特徴抽出機構を模した画像特徴量を定義し、画像から抽出した。心理的レベルの感性のモデル化としては、人間の直感的な画像の分類過程を模したデータベースの階層的分類によりユーザの主観的類似度判断基準を教示し、これを分析することでユーザが重視する特徴量の推定を試みた。

ユーザが重視する特徴は画像平面の領域によって異なると考えられるため、本研究ではまず情報量基準を用いたクラスタリング手法を適用して画像群に属する各画像に共通する構図を推定し、その上で各構図領域ごとにユーザが重視する特徴量を推定した(注目領域モデル)。また、本研究では複数の解像度の画像特徴量集合からそれぞれ注目領域モデルを独立に構築し、ユーザが画像の類似度を判定する際に画像のどの領域のどの大きさ(解像度)のオブジェクトや背景に注目したのか(注目度)を推定した。推定した注目度を用いて複数の解像度から構築した注

目領域モデルを統合し、統合したモデルを用いて階層的に判別分析を行うことで、ユーザが対象画像を分類し認識する過程、すなわち心理的レベルの感性をシミュレートした。

類似画像検索においては、例示画像に階層的判別分析を適用することでユーザが主観的に対象画像を分類する過程をシミュレートし、データベース内で例示画像に最も類似した画像群を特定する。そして、特定された画像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(上位30枚検索時で適合率74.1%、再現率68.5%)を実現した。

## 参考文献

- [1] 多田昌裕, 加藤俊一, “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,” 情処学論, vol.44, no.SIG 8, pp.37-45, 2003.
- [2] L. Spillmann and J.S. Werner, Visual Perception, Academic Press, San Diego, 1990.
- [3] N. Otsu, and T. Kurita, “A new scheme for practical, flexible and intelligent vision systems,” Proc. IAPR Workshop on Computer Vision, pp.431-435, 1988.
- [4] 池田光男, 眼はなにを見ているか—視覚系の情報処理—平凡社・自然叢書 8, 平凡社, 東京, 1988.
- [5] J. Rissanen, “Fisher Information and Stochastic Complexity,” IEEE Trans. Inf. Theory, vol.42, pp.40-47, 1996.