

階層的な判別分析を利用した感性的な類似画像検索システム

Similarity Image Retrieval System based on Subjective Criteria Using Hierarchical Discriminant Analysis

多田 昌裕¹ 加藤 俊一²

Masahiro TADA Toshikazu KATO

本稿ではユーザ毎に異なる視覚感性を反映した類似画像検索システムについて論じる。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法と判別分析を組み合わせた視覚感性のモデル化手法を開発し、類似画像検索システムに応用した。本システムでは、例示画像がユーザの主観的な画像類似度判断基準に従って階層的に分類されたデータベース内のどの画像群に属するのかを最初に判定する。判定された群のみを検索対象とし、動的に最適検索空間を生成することで高い検索精度と計算時間の低減を実現した。

In this paper, we proposed similarity image retrieval system based on users' subjective criteria. We developed a method to model users' subjective criteria for similarity by using hierarchical classification by discriminant analysis. We have applied this method to large image database including various kinds of contents, and developed similarity image retrieval system. Our system, as the first step, selects the most similar group in hierarchically classified database to the given visual image key. By constructing the optimal index space for the given visual key and retrieving only from the selected group, we achieved both better precision and less computation at the same time.

1. はじめに

従来の内容型検索システム（画像自体を検索キーとする類似画像検索システム：図1）では、同一の検索キーを用いた場合の検索結果は、ユーザによらず常に同一である。しかし、画像から受ける印象やその解釈はユーザの経験や知識によって異なるため、たとえ同一の検索キーを用いたとしても、期待する検索結果はユーザによって異なるはずである。それ故、検索結果にも各ユーザの主観的な類似度判断基準が反映されてしかるべきである。

そこで我々は各ユーザの視覚感性をモデル化し、ユーザの主観的な画像類似度判断基準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）^[1]。

本研究で我々は、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、データベースから無作為に抽出した画像群を、ユーザが自ら階層的に分類することで主観的な画像類似度判定基準を教示し、判別分析を段階的に適用して階層をたどることで、その基準をモデル化する。そして構築したモデルをデータベース全体に適用して、自動的にデータベース全体を階層的に分類する。

類似検索アルゴリズムにおいては、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する（粗検索）。そして例示画像に最も類似していると判定された群のみを検索対象とし、動的に生成した例示画像に最適な検索空間内で詳細検索を行うことにより、検索精度の向上と検索時間の短縮を実現した。

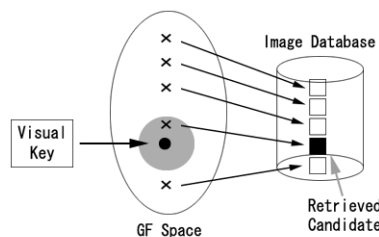


図1 内容型検索システム
Fig. 1 Content-based image retrieval system

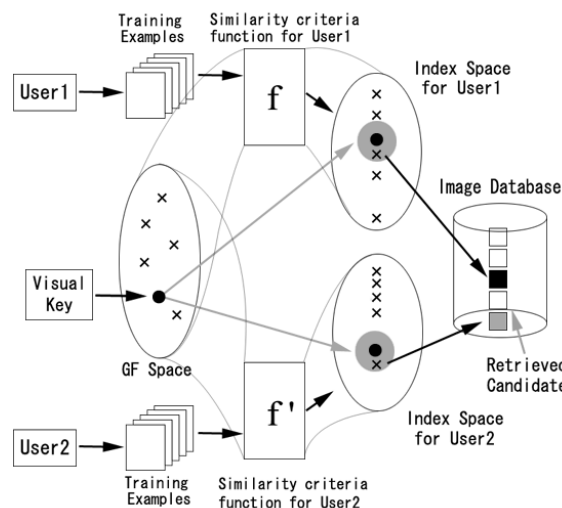


図2 視覚感性を反映した内容型検索システム
Fig. 2 Framework of similarity image retrieval based on subjective criteria of each user

2. 視覚感性

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言われている^[2]。我々は対象から特徴を抽出し、視覚情報をパラメータ化する過程を生理的レベルの（視覚）感性と定義し^[4]、抽出された特徴を取りまとめて対象を分類する過程を心理的レベルの（視覚）感性と定義している^[4]。

生理的レベルの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、その振る舞いをシミュレートできるようにする必要がある。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、特徴抽出機構の働き

¹ 中央大学大学院理工学研究科博士前期課程
umehara@indsys.chuo-u.ac.jp

² 正会員 中央大学理工学部 kato@indsys.chuo-u.ac.jp

も同様である(生理的レベルの感性には個人差はほとんどない)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、つまり心理的レベルの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。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感性モデル化手法については4章で詳述する。

3. 画像特徴量

3.1 局所の特徴量の設計

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。

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渡る色彩の分布をヒストグラム化したもので、画像の全体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。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)。従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。

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。従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 3×3 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3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)の一部を示す。

図中の“+”は参照点 r ，“*”は変位 $a_1, a_2 \in R^2$ を示す。

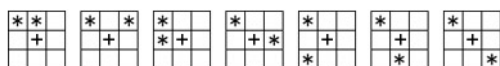


図3 コントラストを測定するパターンの一部

Fig. 3 Part of Local mask patterns for graphical features (GF)

ところで、同一の被写体の写った画像であっても位置ずれなどにより、デジタルデータとしての画像は同一とは言えない。特にテクスチャ画像を扱う場合、位置ずれ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。また、画像にノイズがのった結果、画像から抽出した画像特徴量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は利用できない。

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。

$$C^{(3)}(a_1, a_2, r) = \frac{\{g(r+a_1) - g(r)\} + \{g(r+a_2) - g(r)\}}{|g(r+a_1)| + |g(r+a_2)| + 2|g(r)|}. \quad (1)$$

$g(r)$ は参照点 r の色彩であり、(1)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。 S を刺激強度とすると、

(1)式は $\Delta S / S$ 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為、ノイズに強いという特性をもつ。

本研究では、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 $m(a_1, a_2)$ と分散 $v(a_1, a_2)$ を算出し、粗検索における局所の特徴量とする。

$$m(a_1, a_2) = \frac{1}{N} \int_P C^{(3)}(a_1, a_2, r) dr, \quad (2)$$

$$v(a_1, a_2) = \frac{1}{N-1} \int_P \{C^{(3)}(a_1, a_2, r) - m(a_1, a_2)\}^2 dr. \quad (3)$$

ここで N は画像平面 P 上の全画素数である。 $m(a_1, a_2)$ 及び $v(a_1, a_2)$ はその定義式から明らかに位置不変である。また、(2)式は

$$K \int \frac{1}{S} ds = K \times \log S + C, \quad (4)$$

の計算と同義となり、ウェーバー・フェヒナーの法則

$$I = K \times \log S + C, \quad (5)$$

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色の分散を計測した。

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

3.2 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。逆に解像度を落とすと、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。

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。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. 階層的分類法

4.1 階層的分類の利点

2章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。

- データ数: 対象とする画像数が、統計的に信頼できる解析が行えるくらいに、十分に多くなければならない。
- 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。
- 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一つの方法としては、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、全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(全く似ていない)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。しかしユーザにとって、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。そのため、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とは言い難い。

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点が挙げられる。

- A) 階層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階層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) 複数の解像度(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、一度に全ての判定基準を与えず、段階的に与えていく。そのため、ユーザの興味の変遷に伴って判定基準の重要度が変化しても、その変化に合わせて、影響のある範囲でのみ分類の再教示を行うことで、モデルを修正・再構築することが出来る。

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。この時、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、

$$D_i^2 = (x_0 - \bar{x}_i)^T \Sigma_i^{-1} (x_0 - \bar{x}_i), \quad (6)$$

で定義される^[3]。 $x_0, \bar{x}_i, \Sigma_i^{-1}$ 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、良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^[3]。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の手順

データベースの階層的分類の手順は、以下の通りである。

- (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。

- (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_α の各画像がどの程度教示通りに判別されたかで、感性モデルの当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 とみなす。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を階層的に分類する。

階層的分類の手順において、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出来る。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出来る。ただし、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 N は、説明変数の数 d (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 4 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 $a_0 - C_N^{(K)}$ 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 節の階層的分類の手順を事前に踏むことで、画像データベースはユーザにとってよく目に付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 4 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通りである。

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 4 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

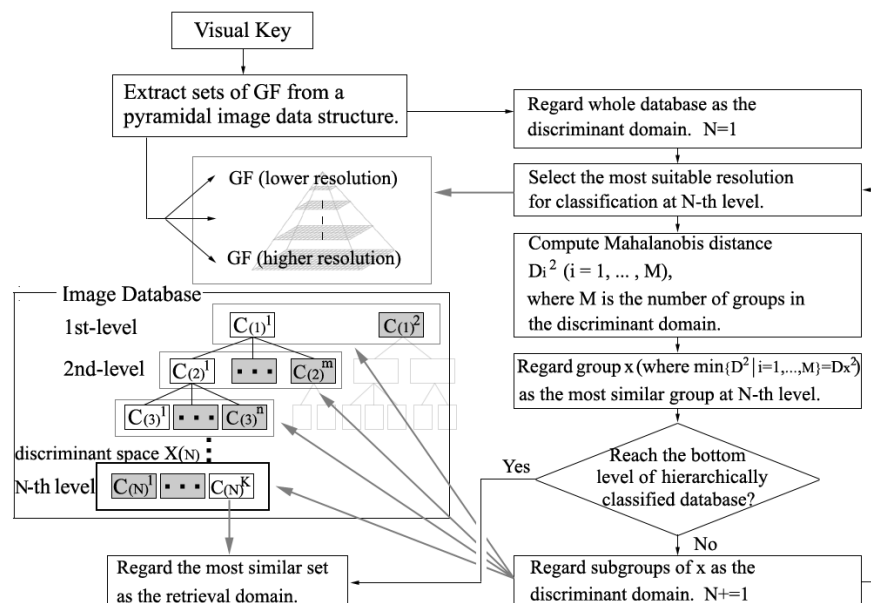


図4 粗検索のアルゴリズム
Fig. 4 Algorithm of rough retrieval

れる。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。

この手順を変数減増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（第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の面でいえば、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、テクスチャ画像、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、階層的な分類法を用いた画像類似検索システムの有効性を検証する。本稿では3章で定義した画像特徴量および4章で提案した階層的な分類法の有効性を評価するため、階層的な分類法（表1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 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、それぞれ類似画像を個別に検索精度を評価した。

表1 適合率と再現率

Table 1 Precision ratio and recall ratio

	Precision ratio	Recall ratio
Our GF+HC	74.2%	50.0%
Color Histogram+HC	49.8%	33.2%
Our GF+Non-HC	43.8%	29.2%
Color Histogram+Non-HC	29.6%	19.7%

まず、我々は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な分類の手順を踏んで階層的に分類し、ユーザの視覚感性モデルを構築する。そして、構築した感性モデルをデータベ

ス全体に適用し、自動的にデータベース全体を階層的に分類した。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、類似画像上位 20 枚検索時の検索精度を再現率と適合率で評価した（表1）。

表1に示すように、我々の定義した画像特徴量と階層的な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、それぞれを単独で用いた場合の値よりも高い水準にある。この結果から、生理的レベルの感性モデル（我々の定義した画像特徴量）と心理的レベルの感性モデル（階層的な分類法）はそれぞれ単独で使用するのではなく、組み合わせて使うことで高い精度の類似検索を実現できることが分かる。

また、検索に要する時間に関して言えば、階層分類法を用いた場合、検索時間は平均 2.1 秒ですんだのに対し、階層分類法を用いなかった場合には、平均 11.8 秒を要した。

6. まとめ

本稿で我々は、ユーザ毎に異なる視覚感性を反映した類似画像検索システムについて論じた。

我々は、データベースの階層的な分類法と判別分析を組み合わせた主観的な画像類似度判定基準のモデル化手法を提案し、それを類似検索に応用した。

我々の類似検索手法では、例示画像がユーザの主観的な画像類似度判断基準に従って階層的に分類されたデータベース内のどの画像群に属するのかを最初に判定する。そして判定された群のみを検索対象とし、動的に最適検索空間を生成することで高い検索精度（類似画像上位 20 枚検索時に、適合率 74.2%、再現率 50.0%）と 80% 以上の計算時間の低減を実現した。

【文献】

- [1] M. Tada and T. Kato: "Similarity Image Retrieval System Using Hierarchical Classification", DEXA 2002 LNCS2453, pp.779-788 (2002)
- [2] L. Spillmann and J. S. Werner: "Visual Perception", Academic Press (1990)
- [3] 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)
- [4] 西尾章治郎, 加藤俊一他: "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)

多田 昌裕 Masahiro TADA

中央大学大学院理工学研究科博士前期課程在学中。2001 中央大学理工学部経営システム工学科卒業。感性工学の研究に従事。情報処理学会、日本感性工学会の学生会員。

加藤 俊一 Toshikazu KATO

中央大学理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。情報処理学会、電子情報通信学会、日本感性工学会の各会員