画像のグルーピングとグループ間類似度に基づく主観的類似検索

中越智哉[†,☆] 佐藤健[†]

本論文は，画像データベースにおける類似検索においてユーザの類似性の観点を学習することで，観点を反映した類似検索を行う手法を提案する．本手法の概要は以下のとおりである．(1) ユーザに複数の画像をグルーピングさせ，さらに，そのグループ間の類似度をユーザに指定させる．(2) グループの類似度を反映した類似性空間を多次元尺度法によって構築する．(3) 分類させた画像から得られる物理的特徴量からその類似性空間への写像を多変量線形回帰分析によって得る．(4) 画像データベース中の画像すべてを上記写像によって類似性空間の点に変換する．(5) 新たな画像が与えられた場合には，その画像を上記の写像を用いて類似性空間の点に変換し，類似性空間上の近傍の画像を類似画像として出力する．本論文では，本手法を用いることにより，栗田ら (1990) の提案したグルーピング情報のみを用いる方法よりも，再現率，適合率，分類正解率の面で向上することを実験によって示す．

User-dependent Similar Image Retrieval Based on Image Grouping and Group Similarity

Tomoya Nakagoshi[†,☆] and Ken Satoh[†]

In this paper, we propose a similar image retrieval method according to user's viewpoint of similarity of images. The outline of our proposed method is as follows. (1) We let a user divide training images into some groups and also let him provide similarity between these groups. (2) We construct similarity space of images based on similarity between these groups by using multidimensional scaling. (3) We obtain a mapping from image feature space to similarity space by multivariate linear regression analysis. (4) We map every image in an image database to a point in the similarity space according to the above mapping. (5) Given a new image for similarity retrieval, we map the new image into a point to the similarity space. (6) We retrieve images of which corresponding points in the similarity space which are nearest to the point correspoding with the new image. In this paper, we show that our method is superior to the previous method proposed by Kurita, et al. (1990), Shimogaki and Katoh using grouping information in terms of recall ratio, precision ratio and correct classfication ratio.

1. はじめに

画像データベースの発達にともない，画像コンテンツの検索の手法の 1 つとして，類似画像検索の手法が重要になってきている．類似画像検索においては，ユーザの観点や使用法によって類似すると感じるものは異なるはずであるから，そのような観点や使用法を反映した手法になっていることが望ましい．

そのような手法としては，主に以下のようなものがある．

(1) ユーザが似ていると思う観点を画像特徴量（の組合せ）を用いて直接指定する．この方法の有名なシステムとしては，QBIC [1)]がある．

(2) 検索キーを人手でつけておいて，ユーザはそのキーを指定し，そのキーに関する類似性によって似ているものを指定する．たとえば，柴田ら[6)]は，画像にキーワードを付与しておき，キーワードが与えられたときに，シソーラスを用いて，似たキーワードが付与された画像を検索する手法を提案している．

(3) ユーザから類似性の観点を学習し，それによって類似検索を行う．栗田らの研究[3)]では，ユーザによる画像のグルーピングからユーザの観点を反映し，検索キーとして画像を用いた類似画像検索を提案している．

しかし，第 1，第 2 の方法については，それぞれ以下のような問題点があると思われる．

- 画像特徴量の指定は，特徴量に関する専門的な知

† 北海道大学大学院工学研究科
Graduate School of Engineering, Hokkaido University
☆ 現在，株式会社テンアートニ
Presently with 10artni

識がないと所望の類似画像の検索が困難であったり，特徴量と自分の類似度との関係を決定するのが難しい．

- キーの付加は妥当性（そのキーが画像を表す適切なものかどうか）およびコスト（キーを付加するための人手）の問題が存在する．とくに衛星画像のようにあらかじめ付加される情報がなくさらに大容量であるような画像データベースを扱う場合には，人手で検索キーを割り当てることには限界がある．

第3の方法は，ユーザの類似性に関する観点を学習する手法がユーザに負担のあまりない形で行われれば，上記2つの方法に比べて，ユーザに対して画像に関する専門的知識の要求や画像データベースを作る際のコストがいらないため有望な方法と考えられる．本論文は，栗田らの手法[3]で用いているユーザからの画像のグルーピングに加えて，ユーザからそのグループ間の類似度を得ることによって類似検索の精度を高める手法を提案する．

栗田らの研究[3]は，以下のように観点を学習する学習フェーズとフェーズと学習した観点による類似画像検索を行うフェーズに分かれている．

学習フェーズ（**図1** 参照）

（1） ユーザの観点をサンプル画像のグルーピングにより獲得する．

（2） 判別分析と同様の方法により，同一グループに属する画像は近く，別グループに属する画像は遠くに位置するような空間（**主観特徴空間**）を構築し，画像特徴量空間から主観特徴空間への写像を近似的に求める．

（3） 上記写像により，画像データベース上の画像を空間上の点に対応させる．

検索フェーズ

（1） 画像が与えられたときにその画像から画像特徴量を計算する．

（2） 学習フェーズで求められた写像を用いて画像特徴量から主観特徴空間の点にその画像を対応させる．

（3） 主観特徴空間上で近傍検索を行い，その画像と近い点に対応する画像を類似画像として提示する．

しかしながら，この方法で求められた主観特徴空間は，グループごとの分離を目的としたものであり，違うグループに属する画像どうしの距離が必ずしも類似性を表しているという保証は存在しないため，主観特徴空間での近傍検索が必ずしも類似画像検索になるとは限らないという問題があると思われる．

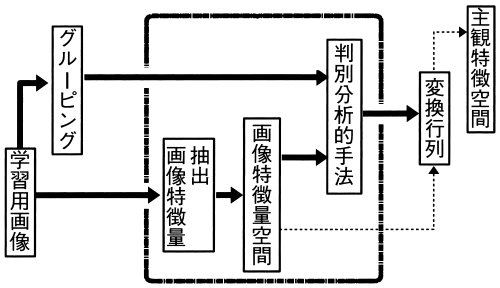


図1 栗田らの手法における学習フェーズ
Fig. 1 Learning phase in Kurita et al's method.

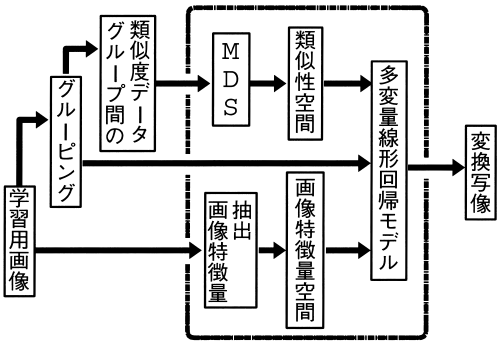


図2 本手法における学習フェーズ
Fig. 2 Learning phase in our method.

本論文では，栗田らの手法の欠点を克服した新しい類似画像検索手法を提案する．本手法では，栗田らの手法で用いるユーザのグルーピング情報に加えて，ユーザからグループ間の類似度を獲得して，その類似度を反映する類似性空間を作り出し，その空間上での近傍検索を行うことを特徴としている．概要は以下のとおりである．

学習フェーズ（**図2** 参照）

（1） ユーザの観点をサンプル画像のグルーピングとグループ間の類似度の形で獲得する．

（2） 多次元尺度法（Multidimensional Scaling; MDS）[2] によってユーザの指定したグループ間の類似性をできるだけ反映する類似性空間を構築する．

（3） 多変量線形回帰分析（Multivariate Linear Regression）[7]を用いることで画像特徴量空間からその類似性空間への写像を近似的に求める．

（4） 上記写像により，画像データベース上の画像を類似性空間上の点に対応させる．

検索フェーズ

（1） 画像が与えられたときにその画像から画像

特徴量を計算する．

（2） 学習フェーズで求められた写像を用いて画像特徴量から類似性空間の点にその画像を対応させる．

（3） 類似性空間上で近傍検索を行い，その画像と近い点に対応する画像を類似画像として提示する．

本論文では，上記手法を **M+R 法**（MDS+Regression 法）と呼ぶ．

M+R 法において用いられる多次元尺度法は，心理学などの方面でよく使われる手法で，類似度が定義された対象に対してその類似度が反映された空間を構築し，その空間の軸に対して研究者が何らかの解釈を与えるという使われ方がなされている．一方，本論文では，多次元尺度法をそのように使うのではなく，その空間自体で近傍検索を行うために，画像特徴量空間との写像を求めるという使い方をしている．このような使用法は我々の知るかぎり，独創的である．本論文では，本手法が，類似性空間を用いることにより，栗田らの手法と比べ，ユーザの類似性の観点をよりよく反映した類似検索手法になることを実験によって示す．

2. M+R 法の動作

M+R 法は，利用者の類似性に関する観点を学習する**類似性学習部**と，実際の検索を行うための**類似画像検索部**の 2 つの部分からなる．

以下では，まず M+R 法で用いる**多次元尺度法**と**多変量線形回帰モデル**について概説した後，M+R 法の手順について解説する．

2.1 多次元尺度法

多次元尺度法（Multidimensional Scaling; MDS）[2] は，複数の対象間の類似性を入力データとし，対象のそれぞれを空間上の点として表現した空間付値を出力する手法である．空間付値の次元数は，利用者があらかじめ指定することができる．

MDS に入力するデータは，対象間の類似性をなんらかの尺度によって表現したものである．この尺度の種類により，MDS は 2 つに分けられる．入力データの尺度が間隔尺度あるいは比例尺度であれば，計量的（メトリック）MDS，順序尺度であれば非計量的（ノンメトリック）MDS という．本研究では，非計量的 MDS を用いており，以下，特にことわらなければ，MDS とは非計量的 MDS のことを指すとする．

MDS の入力データとするためには，ある対象に対して，それ以外のすべての対象との類似性を与える必要がある．したがって，**表 1** のような入力データは行列の形をとる．そのことから，MDS の入力データを**類似度データ行列**と表記することもある．ここでの類似性は対称的であるので，この行列も対称行列となる．また，非計量的 MDS では，行列の要素の大小関係のみ意味を持ち，要素の大きさは，意味がないことに注意されたい．

表 1 類似度データ行列の例
Table 1 Example of similarity data matrix.

	A	B	C	D	E
A	0	9	6	8	3
B		0	4	5	1
C			0	2	7
D				0	3
E					0

出力される空間 X 上の対象 i と対象 j の距離（通常は，ユークリッド距離）を d_{ij}, 非類似度を s_{ij} とすると，これらの間には**単調関係**と呼ばれる以下の関係が近似的に満たされている．

$$\begin{cases} s_{ij} = s_{kl} \Rightarrow d_{ij} = d_{kl} \\ s_{ij} > s_{kl} \Rightarrow d_{ij} \geq d_{kl} \end{cases} \quad (1)$$

多次元尺度法の 1 つである Kruskal の方法では，出力される空間 X が，式 (1) の単調関係をどのくらいよく満たしているかを評価する基準である**ストレス**を測定して，これを最小化するような手順をとる．ストレス S は

$$S = \sqrt{\frac{\sum_{i<j}(d_{ij} - \hat{d}_{ij})^2}{\sum_{i<j}(d_{ij} - \bar{d})^2}} \quad (2)$$

と書ける．このとき，$\bar{d}$ は d_{ij} の平均を示す．また，$\hat{d}_{ij}$ は入力データと出力する空間との比較に用いる媒介変数であり，入力データ s_{ij} の順序関係に矛盾しない理想的な距離を示していると考えてよい．つまり，出力される空間 X 上での対象間の距離関係が，入力データの順序関係に一致すれば，$S = 0$ となり，最も理想的な空間であるといえる．このような空間を逐次近似的解法（最急勾配法）により求めて出力とする．

2.2 多変量線形回帰モデル

分析の対象となるある現象 A について，それを他の現象 B によって説明（推定）しようとする方法を，**回帰分析**[7] という．具体的には分析の対象となる現象 A を変数 Y（従属変数）で表し，それを説明しようとする現象 B を変数 $X_1, X_2, \cdots, X_p$（独立変数）で表したとき，**未知パラメータ** $\beta_0, \beta_1, \beta_2, \cdots, \beta_p$ を使って

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon \quad (3)$$

における ε（残差）を最小にする方法のことをいう．このような β_i は，**最小 2 乗法**によって求める．

観測値は一般には複数あるので，その数を N とし，従属変数を

$$\mathbf{Y} = (Y_1, Y_2, \cdots, Y_N)' \tag{4}$$

で表し，独立変数，未知パラメータ，残差を同様にして，それぞれ

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon \tag{5}$$

と書いたものが線形回帰モデルである．

さらに，従属変数が

$$\mathbf{Y}' = (Y_1, Y_2, \cdots, Y_m) \tag{6}$$

のようなベクトル変数である場合の線形回帰モデルのことを多変量線形回帰モデルという．$Y_1, Y_2, \cdots, Y_m$ を

$$\begin{aligned} Y_1 &= \beta_{01} + \beta_{11}X_1 + \cdots + \beta_{p1}X_p + \varepsilon_1 \\ Y_2 &= \beta_{02} + \beta_{12}X_1 + \cdots + \beta_{p2}X_p + \varepsilon_2 \\ &\vdots \\ Y_m &= \beta_{0m} + \beta_{1m}X_1 + \cdots + \beta_{pm}X_p + \varepsilon_m \end{aligned} \tag{7}$$

と表し，$\mathbf{X}$，β，ε をそれぞれ

$$\mathbf{X} = \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{p1} \\ 1 & X_{12} & X_{22} & \cdots & X_{p2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1N} & X_{2N} & \cdots & X_{pN} \end{bmatrix} \tag{8}$$

$$\beta = \begin{bmatrix} \beta_{01} & \beta_{02} & \cdots & \beta_{0m} \\ \beta_{11} & \beta_{12} & \cdots & \beta_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{p1} & \beta_{p2} & \cdots & \beta_{pm} \end{bmatrix} \tag{9}$$

$$\varepsilon = \begin{bmatrix} \varepsilon_{11} & \varepsilon_{12} & \cdots & \varepsilon_{1m} \\ \varepsilon_{21} & \varepsilon_{22} & \cdots & \varepsilon_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_{N1} & \varepsilon_{N2} & \cdots & \varepsilon_{Nm} \end{bmatrix} \tag{10}$$

と表す．

2.3 類似性学習部

M+R 法の類似性学習部（図 2）は，利用者から画像の類似性に関する観点を

- サンプル画像をグルーピングしたもの，
- そのグループに順序尺度で非類似度を設定したもの，

の形で獲得し，それを画像特徴量を使って表現するための変換写像を得ることが目的である．

本来ならば 1 つの画像に対する，それ以外のすべての画像についての非類似度を設定することが望ましいが，画像の総数が増えるとそのような作業が容易ではなく，また，多次元尺度法の計算時間も膨大になりやすいため，このような入力データを用いることとした．

手順は以下のようになる．

(1) サンプル画像をグルーピングする（サンプル画像数を n，グルーピングされたグループ数を g とする）．

(2) それらのグループどうしに類似度データ行列を設定する．

(3) 類似度データ行列を多次元尺度法に入力することにより，グループ間の類似度の順序尺度の関係を近似的に表現した空間を得ることができる．これを，**類似性空間**とする．このとき，類似性空間の次元数は，ストレス最小となる次元数（m とする）にする．類似性空間上のベクトル表現 y_i $(1 \leq i \leq g)$ は，

$$y_i = (y_1, y_2, \cdots, y_m) \tag{11}$$

と書ける．

(4) 利用者のグルーピングに従い，サンプル画像 1 つ 1 つを，類似性空間上で各グループを示す点に対応させる．このとき，類似性空間を示す行列 $\mathbf{Y}$ は，

$$\mathbf{Y} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1m} \\ y_{21} & y_{22} & \cdots & y_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{nm} \end{bmatrix} \tag{12}$$

と書ける．

(5) サンプル画像自身からは画像特徴量を抽出しておく．画像特徴量の属性数を p とし，これらを行列として書けば，

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \tag{13}$$

となる．

(6) 画像特徴量空間から類似性空間への変換写像 B を，多変量線形回帰モデル

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E} \tag{14}$$

により構成し，これを出力とする．

2.4 類似画像検索部

類似画像検索部（**図 3**）は，利用者が入力した画像そのものをキーとして，類似画像を検索して出力する．

手順は以下のようになる．

(1) まず，入力された画像自身から，画像特徴量を抽出し，これをベクトル表現する．

(2) このベクトルを，類似性学習部で得られた変換写像を用いて，類似性空間上の点に変換する．

(3) そして，類似性空間上で，入力された画像とのユークリッド距離の最も小さい画像を選び，出

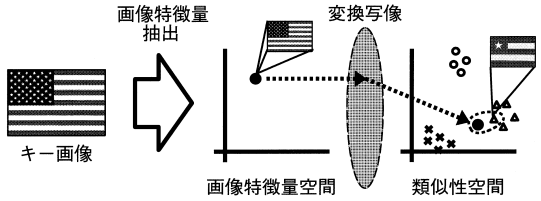


図 3 M+R 法における検索フェーズ
Fig. 3 Retrieval phase in M+R method.

力とする．

2.5 多次元尺度法と多変量線形回帰モデルの連動

M+R 法では，多次元尺度法と回帰分析を連動させて用いるが，これは,「画像の類似性」というユーザの観点を多次元尺度法によって表現したものに対して，画像特徴量を基準とした従属関係を持たせていることになる．そして，多変量線形回帰モデルを適用した結果として得られた写像を使って，画像特徴量のみで画像の類似性を説明する，ということになる．多次元尺度法と多変量線形回帰を連動させて画像の類似性を学習し，類似検索を行う手法は，これまで提案されたことのないものである．

2.6 類似手法との手法的比較

本手法と類似した手法としては，栗田ら[3)]の方法がある．栗田らの方法の特徴を簡単に示すと以下のようになる．

- 利用者がサンプル画像のグルーピングを行うことによって，利用者の観点を与える．
- 利用者の観点を反映するような空間（**主観特徴空間**）が，アフィン写像
  $$y = A'(x - \bar{x}_T)$$
  によって構成できると仮定する．ここで，y は主観特徴空間上のベクトルである．x は画像特徴量ベクトル，$\bar{x}_T$ は x の平均ベクトルである．
- この写像を，判別分析的な手法を用いて，利用者が与えたグルーピングにおいて異なるグループどうしの分散を最大に，同じグループ内の画像どうしの分散が最小になるように構成する．
- 利用者が検索のキーとして与えた画像から，画像特徴量ベクトルを抽出し，上で求めた写像によって，主観特徴空間上に変換し，類似画像を検索する．

しかしながら，この方法で求められた主観特徴空間は，グループごとの分離を目的としたものであり，違うグループに属する画像どうしの距離が必ずしも類似性を表しているという保証は存在しないため，主観特徴空間での近傍検索が必ずしも類似画像検索になるとは限らないという問題があると思われる．これに対して，M+R 法は，分けられたグループどうしに類似度を与えることができるので，類似性空間上で利用者が類似性が高いと判断したグループどうしは近く，類似性の低いと判断したグループどうしは遠くに配置できるため，利用者の観点をよりよく反映した空間を作ることができると考えられる．

3. 実　　験

3.1 評 価 基 準

システムの評価には，次の指標を用いた．

再現率： テスト用の画像に対して，類似度の高い順に上位 n 枚の画像を出力したときに，k 枚の画像が，テスト用の画像と同じグループに含まれていたとき，テスト用の画像と同じグループの総数に対する k の割合．

適合率： テスト用の画像に対して，類似度の高い順に上位 n 枚の画像を出力したときに，k 枚の画像が，テスト用の画像と同じグループに含まれていたとき，n に対する k の割合．

分類正解率： テスト用の画像に対して類似度が最も高い画像が，テスト用の画像と同じグループに含まれる割合．上の適合率で $n = 1$ のときに相当する．

分類正解率の評価は，本来の類似画像検索システムの利用目的とは異なるが，M+R 法の作る類似性空間の良さを測る基準の 1 つと考え，これを従来の分類手法の分類精度と比較した．

3.2 画像特徴量

本論文において，**画像特徴量**とは，計算機上で表現される画像（画素の集合）の持つ特徴を，なんらかの方法によって数値（または，数値の集合）として表現したものであるとする．

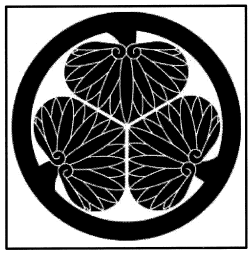
以下に，評価で用いた画像特徴量について述べる．

3.2.1 grayX 特徴量

grayX 特徴量とは，2 値画像（画素値が 0 か 1 で表されるような白黒画像）について，画像をブロックに分解して，そのブロックごとに黒画素数（画素値 1）を数えるものである．これは，局所的な画像の濃度を表すことに相当する．必要であれば，これらの値を正規化して用いる．

ブロックの数は，縦と横のブロック数が等しくなるようにする（ブロックの縦と横の画素数は異なってもよい）．このとき，縦と横のブロック数を X とすると，全ブロック数は X^2 であり，それを，**grayX** と略記する．たとえば，$X = 8$ のときは，**gray8** と略記する．

図 4 に，gray8 の例を示す．



2 値画像(249x252)

0	36	418	621	554	397	34	0
41	569	417	414	406	423	559	31
418	125	402	511	519	369	118	413
619	422	383	385	364	398	411	566
560	423	463	392	375	494	429	563
417	525	518	434	430	457	522	409
34	564	408	202	196	418	532	30
0	37	384	490	493	364	29	0

gray8データ

図 4 gray8 の例
Fig. 4 Example of gray8.

0	36	418	621	554	397	34	0
41	569	417	414	406	423	559	31
418	125	402	511	519	369	118	413
619	422	383	385	364	398	411	566
560	423	463	392	375	494	429	563
417	525	518	434	430	457	522	409
34	564	408	202	196	418	532	30
0	37	384	490	493	364	29	0

gray8データ

	36		382		203		67
41		533		1		207	
	528		152		3		8
337		444		15		97	
	293		277		109		8
201		297		19		126	
	277		39		2		21
59		1		80		7	

contrast8データの一部

図 5 cont8 データの例
Fig. 5 Example of cont8 data.

3.2.2 grayscaleX 特徴量

grayscaleX 特徴量は，基本的には grayX と同様であるが，グレースケールの画像に適用できるよう，ブロックごとの黒画素数を数えるかわりに，ブロックごとの画素の輝度値を合計するものである．これも，必要ならば正規化して用いる．grayX と同様，X の値によって，**grayscale4**（**gs4**），**grayscale8**（**gs8**）などと略記する．

3.2.3 contrastX 特徴量

contrastX は，grayX や grayscaleX の隣接するブロックとの差を要素とする特徴量である．その例を**図 5** に示す．contrastX の要素数は，grayX（grayscaleX）の縦（横）ブロック数を X としたとき，$2X(X-1)$ 個である．contrastX も，他のものと同様，**cont4**，**cont8** などと略記する．

3.3 実験の内容

3.3.1 データ・被験者

実験に用いた画像データは，国旗の画像 194 枚である．実験の被験者は，著者を含め 9 人である．

被験者には，すべての画像を提示し，

- 画像をいくつかのグループに分けさせる．
- それらのグループ間に，非類似度を類似度データ行列として設定させる．

という，2 種類の作業をさせて，学習用データを獲得した．被験者に対して，グループ数や非類似度の基準は，何も指示していない．

表 2 被験者によるグルーピング
Table 2 Grouping Result by Testees.

	グループ数	グループ内画像平均数	グループ内画像最小数	グループ内画像最大数
筆者	18	16.2	3	53
A	5	38.8	4	144
B	6	32.3	1	63
C	10	19.4	2	77
D	4	48.5	3	122
E	10	19.4	3	77
F	9	21.6	3	53
G	8	24.3	7	64
H	15	12.9	3	31

各被験者が分類したグループの数およびグループ内画像数の平均値，最小値，最大値を**表 2** に示す．

3.3.2 類似性空間の次元数

類似性空間の次元数は，多次元尺度法において矛盾が最小となる，ストレス最小の次元数を用いた．実験の結果では次元数は，各被験者のデータによりグループ数 -1 となった．

3.3.3 栗田らの方法の次元数

栗田らの手法において，主観特徴空間を構成するための線形写像は，判別分析における線形判別関数に相当する．栗田らは，この線形判別関数を，寄与率に関して閾値を設けることで，次元数を減らして計算コストの削減を図っているが，ここでは，比較のため，寄与率に閾値を設けず，線形判別関数の次元数を減らさずにそのまま用いた．

3.3.4 属性数および学習用のサンプル画像数

クロスバリデーションを行うために，データセットを以下のように生成した．

(1) 194 枚の画像から，194 枚のグループ分布とできるだけ近い分布になるように，5 つの画像集合に分ける．
(2) 4 つの画像集合を学習用のサンプル画像として用い，残りの画像集合をテスト画像として用いる．

したがって，実験としては，上記の 5 つのデータセットに対して行った．また，画像特徴量の属性数は，gs4，cont4 の計 40 属性とした．

3.3.5 比 較 項 目

上記のデータセットに対し，各テスト画像ごとに，以下の手法での再現率，適合率，分類正解率を計算した．

再現率，適合率： M+R 法，栗田らの手法，画像特徴量空間上での近傍検索．

表 3 再現率（%）
Table 3 Recall ratio (%).

検索数	10	20	30	40	50	60	70	80	90	100
筆者（M+R）	34.3	49.1	58.3	65.8	71.2	76.6	81.0	84.5	87.3	90.2
筆者（栗田ら）	26.0	40.7	48.3	55.8	61.9	66.6	70.8	75.1	79.5	83.0
筆者（画像）	23.3	34.8	41.5	48.2	53.6	57.7	62.7	67.7	72.2	75.8
A（M+R）	9.8	17.4	24.9	31.5	37.9	44.7	51.5	57.7	64.4	70.6
A（栗田ら）	7.9	14.6	21.0	27.5	33.9	40.1	46.6	53.1	59.3	65.5
A（画像）	8.0	15.1	21.2	27.8	34.0	40.7	47.3	53.6	59.3	65.3
B（M+R）	17.6	32.0	42.6	52.0	59.1	65.6	71.2	75.9	79.9	84.1
B（栗田ら）	13.1	22.9	30.9	37.8	44.3	50.9	57.0	62.9	68.7	74.1
B（画像）	12.7	21.7	29.3	36.0	42.4	48.7	54.3	60.0	65.5	71.0
C（M+R）	15.4	23.8	31.0	38.1	45.0	51.2	57.0	63.1	68.7	73.9
C（栗田ら）	13.4	21.5	28.7	36.0	42.7	48.8	54.3	60.7	66.5	72.5
C（画像）	12.5	20.2	26.2	33.2	39.5	45.2	51.5	57.3	63.3	69.1
D（M+R）	7.3	14.1	20.8	27.8	34.3	40.8	47.5	54.1	60.6	67.2
D（栗田ら）	6.8	13.0	19.3	25.8	32.2	39.2	45.4	51.9	58.4	64.9
D（画像）	6.7	12.9	19.0	25.9	32.2	38.9	45.1	50.8	57.2	63.8
E（M+R）	23.6	38.3	47.6	55.1	62.0	67.8	72.5	76.4	80.3	84.2
E（栗田ら）	18.2	28.7	35.9	42.3	49.3	54.8	60.1	65.4	71.0	75.3
E（画像）	19.3	29.3	37.0	43.3	49.0	54.0	59.0	64.7	68.6	72.5
F（M+R）	8.1	15.4	22.9	29.1	36.3	42.4	48.8	55.7	61.4	68.0
F（栗田ら）	7.1	13.5	19.9	26.8	33.5	39.9	45.8	52.4	59.3	65.7
F（画像）	7.9	14.0	20.1	26.5	32.6	38.5	44.8	51.2	57.0	63.4
G（M+R）	16.9	26.9	34.5	41.1	47.9	54.2	60.3	66.2	71.8	76.9
G（栗田ら）	15.5	23.9	31.3	38.1	44.7	51.6	57.4	63.3	68.9	74.1
G（画像）	12.7	20.9	27.6	34.0	40.1	45.7	52.1	57.2	62.7	68.4
H（M+R）	32.6	45.6	54.2	61.2	67.1	71.6	76.8	80.9	83.7	86.6
H（栗田ら）	22.9	36.1	44.6	51.3	57.6	63.7	69.2	75.0	79.9	84.0
H（画像）	21.2	31.9	38.0	44.4	49.5	55.3	59.4	64.6	69.1	73.1
平均（M+R）	18.4	29.2	37.4	44.6	51.2	57.2	63.0	68.3	73.1	78.0
平均（栗田ら）	14.5	23.9	31.1	37.9	44.5	50.6	56.3	62.2	67.9	73.2
平均（画像）	13.8	22.3	28.9	35.5	41.4	47.2	52.9	58.6	63.9	69.2

分類正解率： M+R 法，栗田らの手法，画像特徴量空間上での近傍探索による分類，および，判別分析[7]，C4.5[5]による分類．

3.4 実験結果

上記の条件のもとで，再現率，適合率，分類正解率の測定を行った結果が，**表 3**，**表 4**，**表 5** である．表 3，表 4 では，各データセットの各テスト画像に対して検索画像数を 10 枚から 100 枚まで変化させ，各被験者ごとに，M+R 法，栗田らの方法，画像特徴量による検索を行い，各々の手法での再現率，適合率を計算し，その平均をとった．平均に関してのグラフを**図 6**，**図 7** に示す（グラフでは，155 枚までのデータを示してある）．また，同じ検索枚数での再現率，適合率を座標上にプロットしたグラフを**図 8** に示す．

再現率に関しては，検索数を変化させても，また被験者を変えても，栗田らの方法，画像特徴量による検索と比較して，M+R 法の方が優位であった．また，検索数 10 枚で，栗田らの方法に比べ，平均で，3.9 ポイントの上昇が見られ，また，画像特徴量による検索に比べ，平均で，5.6 ポイントの上昇が見られた．

適合率に関しても，再現率からの当然の帰結として，検索数を変化させても，また被験者を変えても，栗田らの方法，画像特徴量による検索と比較して，M+R 法の方が優位であった．また，平均で，検索数 10 枚で，栗田らの方法に比べ，平均で，6.9 ポイントの上昇が見られ，また，画像特徴量による検索に比べ，平均で，8.2 ポイントの上昇が見られた．

これらの再現率，適合率の上昇は，グループ間類似度情報の効果と考えられる．

分類正解率については，表 5 より，栗田らの方法，画像特徴量のみを使う分類，判別分析，C4.5 による分類に対して，ほぼ同等かそれ以上の正解率となった．これは，本手法が分類というタスクにも有効であることを示している．

また，**図 10** と**図 11** は，M+R 法でメキシコ国旗（**図 9**）をキーとして検索した場合の上位 10 枚の画像を，ある 2 人の被験者について示したものである．図 10 の被験者はメキシコ国旗を「マークのある画像」

表 4 適合率（%）
Table 4 Precision ratio (%).

検索数	10	20	30	40	50	60	70	80	90	100
筆者（M+R）	48.0	37.8	31.5	27.1	23.4	20.9	18.8	17.1	15.7	14.5
筆者（栗田ら）	36.8	31.0	25.9	22.9	20.4	18.5	17.0	15.8	14.8	13.9
筆者（画像）	34.8	27.5	22.2	19.1	16.9	15.1	13.9	13.1	12.4	11.8
A（M+R）	63.1	62.5	62.5	62.5	62.8	63.0	63.2	63.0	63.0	62.7
A（栗田ら）	60.2	60.4	60.0	59.6	59.4	59.3	59.2	59.3	59.2	59.1
A（画像）	59.0	58.1	57.4	57.5	57.6	58.4	59.0	59.3	59.4	59.2
B（M+R）	59.5	56.3	51.8	48.8	45.2	42.2	39.5	36.9	34.7	32.8
B（栗田ら）	45.6	40.9	37.8	35.5	33.9	32.7	31.5	30.5	29.7	28.8
B（画像）	44.4	39.8	36.9	34.6	32.7	31.3	29.9	28.9	27.9	27.3
C（M+R）	37.1	34.3	32.5	31.6	31.2	30.6	30.0	29.7	29.3	28.8
C（栗田ら）	32.1	30.6	29.7	29.6	29.2	28.6	28.1	28.0	27.5	27.4
C（画像）	30.9	29.6	28.6	28.2	27.8	27.3	27.0	26.2	25.9	25.7
D（M+R）	53.2	52.8	52.9	52.9	52.7	52.6	52.5	52.2	52.2	52.2
D（栗田ら）	52.8	50.8	50.7	50.8	50.7	50.7	50.4	50.2	50.1	50.1
D（画像）	52.2	52.2	52.5	52.8	52.4	52.3	52.1	51.2	51.0	51.2
E（M+R）	54.8	50.4	45.7	42.5	40.0	37.2	34.5	31.9	30.1	28.4
E（栗田ら）	44.7	39.8	35.8	33.0	31.3	29.7	28.3	27.2	26.4	25.3
E（画像）	45.8	38.3	33.3	29.5	26.9	24.9	23.5	22.6	21.7	21.1
F（M+R）	24.2	23.4	23.0	22.3	22.2	21.7	21.2	20.9	20.5	20.3
F（栗田ら）	21.0	20.2	19.9	19.9	20.1	20.0	19.7	19.8	19.8	19.9
F（画像）	21.5	19.6	18.4	18.0	17.8	17.6	17.5	17.5	17.5	17.6
G（M+R）	36.6	34.3	32.6	30.8	29.8	28.9	28.2	27.7	27.0	26.3
G（栗田ら）	35.3	31.2	29.1	28.2	27.0	26.6	25.8	25.3	25.0	24.6
G（画像）	30.0	28.3	27.0	26.4	25.7	25.0	24.5	23.6	23.2	23.0
H（M+R）	42.4	32.0	26.3	22.3	19.6	17.4	16.0	14.6	13.5	12.5
H（栗田ら）	28.4	23.3	19.9	17.7	16.2	15.1	14.2	13.6	12.9	12.3
H（画像）	27.0	21.7	17.7	15.7	14.0	12.9	11.9	11.3	10.7	10.2
平均（M+R）	46.6	42.6	39.9	37.9	36.3	35.0	33.8	32.7	31.8	30.9
平均（栗田ら）	39.7	36.5	34.3	33.0	32.0	31.2	30.5	30.0	29.5	29.0
平均（画像）	38.4	35.0	32.7	31.3	30.2	29.4	28.8	28.2	27.8	27.5

表 5 分類正解率
Table 5 Correct classfication ratio.

方法	M+R 法	栗田ら	画像	判別分析	C4.5
筆者	60.1	51.9	53.4	55.2	49.5
A	63.4	57.3	69.6	61.3	63.6
B	59.6	53.4	59.1	62.4	54.1
C	48.4	39.6	34.5	45.3	36.6
D	51.5	54.7	62.9	51.6	57.2
E	60.3	52.1	67.0	59.8	55.7
F	27.3	22.2	33.5	22.2	27.3
G	43.3	41.8	39.6	43.8	37.1
H	54.6	45.4	48.5	53.1	55.6
平均	52.1	46.5	52.0	50.5	48.5

という観点でグルーピングしており，検索結果も縞模様やマークの位置にはかかわらず上位にマークのある画像が検索されている．それに対し図 11 の被験者はこれを「縦縞で中央にマークのある画像」という観点でグルーピングしており，検索結果もやはりその観点にそった画像が多く検索されているといえよう．

さらに，類似性空間を用いない（画像特徴量空間上での）類似検索と比べると，M+R 法で検索に用いる類似性空間の次元数は最大でも利用者の行うグルーピングのグループ数 −1 である．国旗データの場合，被験者のグルーピングのグループ数は，最大 18，最小 4 であるので，類似性空間の次元数は，すべてグループ数 −1 であったとしても，最大で 17，最小で 3 となる．画像特徴量の属性数は，国旗データでは 40 属性であるから，類似性空間の次元数はこれと比べ最小でもおよそ $\frac{1}{2}$，最大でおよそ $\frac{1}{13}$ であり，検索にかかる計算時間の短縮をはかることができるといえる．

4. おわりに

本論文の貢献は以下である．

- ユーザの画像のグルーピングとそのグループ間の類似度と画像の物理的特徴量からユーザの観点を反映する類似性を学習して類似検索を行う手法を提案した．具体的には，グループ間の類似度の情報から類似性空間を多次元尺度法によって構築し，さらに画像の物理的特徴量から類似性空間への写

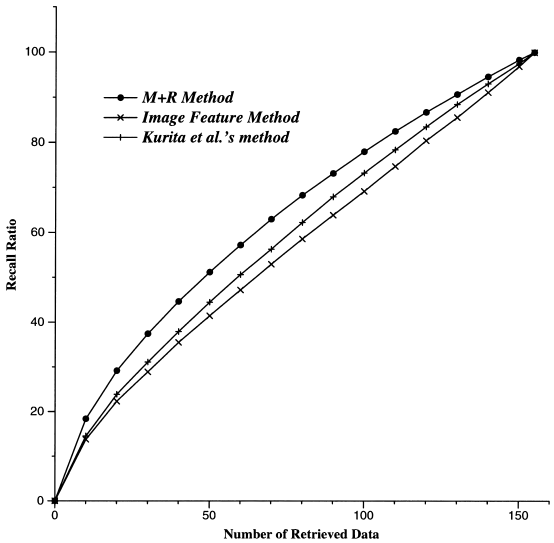


図 6 各手法の平均再現率
Fig. 6 Average recall ratio.

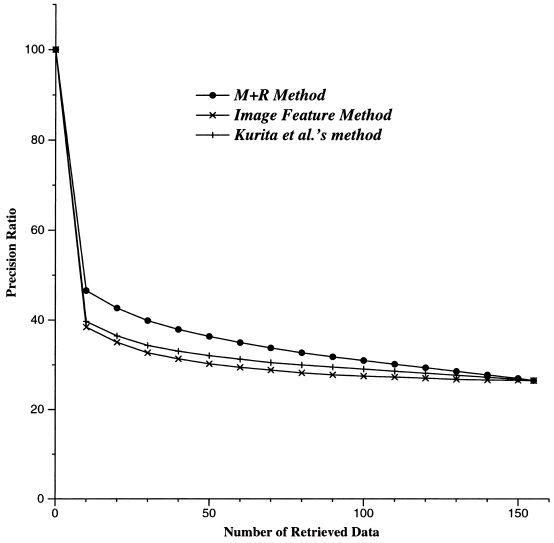


図 7 各手法の平均適合率
Fig. 7 Average precision ratio.

像を多変量線形回帰によって作り出すことにより，ユーザの観点を反映した類似検索手法を提案した．

- グループ間の類似度を用いない既提案手法[3]，画像特徴量のみを使う手法との比較を行い，再現率，適合率，分類正解率における有効性を実験的に検証した．
- 画像特徴量のみを使って類似検索する手法に比べて，類似性空間上での類似検索を行うことで，特徴量の圧縮ができることを実験的に示した．
- 従来ある分類手法と比較しても同程度か，それより優れた結果が得られることを実験的に示し，本手法の画像の分類器として使える可能性を示唆した．

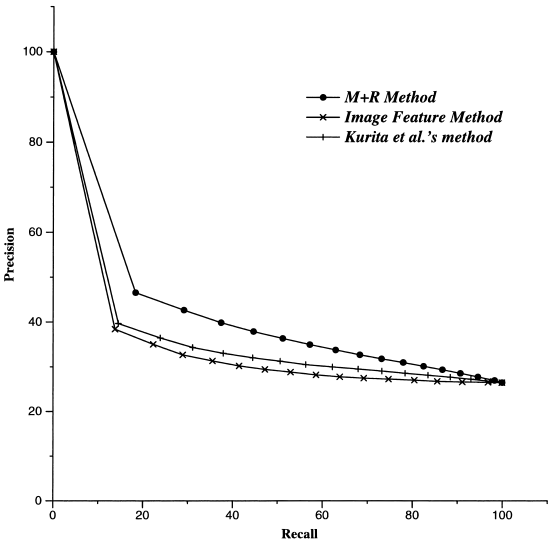


図 8 各手法の再現率と適合率の関係
Fig. 8 Recall-precision graph.

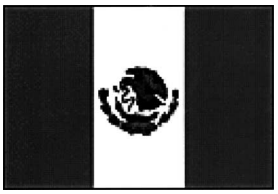


図 9 検索例のキー画像（メキシコ国旗）
Fig. 9 Example of key image (national flag of Mexico).

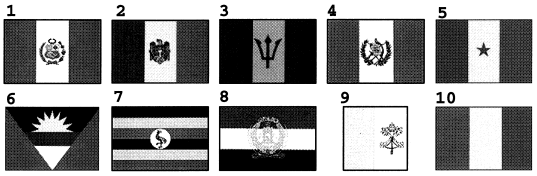


図 10 検索例 1
Fig. 10 Retrieval Result 1.

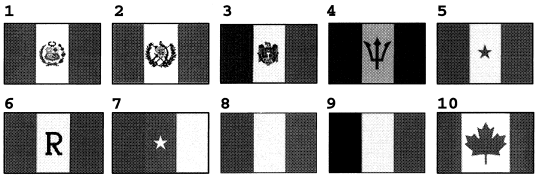


図 11 検索例 2
Fig. 11 Retrieval Result 2.

今後の課題としては以下を考えている．

より多くの画像データでの評価 今回の実験では，194 枚という比較的小規模の画像セットに対し

ての評価を行った．これは，評価のためにデータすべてを分類できるものを選んだためである．今後は，大規模なデータに対して，その1部のみを学習例として用いた評価を行う必要がある．

類似性データ獲得の方法　今回の実験では，グループ数が最大18，最小4であり，筆者の行った18グループに関しての類似度データ行列でさえ，1時間ほどで作成することができた．MDSは心理学方面でよく使われている手法であり，類似性を表すのに適した方法と考えられるが，グループ数が多くなれば，類似度データ行列の作成には，手間がかかると思われる．たとえば，現在用いている非計量的MDSであっても，グループ数をnとすれば，$\frac{n(n-1)}{2}$個の大小関係の情報が必要になるため，nが大きくなれば，その情報の量はかなり大きくなってしまう．

したがって，各グループを平面または3次元空間上へ配置するようなマンマシンインタフェースのようなものを開発する必要がある．

非線形回帰による再現率，適合率の向上の検討　本手法において，多変量線形回帰分析は，画像特徴量空間と類似性空間との写像を作るために使われているが，その写像は，必ずしも線形回帰を行う必要はなく，類似している画像どうしが，類似性空間上で距離が近くなるように写像すればよい．このため，ニューラルネットワークを用いるような非線形回帰のような方法も考えられるため，このような手法による再現率，適合率の向上が可能か検討していきたい．

M+R法の他メディア類似検索への応用の検討　本手法では，検索対象の物理的な特徴量空間と対象のグルーピングから構築された類似性空間との写像を作り出すことで類似検索を可能にしているため，検索対象が画像である必然性はなく，他のメディア，たとえば，音声などへの応用も可能であると考えられるため，そのような別なメディアでの類似検索にも使用できるか検討していきたい．

本手法の画像事例ベース推論への応用　本手法で述べたような類似画像検索を用いて問題解決をすることも可能と思われる．たとえば，医療においてX線写真によって病巣を見つけようとするときに過去の類似X線写真を参考にしたり，過去の台風の類似進路画像を参考にして，現在の台風の進路を予想することなどである．本方式のグルーピングとその類似度の部分を，たとえばX線写真においては，病気の進度によって分けることに対応させることによって，画像中のどんな属性がその病気を見つけるのに役に立つのかを決定できると思われる．今後はこのような画像事例ベース推論の応用を考えていきたい．

謝辞　本研究について議論していただいた北海道大学工学部原口誠教授，大久保好章助手，公立はこだて未来大学の川嶋稔夫教授に感謝いたします．また，マルチメディアを含む事例ベース推論の重要性について指摘していただいた（株）ドーシスの仲谷善雄氏に感謝いたします．

参 考 文 献

1) Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., Steele, D. and Yanker, P.: Query by image and video content: The QBIC system, *IEEE Computer*, Vol.28, No.9, pp.23–32 (1995).
2) Kruskal J.B., Wish, M.（著），高根芳雄（訳）：人間科学の統計学 I—多次元尺度法，朝倉書店 (1980).
3) 栗田多喜夫，下垣弘行，加藤俊一：主観的類似度に適応した画像検索，情報処理学会論文誌，Vol.31, No.2, pp.227–237 (1990).
4) 中越智哉，佐藤　健：画像特徴量空間から類似性空間への写像を利用した類似画像検索，情報処理学会研究報告，コンピュータビジョンとイメージメディア研究会，99-CVIM-115, pp.9–16 (1999).
5) Quinlan, J.R.（著），古川康一（監訳）：AIによるデータ解析，トッパン (1995).
6) 柴田正啓，井上誠喜：画像データベースの連想検索方式，電子情報通信学会論文誌（D-II），Vol.J73-D-II, No.4, pp. 526–534 (1990).
7) 塩谷　實：統計ライブラリー—多変量解析概論，朝倉書店 (1995).

(平成12年 6 月20日受付)
(平成12年10月14日採録)

（担当編集委員　田中 克己）

中越　智哉

昭和49年生．平成11年北海道大学大学院工学研究科修士課程修了．同年（株）テンアートニ入社．

佐藤　健（正会員）

昭和 56 年東京大学理学部情報科学科卒業．同年（株）富士通研究所入社．昭和 59 年英国インペリアル大学客員研究員．昭和 62 年〜平成 4 年（財）新世代コンピュータ技術開発機構出向．平成 7 年北海道大学工学部助教授．現在，同大学大学院工学研究科助教授．人工知能の基礎研究に従事．博士（理学）．