

イラストスタイルの認知的分類に基づく識別モデルの構築

栗山 繁^{1,a)}

概要：関連する見出し・説明文やキーワードに頼らず、入力画像の特徴量のみから類似内容の画像を検索する手法が多く提案されている。しかし、表現方法に一貫性を有するビジュアルコンテンツを制作するためには、画風や画調が類似したクリップアートやイラスト等を素材として用いるのが望ましいので、人のスタイルに対する認識に基づいた検索機構が望まれる。本稿では、イラストスタイルの相違度判定に基づく定量化を試み、データに基づく学習的分類法に頼らないスタイル特徴量の算出法を提案する。

1. はじめに

イラストやアイコンの画風や画調をその画像特徴量から解析する手法は、筆者の知る限り未だ提案されていない。最も関連する研究として、文字を画像とみなした特徴量からフォントの書体を識別する手法が既に提案されている。例えば、同じ書体フォントの文字多数個を密に配置した画像のテクスチャ特徴量を解析したり [1]、各文字の画像の周波数特徴量を用いて [2]、書体の種類を識別する手法が提案されている。

これらの手法は、各書体に対して計算した画像特徴量の識別器を線形判別分析等により構築しているが、イラストのスタイルに関しては書体のように決められた分類が存在しないので、識別器の学習は適用できない。ゆえに本研究では、イラストスタイルに関する人の認知的尺度を反映できるような画像特徴量を抽出し、その特徴量から計算される相違度（または類似度）のみからスタイルの類似したイラスト素材群を検索できる機構を開発する。

まず、イラストスタイルの相違度を相対的に比較する実験から認知的な距離の定量化を試み、2種類の分類画像群に対して得られた結果を報告する。次に、この定量化された認知的距離に基づき、フォント書体の識別等で用いられている画像特徴量と、新たに提案する特徴量から算出される距離に基づいてスタイルの類似した画像を識別する性能を比較する。ただし、各画像特徴量に対するスタイル識別性能は情報検索のランキング性能を評価する指標を用いた。さらに、予めスタイルが分類化された中規模なデータ集合群に対しても識別性能を比較した。

2. イラストスタイルの認知的距離の推定

多くの画像検索手法では、画像に対する注釈のテキスト情報や各種の画像特徴量から類似度（または相違度）を計算している。しかし、イラストの画風や画調を共通の言葉やキーワードで的確に表現するのは困難である。さらに、スタイルに関する感じ方には個人差があるので、一貫性のある連続的な数値量で表現するのも容易ではない。したがって本手法では、3個の画像間の相違度の大小関係を被験者に判定させて、画像全体に対する認知的な距離関係を推定する手法を導入する。具体的には、画像集合の部分的組み合わせに対する相違度の大小関係を指定するだけで全画像間の距離の値を自動的に推定する手法 [3] を導入する。これにより、各イラスト画像間のスタイルの相違度を反映した距離行列が得られ、その値から各画像を2次元パラメータ化する事によりスタイルに関する2次元座標のマップが得られる。

2.1 一般化非計量多次元尺度法

イラスト画像間のスタイルの相違度を定量化する方法として、Agarwal らによって提案された一般化非計量多次元尺度法 [4] を用いる。この手法は、イラスト画像 I_i, I_j 間のスタイルの相違度 $\delta_{i,j}$ の3個の画像間での大小関係 $\delta_{i,j} < \delta_{j,k}$ を指定して、対応するスタイルをパラメータ化したベクトル量 $\mathbf{x}_n, n = i, j, k, \dots$ を計算する。この計算には、線形計画法を一般化した凸最適化問題の計算法である半正定値計画問題（semidefinite programming）を用いる（詳細に関しては文献 [3], [4] を参照されたい）。このとき、ベクトル間のユークリッド距離 $\|\mathbf{x}_i - \mathbf{x}_j\|$ がスタイルの相違度 $\delta_{i,j}$ に相当する。

¹ 豊橋技術科学大学
Toyohashi University of Technology
^{a)} sk@tut.jp



図 1 イラストスタイルの相違度識別実験の表示例
Fig. 1 Interface of style discrimination experiment

2.2 実験用データ集合

今回の実験では、市販データ集の「子供と二人」および「動物・キャラクター」と分類されたイラスト画像集から各々 15 個と 30 個の画像を選別し、各分類から 3 個のイラスト画像を図 1 の様に表示して、そのスタイルが最も異なっている画像を選択させた。ただし、3 個のスタイルがどれも同程度に似ている、あるいは同程度に異なっており最も異なったものが特定できない場合は、その選択をスキップしてよいとした。

被験者は 21 歳から 36 歳までの成人男性 15 名に対して実施され、各被験者に対して「子供と二人」と「動物・キャラクター」の 2 種類の分類画像群（以後、簡略化して人分類、動物分類と呼ぶ）に対して、上記の選択操作を 100 回行わせた。ただし、選択がスキップされた場合は回数には含めなかった。また、スタイルの判断に関する色の影響を調査するために、最初の 50 回は濃淡変換した画像を表示し、残りの 50 回は原版のカラー画像を表示した。

2.3 認知的マップの生成結果

図 2 と図 3 に、2 種類の分類画像群に対する相違度の判定から生成された、2 次元マップ上の各画像の位置を示す。ただし、濃淡画像とカラー画像の両方の選択結果を全被験者から収集したデータを用い、MATLAB 用のライブラリ (SeDuMi パッケージ) で算出した全画像間のスタイルの距離 $\forall i, j \delta_{i,j}$ の値を用いて、古典的な多次元尺度法から 2 次元座標値（図中の○印で示されている）を算出した。

両図において、色付き楕円で囲まれた部分はマップ上の位置が近接しているため類似したスタイルと見なされる。また、各楕円内で○印（横の数字は画像番号を表す）に対応する画像を表示している。

2.4 データ集合の規模と色の影響に関する考察

図 2 の結果より、人分類の画像群に対してはスタイルの類似した画像が近い位置にあるのが確認できる。一方、動物分類の画像群に対してはスタイルの異なるものが近い位置に存在するケースが多く見られ、マップ上での距離が必ずしも認知的な距離を反映していないものと考えられる。この原因として、選択結果のデータ数の不足と、選択肢の増加に伴う各個人の選択傾向の分散化が考えられる。

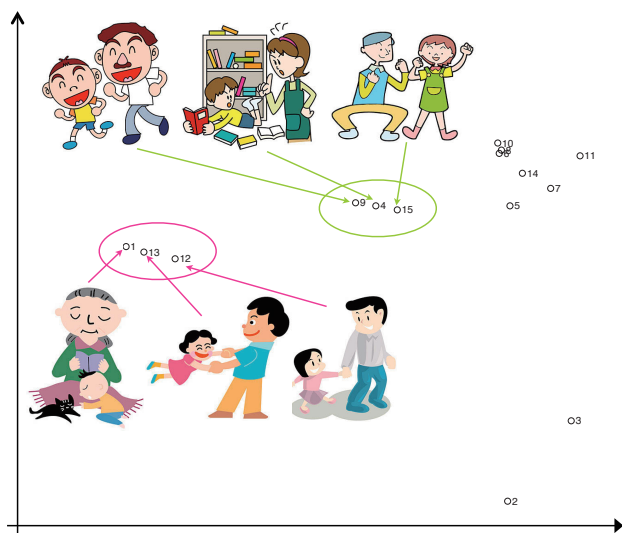
選択結果のデータ数に関しては、動物分類の画像数 30 個からの 3 個の組みに合わせに対しては $30 \times 29 \times 28 / 2 = 12,180$ 通りの距離の大小関係が存在するが、今回の 15 人の被験者の各 100 回の選択から得られるデータ数は、各選択に対して 2 通りの大小関係が与えられる点を考慮すると、総数 3000 通りとなる。この数は、全画像間での距離を求める際の自由度の 25% 程度であるが、20% 程度のデータがあれば十分な精度が得られるという結果が関連研究で報告されている [3] ので、選択数の不足が生成されたマップの信頼性の低下の原因とは断定できない。

一方、各被験者の選択データから個別に算出した距離 $\delta_{i,j}$ を要素とする行列と全データから同様に求めた行列との相関係数を求めた結果、人分類画像から計算した結果（平均値 0.73）よりもかなり低い値（平均値 0.48）となったので、判定に関する個人差の大きさが信頼性低下の原因として考えられる。

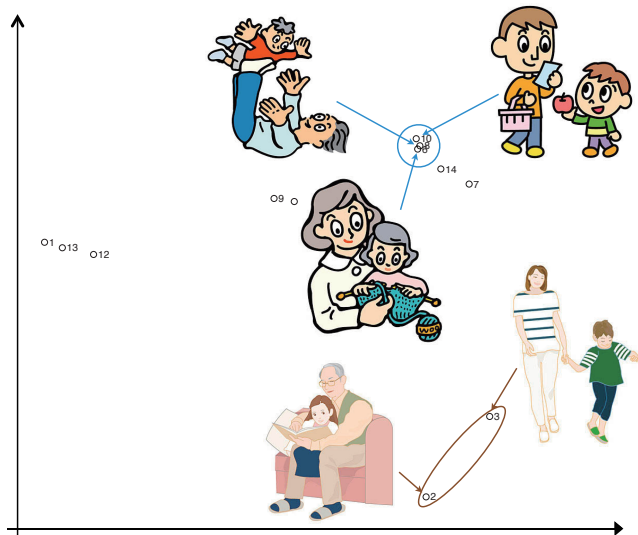
次に、濃淡画像とカラー画像を用いた比較結果を個別に用いて距離行列を算出してその相関係数を求めた結果、人分類と動物分類では各々 0.81, 0.68 という値が得られた。ゆえに、信頼性の高い人分類の画像群に関しては、色の情報がスタイル相違度の判定に与える影響は小さいと判断できる。一方、動物分類の画像に対しては高い相関は認められなかったが、上述した被験者間の相関の低さ (0.48) と比較すれば、色の影響は比較的小さかったものと考えられる。

3. スタイルに関する画像特徴量

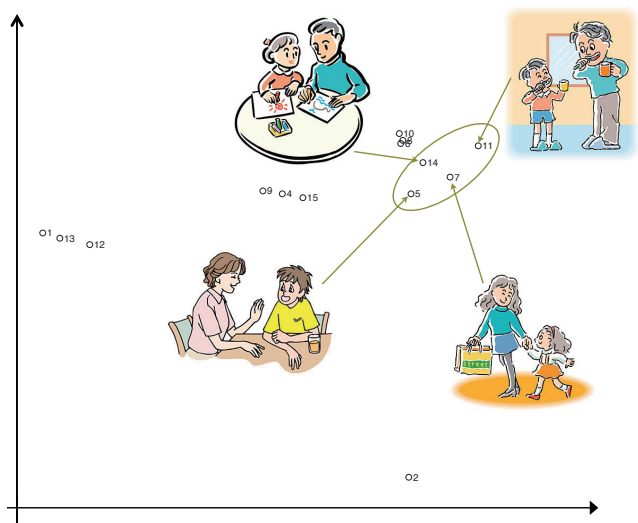
ここでは、スタイルの識別に有用と思われる画像の特徴量を調査する。スタイルは画風とも捉えられるが、一般的には筆使いや色使いにその特徴が現れると予想される。特に、イラストの輪郭線の描き方（太さ、コントラスト、曲がり具合、形状の揺らぎ等）はスタイルに影響を及ぼすものと考えられ、その変換手法も提案されている [5]。本報告では、輪郭線の情報を反映すると考えられる特徴量に焦点を当て、認知的な距離に基づく識別との相関を比較する事によりスタイルの識別に適した特徴量を調査する。



(a)

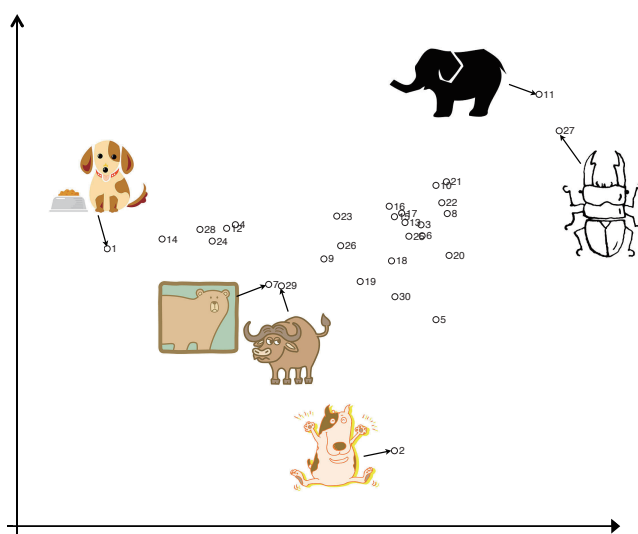


(b)

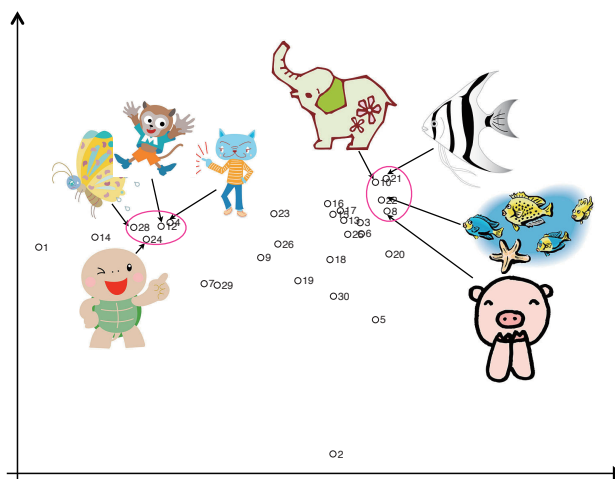


(c)

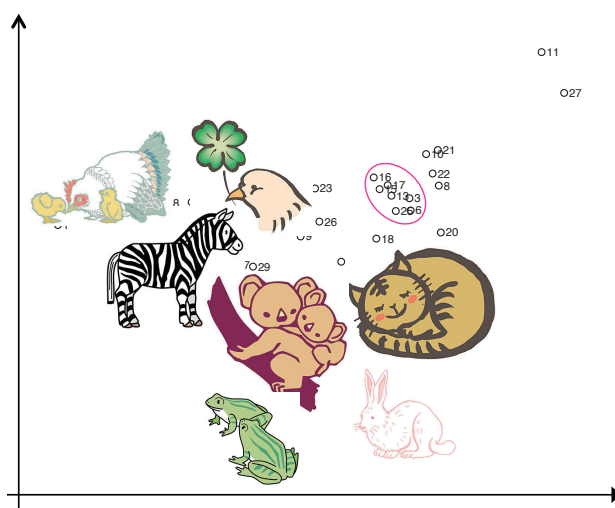
図 2 認知的距離マップの生成結果 (子供と二人)
 Fig. 2 Synthesis of cognitive distance map



(a)



(b)



(c)

図 3 認知的距離マップの生成結果 (動物・キャラクター)
 Fig. 3 Synthesis of cognitive distance map

スタイルの特徴は異なる解像度で算出したものを多重に用いるのが望ましいと考えられるので、画像をアンチエイリアシングを施した双3次内挿法を用いて半分のサイズにダウンサンプリングし、各解像度での画像で計算した特徴量を併合する共通のアプローチを採用する。今回の実験では画像のサイズを考慮して、全ての特徴量に関して多重の解像度を5層で計算した。ただし、後述する離散ウェーブレット変換に関しては、ダウンサンプリングではなく既存の分解法に依って5層分のデータを求めた。

3.1 フォントの書体識別に用いられた特徴量

イラストスタイルの識別に関連した手法として、文字画像からその書体を識別する手法が挙げられる。書体は濃淡画像であり同じ文字での輪郭形状の違いを反映しているので、その特徴量がイラストスタイルの特徴量としても利用できると予想される。ゆえにここでは、書体識別の手法で導入されている Gabor Filter を用いて算出される特徴量 [1] と、離散ウェーブレット変換を用いて算出される特徴量 [2] を取り上げる。

Gabor Filter はガウシアン基底を用いた多方向での周波数成分を解析できる手法であり、画像 $I(x, y)$ に畳み込み積分として適用されるフィルタ $G_{f, \theta}(x, y)$ が以下の式で与えられる。

$$G_{f, \theta}(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) M_{f, \theta}(x, y),$$
$$M_{f, \theta}(x, y) = \cos(2\pi f d_{\theta}(x, y)) + \sin(2\pi f d_{\theta}(x, y))$$
$$d_{\theta}(x, y) = x \cos \theta + y \sin \theta$$

本報告での実験では関連手法 [1] と同様に4方向 ($\theta = 0, \pi/4, \pi/2, 3\pi/4$) とし、周波数と偏差の値は $f = 1, \sigma = 4$ に設定した。

今回は、各方向と周波数成分の組合わせで計算されたフィルタ出力値の画像全体での平均と標準偏差の値を特徴量として用いたので、そのベクトル量は5層の多重解像度を考慮すると5層 $\times$ 4方向 $\times$ 2統計量 = 40次元となる(以後、この特徴量を Gabor と略記する)。

一方、離散ウェーブレット変換を用いた特徴量では多層(多解像度)でのウェーブレット係数を算出した。具体的には、倍直交基底である Biorthogonal spline2 wavelet [2], [6] を用いて、x-微分、y-微分、xy-2階微分のウェーブレット係数を5層分求め、各層での画像全体に対する平均値を特徴量とした(以後、本特徴量を示すのに Discrete Wavelet Transform の略語 DWT を用いる)。したがって、特徴量は5層 $\times$ 3微分量 = 15次元となる。

3.2 エッジの方向に非依存性な画像特徴量

前節で導入した Gabor Filter と離散ウェーブレット変換を用いた手法は、一定方向に関する微分量からイラスト

画像の線分の特徴量を算出している。しかしながら、イラストには様々な形状が存在し、文字画像とは異なって線分(エッジ)の方向性自体はスタイルとは独立なものと考えられる。ゆえにここでは、方向に依存しない微分特徴量として、SIFT 特徴量等で特徴点を計算するのに用いられる DoG 特徴量 (Difference of Gaussian) [7] を導入する。

DoG 特徴量の算出に用いられるフィルタ関数 $D_n(x, y)$ は以下の式で表される。

$$D_n(x, y) = G(x, y, k\sigma_n) - G(x, y, \sigma_n),$$
$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right), \sigma_n = k\sigma_{n-1}$$

本実験では、増加率を $k = 2^{1/s}$ と設定し、初期偏差値 $\sigma_0 = 1.6$ に対してスケールを3段階で変化 $s = 3$ させ、5層のダウンサンプリング画像に対する平均値を用いたので、特徴量は15次元となる(以後、DoG と略記する)。

3.3 局所的なヒストグラム分散を用いた特徴量

ここでは、線分の局所的な特徴(疎密度、コントラスト等)を、局所的に求めた濃淡ヒストグラムから算出する方法を導入する。

まず、 $N \times N$ 画素で構成されるブロック領域に関して、濃淡化された画像の濃度ヒストグラムを H 個の区間で求めて画像全体でのヒストグラムの平均値を求める。ただし、全画素値が同じ濃度区間の値となる変動の無いブロックの値は平均値の計算から除外する。ブロックとヒストグラムのサイズに関しては、経験的に良い性能が得られた値として $N = 4$ および $H = 16$ を設定した。他の特徴量と同様に5層分の多重解像度での値を併合するので、その次元数は5層 $\times$ 16区間数 = 80となる(以後、本特徴量を Local Histogram の名称から LH と略記する)。

さらに、各ブロックで計算されるヒストグラムの平均と標準偏差を計算し、その値を画像全体に関して平均化した特徴量も求めた(本特徴量は LH の平均と偏差を用いるので、LHMD と略記する)。この場合の次元数は、5層 $\times$ 2統計量 = 10となる。

3.4 提案特徴量のまとめ

上述した5種類の特徴量は、いずれも多重解像度での各濃淡画像に対する統計量(平均または標準偏差)に関する値である。各々の特徴量の性質を以下にまとめる(ただし、【】内の数字は次元数を表す)。

Gabor 【40】 雑音に頑健な多方向での微分量算出

DWT 【15】 基底関数を用いた直交性を有する多重分解

DoG 【15】 雑音に頑健な多解像度での微分量算出

LH 【80】 局所的なヒストグラムの平均分布

LHMD 【10】 局所的なヒストグラムの統計量の平均

4. 各特徴量の識別性能比較実験

4.1 ランキング性能指標を用いた比較

ここでは、同じスタイルのイラスト画像を検索する応用を考慮して、情報検索の分野でランキング結果の性能指標として広く用いられている Discounted Cumulative Gain (以後、DCG) を用いた。この指標は、 p を性能評価の際に考慮するランキングの数、 R_i を質問画像と検索された画像間のスタイルの相関度とすると、

$$DCG_p = \sum_{i=1}^p \frac{2^{R_i} - 1}{\log_2(i + 1)} \quad (1)$$

で与えられる、ただしスタイルの相関度 R_i は、被験者の比較実験から推定された認知的距離 $\delta_{i,j}$ を用いて、

$$R_i = \frac{\min(\delta_{q,i})}{\delta_{q,i}} \quad (2)$$

で計算した。ここで、 $\delta_{q,i}$ は検索質問に用いられた q 番目の画像と i 番目のランキングで得られた画像との認知距離であり、 $\min(\delta_{q,i})$ は q 番目の画像とスタイルの最も類似した画像との認知距離を表す。また、この値を正規化した Normalized DCG (以後、nDCG) は、理想的な検索結果の値に対する DCG 値の割合として計算される。

4.2 人分類画像群に対する特徴量の性能比較

2章で求めた認知的距離の推定値は人分類の結果に関しては高い信頼性が期待できるが、動物分類に関しては認知的距離自体の信頼性が低いと判断し、今回の実験では人分類で求められた認知距離のみを用いた。前章で述べた各特徴量を用いて計算した距離から各画像に対するスタイル類似度に基づく検索順位を求め、2章で推定した認知的距離を nDCG の $\delta_{q,i}$ に用いて各ランキング数 p に対して求めた結果を図 4 に示す。

ただし、本実験では全ての画像データは JPEG に変換されており、長辺が 320 画素になる様に同一の縦横比で事前にスケール変換され、全ての背景は不透明の白色に設定された。また、特徴量を計算する際には全画像は濃淡変換され、色の特徴は用いなかった。また、背景の値が 0 となる様に、白黒反転 (255-濃淡値) してから各特徴量を計算した。

4.3 スタイル分類された公開データ集を用いた性能評価

人分類のデータは 15 個の画像しか含まないので、スケラビリティの観点からは不十分である。一方、動物分類の例で示された様に、画像データの数が増加した程度でも指定すべきスタイルの相違度の比較選択数が膨大なものとなり、個人毎の主観のバラツキの影響等から、信頼性の高い距離関係の値を得るのは困難であると考えられる。

そこで、一般公開されているクリップアート画像のアー

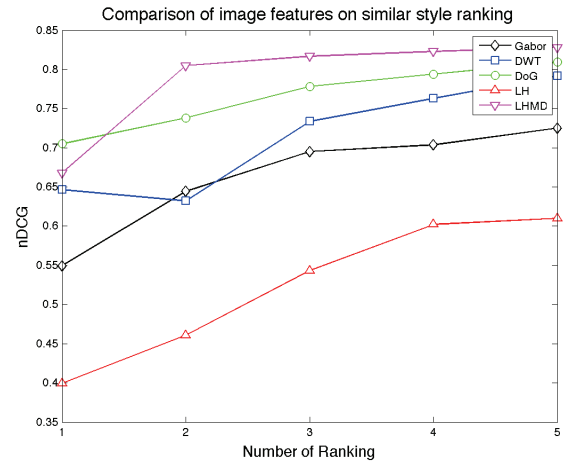


図 4 人分類画像の認知距離を用いたランキング性能比較

Fig. 4 Ranking performance with cognitive distances of human category images

カイブサイト*1からイラスト画像のデータセットを選択し、特定のイラスト画像に付与された Style 属性から画像群を再検索し、多数の画像が含まれている 30 種類のスタイルの中から 10 個の画像を選別した 300 個の画像データ集を用いた。図 5 に、同一スタイルのクリップアートの例と、全てのスタイルから無作為に選択された画像の例を示す (ファイル名の s に続く数字が、実際の Style 属性の値を示す。ただし、ハイフン後の数字は各スタイルでの画像の連番である)。

全特徴量に対して識別性能を評価した結果を図 6 に示す。この実験においても、前節と同じ条件で全画像の大きさを正規化するスケール変換と、濃淡変換+白黒変換を特徴量計算の前処理に用いている。また、検索性能の試験の際には、同じスタイル属性値の画像が検索された場合に正解画像として、nDCG を計算する際の相関度を $R_i = 1$ とし、他のスタイル属性の場合は全て $R_i = 0$ とした。

4.4 考察

比較的小規模のデータと認知的距離を用いた実験結果と、中規模のスタイルで分類されたデータを用いた結果に共通の特性は確認できなかった。しかし、局所的なヒストグラムの統計量を用いた特徴量 (LHMD) は上述した 2 通りの実験において共に高い性能を示しており、そのデータ次元数も最も小さいのでメタデータやタグとして効率良く保存できる。ゆえに、現時点では LHMD が特徴量としては最も適していると考えられるが、局所的なヒストグラムをそのまま平均化した LH は中規模データの実験においては高い性能を示しているため、ヒストグラムから抽出すべき特徴量に関して、今後さらに検討する必要がある。

各特徴量の計算に要した時間の比較結果を表 1 に示す。

*1 <http://office.microsoft.com/en-001/images/>



(a) 各行 Style 555, 655, 1191 の属性画像群



(b) 全 30 スタイルから選択された画像群

図 5 Style 属性を選択して表示されたクリップアート群

Fig. 5 Display of clip-arts by selecting a style attribute

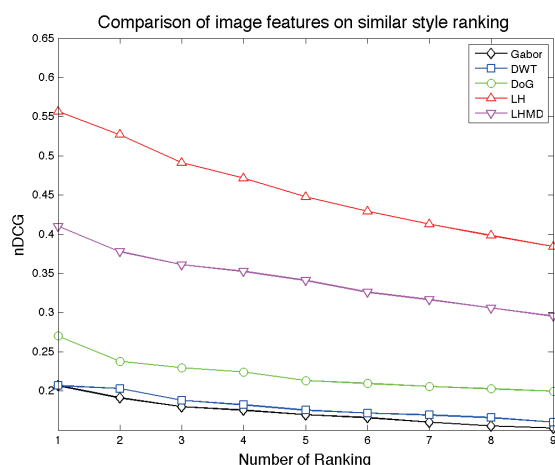


図 6 スタイル分類されたクリップアート群に対する性能比較

Fig. 6 Performance comparison for categorized clip-arts

ただし、計算時間は MacBook Pro 2.66 GHz Intel Core i7 + MATLAB R2012a の条件で計測した値である。この結果より、ヒストグラムに基づく特徴量 (LH, LHMD) はい

表 1 各特徴量の次元数と計算時間 (画像 300 個分) の比較

Table 1 Comparison of dimension and calculation time

特徴量	Gabor	DWT	DoG	LH	LHMD
計算時間 (秒)	18	7	22	75	53

ずれも他の特徴量と比較して多くの計算時間を要しているため、ヒストグラムの計算部分等を最適に並列化して計算時間を短縮化することが望まれる。

その他、今回の実験を通じて得られた知見は以下に要約される。

- スタイルの認知的な相違度の大小関係の判定のみからスタイルに関する相対的な定量化が可能になる。しかし、個人間の主観的なバラツキが増加すると共にその精度は悪化する。
- 輪郭線の濃淡、疎密、コントラスト等に関連する特徴量はスタイルの認知的な識別結果との関連性が認められる。
- 色の情報の有無はスタイルの認知的な識別には大きく影響しない。
- スタイルの識別には画像全体での平均値等の統計量が特徴量として機能する。
- 輪郭線の方に依存した特徴量はスタイルの識別には大きく影響しない。
- 特徴量の次元数と識別性能には有意な相関は認められない。

5. おわりに

本報告では、イラスト画像のスタイルを識別する特徴量を提案し、実験的に得られた認知的距離とスタイル分類されたデータセットの双方に対して識別性能を実験的に比較した。

今回の実験により、局所的なヒストグラムの特徴量 (LH, LHMD) は他の特徴量に対して、中規模のデータに対しては nDCG の指標値で比較的に優れた性能が確認できたが、その絶対的な値を考慮した場合、検索システムとしては未だ実用レベルには達していない。また、手法の信頼性の向上のために、数と種類において大規模なデータセットを用いた検証実験が必要なので、ベンチマーク用のデータセットの整備も今後の重要な課題である。

今回は輪郭等の線のスタイルを反映できる特徴量にのみ着目したが、スタイルには色彩の特徴も含まれるので、認知的な影響力は少ないという結果が得られたものの、色情報の活用方法も検討すべき課題である。また、描画物体の形状や色に関する高次の特徴量も考慮する必要があるが、スタイルの特徴を階層的に分類することにより、構造的な識別手法を導入できる可能性がある。また、相違度に相当する距離計算にも非多様体学習を用いた手法を導入する等の拡張が必要であろう。

今回の実験では、イラスト画像に着目してスタイル分類を試みたが、背景が単純な画像であれば汎用的に用いられる可能性がある。例えば、カタログからスタイルの共通した商品のみを絞り込んで検索する機能等への応用が考えられる。また、イラストのスタイルを書体のスタイルと連動させる応用等も将来的な検討課題である。

参考文献

- [1] Zhu, Y., Tan, T., and Wang, Y.: Font recognition based on global texture analysis, *Pattern Analysis and Machine Intelligence*, IEEE Transactions on, 23 (10), pp.1192–1200 (2001)
- [2] Ding, X., Chen, L., and Wu, T.: Character Independent Font Recognition on a Single Chinese Character, *Pattern Analysis and Machine Intelligence*, IEEE Transactions on, 29 (2), pp.195–204 (2007)
- [3] Wills, J., Agarwal, S., Kriegman, D. and Belongie, S.: Toward a perceptual space for gloss, *ACM Trans. Graph.*, 28 (4), pp.103:1–103:15 (2009).
- [4] Agarwal, S., Wills, J., Cayton, L., Lanckriet, G., Kriegman, D., and Belongie, S.: Generalized non-metric multidimensional scaling, In *Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS)* (2007).
- [5] Freeman, W., T., Joshua, B., Tenenbaum, J., B., and Pasztor E., C.: Learning style translation for the lines of a drawing, *ACM Transactions on Graphics*, 22, pp.33–46 (2003).
- [6] Castleman, K.R., *Digital Image Processing*. Prentice Hall (1996).
- [7] 藤吉弘亘: Gradient ベースの特徴抽出 - SIFT と HOG -, 情報処理学会研究報告 CVIM 160, pp.211–224 (2007).