

クリップアート画像のスタイル分類特徴量

Style-Oriented Features for Classifying Clipart

- スタイルに基づくランキング -

- Toward Style-Based Ranking -

栗山 繁

Shigeru KURIYAMA, E-mail: sk@tut.jp

豊橋技術科学大学 情報・知能工学系

Department of Computer Science and Engineering, Toyohashi University of Technology

1 はじめに

プレゼンテーションの挿絵等に用いられるイラスト画像はクリップアートと呼ばれ、近年ではその画像を無償で利用できる Web サイトの数も増加している。多くのサイトでは画像ごとに付与されている描画内容を表すキーワードを用いて所望する画像を検索できる。また、Google 等の画像専用の検索機能で「イラスト」や「クリップアート」等の単語と伴に所望の描画内容を表す単語を入力して検索すれば、大量の関連画像が得られる。しかしながら、検索された大量の画像から意図する描画様式（以後、これをスタイルと呼ぶ）を有する画像を探す際には、描画のスタイルに基づく絞り込み検索やランキングの機能が望まれる。

本研究では、クリップアート画像にそのスタイル属性を表すタグデータを付与し、画像を検索する際に内容を表すキーワードと共にスタイル属性も参照し、スタイルの類似順に検索結果をランキングするための基盤技術を開発する。以後、本稿ではラスタデータ形式のクリップアート画像を、単に「画像」と呼ぶ。

クリップアート画像の内容（コンテンツ）に基づく検索手法 [1, 2] が提案されているが、実用的な性能には未だ達していない。フォント書体の種別を識別する手法 [3, 4] は本提案手法と類似の目的で設計されているが、分類に用いる特徴量が 1 種類のみであり、様々な内容を含むイラスト画像全般を扱うには、特徴量を多面的に捉える必要がある。

著者の先行研究 [5] において、濃淡値の確率分布（ヒストグラム）による特徴量が、既存の手法と比較して画像スタイルの識別に高い性能を示す実験結果が得られた。また、画像間の認知的な距離を心理実験により数値化したデータを用いて、その特徴量が認知的距離を精度良く再現する事を確認した。本稿ではその特徴量の計算方法の基本的枠組みを継承しながら、エッジ、テクスチャ、および色に関連する特徴量を新たに導入し、それらの組み合わせによりスタイルのランキング性能を大幅に向上させた結果を報告する。さらに、特徴量間の干渉を軽減しながら特徴量の次元数を削減するために、有限個の識別器の出力から構成され

る特徴量をスタイル属性を表すタグデータとして用いる手法を提案し、そのランキング性能を定量的に評価する。

2 スタイル識別に適した特徴量

2.1 局所的な輝度分布による特徴量

先行研究 [5] では、フォント画像からその書体を識別する手法 [3, 4] で使用された特徴量を参考にして、5 種類の特徴量を比較した。その結果、描画スタイルの識別には多重解像度の特徴量が適しており、各解像度での輝度の確率分布を局所ブロック領域内で平坦となる部分を除いて求めた値を特徴量とする事により、識別性能が向上するという知見を得た。

この特徴量は、カラー画像をグレースケール変換した輝度画像 $\mathbf{I}$ から生成した画像ピラミッドの、各層でのヒストグラムを計算して得られる。具体的には、 N_b 段階の k 番目の濃淡レベル $\mathbf{I} \in [B(k-1), Bk]$, $B = 256/N_b$ の画素数を $b_k(\mathbf{I})$ とすると、 m 次元の特徴ベクトル $\mathbf{F} = [f_1, f_2, \dots, f_m]$ は以下の式で与えられる。

$$f_{(i-1) \times N_b + k} = \frac{b_k(\mathbf{I}^i)}{\sum_k b_k(\mathbf{I}^i)}, \quad k = 1, 2, \dots, N_b, \quad (1)$$

$$\mathbf{I}^1 = \mathbf{I}, \quad \mathbf{I}^{i+1} = \downarrow_{\frac{1}{2}} \mathbf{I}^i, \quad i = 1, 2, \dots, L \quad (2)$$

ただし、輝度画像 $\mathbf{I}^i$ は $\mathbf{I}^{i-1}$ を双三次補間を用いた半分のダウンスケール変換 $\downarrow_{\frac{1}{2}}$ で得られる。特徴量の次元数は画像ピラミッドの層数 L に対して $m = N_b L$ となり、本手法では経験的に $L = 5$ 層の画像ピラミッドに対して $N_b = 16$ 段階のヒストグラムを使用したので、80 次元の特徴ベクトルとなった。ただし、ヒストグラムの値は確率分布となる様に、値の総和が 1 となる様に正規化している。

アイコンやシンボリックなクリップアート画像は特徴となる情報が疎に分布しており、それを効果的に捉える必要がある。また本手法では、局所的に色変化の無い平坦部分の面積比率はスタイル特徴としては重要でないとして仮定し、特徴量の計算には反映させない。したがって、特徴抽出の際

に画像を 4×4 画素のブロック領域に分割し、各ブロック内での輝度変化が無い平坦な領域はヒストグラムの計算から除外した。図1の画像の赤色部分は、各解像度の画像において除外される平坦な領域を示しており、解像度が高い程、除外される領域の割合が増大する事が確認できる。本稿では、この平坦領域を除外して計算される輝度画像の各層でのヒストグラムに基づく特徴量を輝度 (Intensity) 特徴量と呼ぶ。

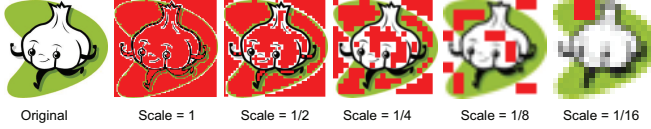


図 1: 特徴量計算から除外される平坦領域部分

以後の節では性能のさらなる向上を目指して、輝度を他の画像特徴に置き換えた特徴量を提案する。

2.2 エッジ特徴量の導入

似顔絵の特徴を描画線の特徴から識別する研究 [6] や描画線のスタイルを置換する研究 [7] が提案されている様に、線 (エッジ) の特徴は描画スタイルに影響すると考えられる。先行研究 [5] では、エッジ (微分) 特徴に関連した値として DoG (Difference of Gaussian) とウェーブレット係数の識別性能を調査したが、高い性能は得られなかった。ここでは Laplacian フィルタの出力値をエッジの特徴量として用い、輝度の特徴量と同様に 5 層の画像ピラミッドに対して算出した値のヒストグラムを新たな特徴量とする。ただし、自然画像のエッジ強度分布が正規分布に近づく性質をイラスト画像にも仮定し、フィルタ出力値の対数値を等間隔に分割した $N_b = 16$ 段階のヒストグラムを用いる。したがって、エッジ画像のレベル $\hat{\mathbf{I}} \in [\exp(\hat{B}(k-1)) - 1, \exp(\hat{B}k) - 1]$, $\hat{B} = \log(256)/N_b$ の画素数を $\hat{b}_k(\hat{\mathbf{I}})$ とすると、特徴ベクトルは以下の式で与えられる。

$$f_{(i-1) \times N_b + k} = \frac{\hat{b}_k(\hat{\mathbf{I}}^i)}{\sum_k \hat{b}_k(\hat{\mathbf{I}}^i)}, \quad k = 1, 2, \dots, N_b \quad (3)$$

$$\hat{\mathbf{I}}^i = |\Delta \circledast \mathbf{I}^i|, \quad i = 1, 2, \dots, L$$

ここで、 Δ は非零の対角成分を有するラプラシアン演算の行列、および $\circledast$ は畳み込み積分を表す。

ここでも輝度特徴量と同様にイラストの疎な情報分布を考慮して、 $\hat{\mathbf{I}}^i$ の値変化の無い平坦な 4×4 画素のブロック領域はヒストグラムの計算から除外した。以後、上記のエッジに基づく方法で算出される 16 段階 $\times$ 5 層 = 80 次元のベクトル値をエッジ (Edge) 特徴量と呼ぶ。

2.3 テクスチャ特徴量の導入

先行研究 [5] で性能を調査した、テクスチャ画像の解析に用いられる Gabor フィルタで算出された特徴量は、スタイルに対しては高い識別性能が得られなかった。これはウェーブレット特徴量と同様に、その局所的な方向性を含む特徴が描画スタイルの識別には悪影響を及ぼした結果と考えられる。ゆえに、テクスチャ解析に適した特徴量として、方向に非依存な LBP (Local Binary Pattern) を導入し、前節で提案した特徴量と同様に 5 層の画像ピラミッドに対して平坦領域を除外して計算したパターンヒストグラムを特徴量として用いる。

具体的には、各画素に対して一画素分の距離の 8 近傍の画素値 (ただし、斜め方向は線型補間した値を用いる) との輝度の大小関係から 256 パターンの LBP を得る。このパターンに回転の不変性と、隣接ビットの符号変化が 5 回以上のパターンを除外するという緩い一様性を適用する事により 28 パターンに縮退し、各層に対してヒストグラムを計算して 28 パターン $\times$ 5 層 = 140 次元の特徴量を得る。ただし、濃淡の相違を反映できる情報を扱うために、輝度情報のビット幅を 1/2 および 1/4 に縮退した値に対しても同様に LBP を計算してヒストグラムを算出した。すなわち、各層に対して以下の式で得られる 3 点の画像に対して計算したパターンヒストグラムの平均値を特徴量として用いた。

$$\mathbf{I}_1^i = \mathbf{I}^i, \quad \mathbf{I}_{m+1}^i = \lfloor \mathbf{I}_m^i / 8 \rfloor, \quad m = 1, 2 \quad (4)$$

ただし、 $\lfloor \cdot \rfloor$ は打ち切り関数を表す。また、この特徴量でも平坦な色領域を除外するために、中心と全近傍の画素が同じ輝度となる場合には LBP 計算の対象外とした。以後、上記のテクスチャに基づく方法で構成される 140 次元のベクトル値をテクスチャ (Texture) 特徴量と呼ぶ。

2.4 色特徴量の導入

これまでに導入した特徴量は全て、グレースケール変換した輝度画像から算出される特徴量であるが、色の特徴も描画スタイルと深く関連するものと考えられる。認知的な相違度を反映する色空間として、 $L^*a^*b^*$ や HSV が挙げられるが、今回は RGB 画像をこれらの色空間に変換した 0 から 1 の範囲の画素値 $\mathbf{I}_c$ を 12 区間 $\mathbf{I}_c \in [(k-1)/12, k/12]$, $k = 1, 2, \dots, 12$ (ただし、最後の区間は 1 も含む) に分割したヒストグラムで特徴量を構成した。ただし、輝度 (L)、彩度 (S)、および明度 (V) に関しては、その上下限の値 (0,1) にはスタイルを特徴付ける特別な意味が含まれると仮定し、最初と最後の区間は 0 と 1 の値だけの画素数を計算し、残りは区間 $\mathbf{I}_c \in [(k-2)/10, (k-1)/10]$, $k = 2, 3, \dots, 11$ (ただし、 $k = 2$ の区間は両端とも开区間とする) の画素数からなるヒストグラムを用いた。

色特徴量のスタイル分類性能を調査するために、RGB, $L^*a^*b^*$, および HSV の各色空間の成分毎のヒストグラムを結合したベクトルで特徴量を構成し、次章で述べる実験と同じ設定でランキング性能を比較した結果、HSV 色空間での特徴量が最も高い性能となった。以後、この HSV 色空間での各色成分の特徴量を結合して得られる 36 次元のベクトル値を、カラー (Color) 特徴量と呼ぶ。

3 特徴量の性能比較

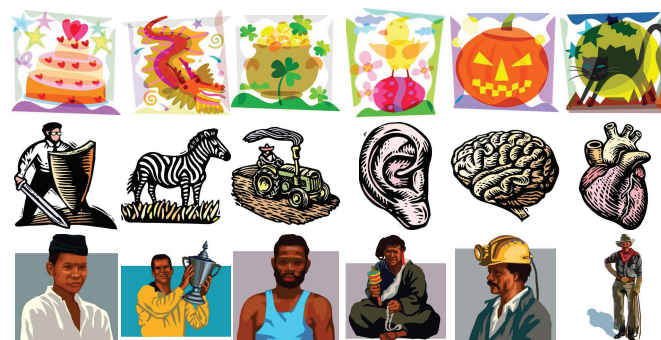
3.1 ランキング性能の比較実験

ランキング性能の比較実験には先行研究 [5] と同様に、マイクロソフト社のクリップアート用サイトからスタイル属性が予め数値として定義された画像群から、図 2 に示す 30 種類の各スタイルに対して 10 点ずつの画像を選別し、総数で 300 点の試験画像を jpeg 画像に変換して収集した。



(a) スタイルで分類された 30 種類の画像データ

画像下の数値は、3 種類の特徴量を全て用いて各スタイルで計算された評価 (MAP) 値を示す。



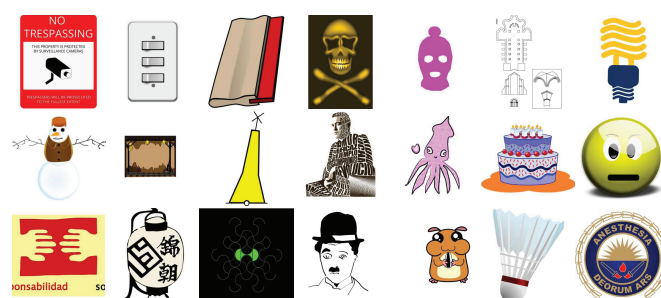
(b) 同一スタイル画像群 (各 10 点中の 6 点を抜粋)

図 2: 利用した試験用画像データ群

これらのスタイル分類された画像群に加え、国内向け有料の画像素材集から人やキャラクタが含まれたイラスト画像¹(DEX) と、フリーで配布されているクリップアート素材集²(OpenClipart), および海外向け有料のクリップアート素材集³ (ArtExplosion: Elena Garcess 氏のプロジェクトページから研究用に利用可能) から各々無作為に 1,000 点を選択し、総計 3,300 点の画像データを性能評価に用いた (図 3 参照)。ただし、ArtExplosion は png, 他の素材は jpeg 形式であり、全ての画像データは長辺が 320 画素になる様に事前に等比でスケール調整している。



(a) DEX イラスト素材



(b) OpenClipart



(c) Art Explosion

図 3: 混入させる画像素材集 (各 1,000 点からの抜粋)

評価実験として、スタイル分類された図 2 の画像群から抜粋した 1 点のクエリ画像に対して他の全画像との特徴量間の距離 (すなわち、非類似度に相当する値) を算出し、その値を昇順に並べた画像を検索結果と見なす。

ランキング性能の評価には、各クエリ画像に対する平均適合率を平均した値である、以下の式で表される MAP 値

¹<http://www.dex.ne.jp/product/mantan/gman.html>

²<http://openclipart.org>

³<http://www.novadevelopment.com>

(Mean Average Precision) を用いた.

$$\text{MAP} \stackrel{\text{def}}{=} \frac{1}{Q} \sum_{q=1}^Q \left(\frac{1}{S_q} \sum_{i=1}^{S_q} \text{prec}(i) \right),$$
$$\text{prec}(i) = \frac{i}{\text{rank}(i)}$$

ただし, Q は実験に用いたクエリ画像の数, S_q は q 番目のクエリ画像と同じスタイルの画像数, および $\text{prec}(i)$ は i 番目に高い順位で $\text{rank}(i)$ 番目にランキングされた画像の適合率である.

特徴量間の距離には, 実験に用いた全画像の特徴量の標準偏差の逆数で差分値が重み付けされる正規化ユークリッド距離を用いた. 表 1 に, 図 2 の全 300 点の画像に対して求めた MAP 値の比較結果を示す. ただし, 図 3 の各 1,000 点のクリップアート素材集を個別に含めた場合と, それらの全てを含めた結果が示されている.

表 1: スタイルのランキング性能 (MAP 値) 比較

特徴量	Edge	Texture	Color	Intensity
次元数	80	140	36	80
DEX	0.43	0.44	0.27	0.27
OpenClipart	0.44	0.38	0.30	0.28
ArtExplosion	0.51	0.46	0.35	0.26
全データ素材	0.37	0.34	0.24	0.21

この結果より, 先行研究での特徴量 (Intensity) と比較して, 今回提案するエッジおよびテクスチャに関連した特徴量の相対的に高いランキング性能が確認できる. また, ランキング性能は混入した素材集の種類に若干の影響を受けているが, 性能の大小関係の傾向に大きな変化は認められない. また, 全 3 種類の素材集を混入した場合でも性能の大幅な低下は見られなかった.

各特徴量から算出される画像間の非類似度の値の相関係数を調査した結果, 先行研究で用いた輝度特徴量とカラー特徴量の間には高い相関が認められた. ゆえに, 輝度特徴量から得られる情報には冗長性が含まれ, エッジやテクスチャの特徴量と比較して特徴量単体としての性能も低い. ゆえに, 以後の実験では輝度特徴量は含めない事とした.

以上に述べた, スタイルの抽出に適したエッジ, テクスチャ, およびカラーの特徴量の組み合わせに対して, 上記の実験と同じ条件 (ただし, 全ての素材集を用いた場合のみ) を用いて計算した MAP 値を表 2 に示す. ただし, E, T, C は各々スタイルのエッジ, テクスチャ, およびカラー特徴量を表し, + 記号はそれらの特徴量の結合を表す.

図 2(a) の各画像の下の数値は, 全特徴量の組み合わせで得られた, 各スタイルに含まれる画像のみで計算した MAP 値を, 左上から右下に降順で示している.

表 2 の結果が示す様に, 色特徴との組み合わせによりも

表 2: 各特徴量の組合せによるランキング性能 (MAP 値)

特徴量の組合せ	E+C	T+C	E+T	E+T+C
次元数	116	176	220	256
全データ素材	0.40	0.41	0.45	0.46

エッジとテクスチャ特徴量の組み合わせに対して, より大きな性能の向上が認められる.

全ての特徴量を用いた場合には, 有意な性能向上は見られなかった. これは, スタイルの種別によって分類に影響する特徴が異なるので, 特徴量間の影響の打ち消し合いの負の効果が考えられる. したがって, スタイルの種別毎に特徴量の重み付けを最適化させる必要がある. この最適化に関しては, 4 章で言及する.

3.2 スタイルと内容に関する特徴の分離性能

画像データを撮影あるいは描画されている内容に基づいて検索する一般的な手法は, SIFT や HOG 等の局所特徴量やそれらを辞書化したベクトル量を用いて, 内容の一致する物体を検索する. 一方, 本研究においては線分や色のスタイルの類似度の識別を対象としているので, 類似度の定量化において内容類似性の影響を排除する事が望ましい. ゆえに本節では, 本研究で提案した特徴量群が, スタイルと内容の双方の類似度に対して及ぼす影響の度合いを比較調査する.

実験用データにはマイクロソフト社の同じサイトを利用し, 21 歳から 36 歳までの 15 名の被験者により以下の要領で実験用画像を収集した. まず, 適当なキーワードで画像を検索し, その画像と類似する画像群を同サイトの類似画像検索機能を用いて図 4 の様に表示させる. 被験者はその画像群から描画スタイルが最も類似していると判断した画像を選択し, その画像に対して再び類似検索を実行する. 次に被験者は表示画像群から描画内容が類似した画像を選択し, 以後同様にスタイルが類似した画像と内容が類似した画像を交互に選択して 6~12 点の画像系列を収集する. この収集作業により 78 系列, 総数 700 点の画像を試験用に収集した (その系列の一部を図 5 に示す).

表 3 に分類性能の評価に MAP 値を計算した結果を示す. 各々の類似度に対する性能を比較した結果, 本研究で提案した特徴量はスタイルの類似度への影響を及ぼし, 内容の類似度に与える影響はかなり小さい事が判明した. ただし, 内容の類似度に関して得られた MAP 値は, 正解画像を無作為に選択した場合の推定値; $(\sum_{i=1}^{699} 1/i)/699 \approx 0.0102$ より若干大きい値となっており, これらの特徴量は内容の類似度と全く無関係ではない事も示唆された.



図 4: 類似画像検索機構による結果表示例



図 5: スタイルと内容の類似性で収集された画像群

4 次元圧縮に基づくスタイル・タグ生成

画像データサイトの検索機能にスタイルに基づくランキングやフィルタリング機能を実装するには、スタイル属性を示すタグデータを各画像に対して登録する必要がある。しかしながら、提案する 256 次元の特徴量をそのままの形式でタグデータとして付与する場合には、内容を表すキーワードのタグデータと比較して大きなデータ量が必要となる。

そこで本章では、特徴量の次元数を効果的に削減するために、以下の仮定：

スタイル特徴の分解可能性

全てのスタイル特徴は有限個のスタイル種別毎に定義される類似度の組み合わせで表現可能である。

に基づいた新たな特徴量を提案する。この仮定が正しければ、スタイルの種類が増大しても一定次元数の特徴量でランキングやフィルタリングが可能になる。具体的には、事前にスタイル分類された試験画像から構築された一定数の識別器の出力結果から、全ての画像スタイルを表現できる

表 3: スタイルと内容の類似性に基づく性能評価 (MAP) 値

特徴量	Edge	Texture	Color	E+T+C
スタイル類似	0.36	0.30	0.37	0.44
内容類似	0.051	0.048	0.036	0.060

特徴量を計算する。このアプローチは Classeme [8] と呼ばれ、自然画像のカテゴリ分類においてはその有効性が既に実証されている。

4.1 識別器の出力集合による特徴量の構成

本手法では、あらかじめスタイル分類された画像を教師データとした SVM を識別器 $\phi_i(\mathbf{F}) \subset \mathbb{R}$ として用いる。ただし、SVM の計算にはガウス放射基底関数のカーネルを用い、基底関数の標準偏差値は経験的に 10 に設定した。この SVM 識別器は判定結果のラベルではなく識別超平面からの符号付き距離の値を出力する関数とし、 c 番目のスタイルに関して生成された SVM の出力関数 $\phi_c(\mathbf{F})$ を用いて特徴量 $\tilde{\mathbf{F}}$ を、

$$\tilde{\mathbf{F}} = [\phi_1(\mathbf{F}), \phi_2(\mathbf{F}), \dots, \phi_N(\mathbf{F})] \quad (5)$$

で算出する。ただし、 N は用いる識別器の数であり、識別器の関数 $\phi_i(\mathbf{F})$ は識別超平面からの符号付き距離を出力する点に注意されたい。

この特徴量 $\tilde{\mathbf{F}}$ は教師データのスタイルの種別数と同次元数となるが、SVM の学習は一つのスタイルと他の全てのスタイルの識別平面を計算するので、学習させるクラス種別数の増大に伴い種別は異なるが実際にはスタイルの類似する複数のクラスが含まれる事により、各 SVM 識別器の性能が低下してしまう。しかしながら、スタイルの重複の無い少ないクラス種別数で構成した特徴量は、次元数の不足によって十分な分類性能が得られなくなるというトレードオフの関係が存在する。そこで本手法では、各識別器 $\phi_n(\mathbf{F})$ を構築する際にはエッジ、テクスチャ、およびカラーに関する特徴量を個別に適用する。ゆえに、識別器の最大数は学習に用いるスタイル数の 3 倍の値となる。

この識別器の個数 N を十分なランキング性能が得られるまで削減すれば、スタイル属性として保存するデータ量が削減できる。さらに、各実数値を離散化して整数で表現することにより、データ量のさらなる圧縮が可能になる。

識別関数の値は $-4.0 < \phi_i(\mathbf{F}) < 2.5$ の範囲を超える可能性が極めて低い事を実験的に確認したので、以下の式により実数値を整数値 $z_i \in \mathbb{Z}$, $0 \leq z_i \leq 255$ に変換する。

$$z_i = \begin{cases} 0 & \text{for } \phi_i(\mathbf{F}) \leq -4.0 \\ \lfloor \frac{255}{6.5}(\phi_i(\mathbf{F}) + 4.0) \rfloor & \text{for } -4.0 < \phi_i(\mathbf{F}) < 2.5 \\ 255 & \text{for } \phi_i(\mathbf{F}) \geq 2.5 \end{cases}$$

この離散化により、スタイル属性を保存するために必要なデータ数は N バイトに大幅に削減できる。以後、この N バイトのデータ $\mathbf{z} = [z_1, z_2, \dots, z_N]$ を**スタイルタグ**と呼称し、次節の性能評価実験においては全てこのスタイルタグの特徴量を用いる。

4.2 ランキング性能評価

スタイルタグによるランキング性能は識別器の数 N に影響を受ける事が予想されるので、識別器の数に伴うランキング性能の変化を調査する。ここで、図2に示される30種類のスタイル分類された画像データからSVM識別器を構築し、同スタイル画像での出力の平均値から異スタイルの出力の平均値を減算した値が大きい識別器から優先的に使用した。

この実験では識別器の構築に試験データの影響を排除するために、図2の30種類のスタイル中の1種類の10点の画像を試験データとして、識別器は他の29種類に含まれる全画像から構築し、この交差検証を30回実行して計算したMAP値 (Style Tag (29x3)) をランキング性能評価に用いた。また、3種類の特徴量毎に分割して識別器を構成して生成したスタイルタグ効果を調べるために、マイクロソフト社の同じクリップアート用サイトから、さらに60種類の画像を各スタイル10点ずつ収集し、試験用データを除いた計89種類の画像で構築した識別器から構成されるスタイルタグで計算したMAP値 (Style Tag (89)) と比較した。

図6に、これら特徴量に用いる識別器の数 (次元数) に対する各MAP値の変化を示す。ただし、3.1節での実験と同様に3種類の素材集 (総数3,000点) の画像を混入し、比較のために主成分分析 (PCA) を用いて同じ次元数に縮退した特徴量に対するMAP値も求めた。PCAの縮退次元の座標 (固有値ベクトル) の値は、上記のSVM識別器と同様に試験データ以外の29種類のスタイルに含まれる全画像から計算した。さらに比較のために、256次元の特徴量をそのまま用いたMAP値 (Raw Feature) も描いている。

図6の結果が示す様に、スタイルタグは同じ次元圧縮数に対して、PCAと比較して大きなランキング性能が得られている。また、スタイルタグは45程度の次元数を確保すれば、元の特徴量を用いた場合と同等のランキング性能が得られ、67以上の次元数では性能値は増減で振動している。したがって、今回の実験条件の場合、十分なランキング性能を得たい場合には70次元 (バイト) 程度に、タグとしてのデータ量削減を重要視する場合には45次元 (バイト) 程度に設定するという指針が得られる。

また、識別器を3種類の特徴量毎に構成する効果は、Style Tag (29x3) の方が全ての次元数に置いて Style Tag (89) の

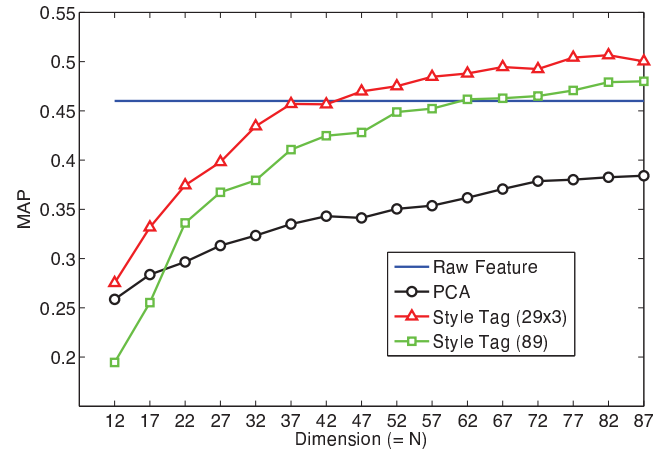


図6: 特徴量の次元数によるランキング性能の変化

値を上回っている事により確認できた。これは、スタイルの種別が増加するに伴いSVM識別器の性能が低下したのが原因と考えられるが、追加した60種類の教師データを等分して1対29のクラスを分類する識別器を構成した場合の性能を調査した結果、大幅な性能低下が見られた。したがって、教師データのクラス数よりも、その種類の構成が大きな影響を及ぼす事が示唆された。

スタイルタグの離散化がランキング性能に与える影響を調査するために、離散化を施さない特徴量でもMAP値を計算した結果、図6の各次元数に対するその増加分の最大値は 6.1×10^{-4} であり、離散化有りの方が性能が若干上回る次元数もあった。ゆえに、離散化に伴う性能の悪化は無いものと考えられる。

図7、に上記実験で用いた70次元のスタイルタグで上位の5番以内に同じスタイルの画像がランキングされなかった例を示す。ただし、最左のクエリ画像に対してランキング順に左から右へ検索結果を配置している。これらのクエリ画像は、3.1節の実験においても低い評価 (MAP) 値を示しているが、図7の結果が示す様に、同じスタイルの教師データではない場合でも、スタイルの類似した画像が検索されていると判断できる。あるいは、検索された画像間でのスタイルの類似性も確認できる。

図8に、図2(a)の画像群に対してスタイルタグの値を基に多次元尺度法を用いて2次元に次数削減した結果を示す。

図9に混入した3種類の画像集から無作為に選択したクエリ画像に対する検索結果を示す。ただし、SVM識別器の学習には図2の30種全てのスタイル分類された画像から計算された70次元のスタイルタグを用いた。また全ての例において、最下段には著者が不自然と判断したランキング結果を示している。

また、図10には、各画像集から無作為に選んだ100点の画像に対して同じ識別器を用いてスタイルタグを計算した結果を、t-SNE法 [9] を用いて可視化した例を示す。



図 7: 教師データの同一スタイルがランキングされない例

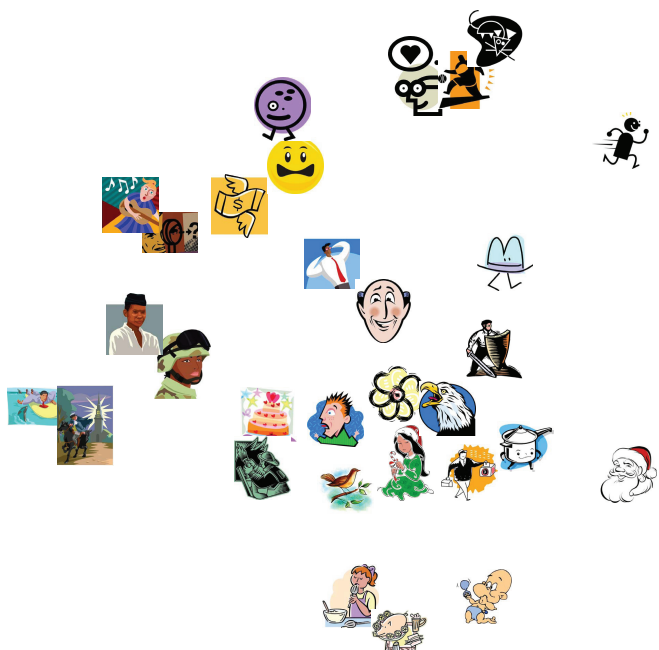


図 8: 多次元尺度法を用いたスタイルタグの類似度配置

5 むすび

本稿ではクリップアート画像をその描画スタイルに基づいてランキングする手法を提案した。スタイルに特化した特徴量をタグデータとして同時保存すれば、クリップアート画像を検索する際にスタイルの類似度に基づくランキング表示が可能となる。今回提案した識別器の出力に基づくスタイルタグは、各画像に対して 70 バイト程度のデータ量でスタイル属性を登録できる。これは、5 万点以上の SVG 画像を有する Openclipart の全画像データが非圧縮で約 4.5GB (圧縮で約 1GB) である点を考慮すると、画像データのサイズに対して 0.08 % 程度という少ない比率であり、十分に実用的な範囲であると考えられる。

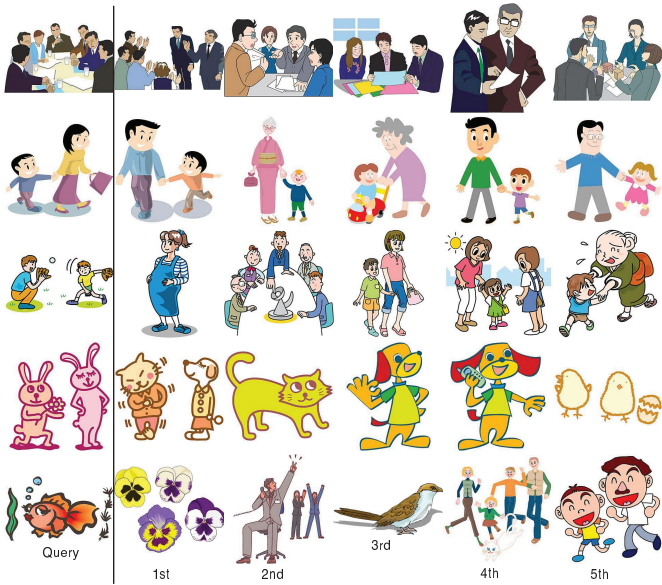
識別器を構築する際には、教師データ画像のスタイルの組み合わせを任意に選択できるので、最適な性能が得られ

るスタイルの種類の組み合わせを自動選択する機能が望まれる。また、ユーザの対話的な操作によって主観的なスタイルの類似関係を推定し、その結果を反映させた識別器を構築できれば、個人的なカスタマイズの可能性も考えられる。その方法論と検証に関しては今後の課題である。

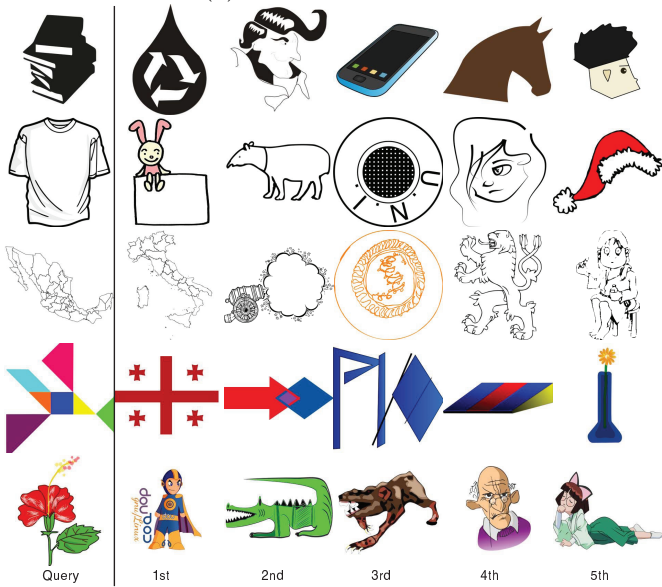
クリップアート画像の多くは、ラスタ形式の画像ではなく SVG や WMF 等のベクタ形式で与えられる事が多いので、ベクトル量から直接的にスタイル特徴を計算する方法に関しても今後検討していく予定である。

参考文献

- [1] M. J. Fonseca, B. Barroso, and P. Ribeiro, and J. A. Jorge, "Retrieving ClipArt Images by Content," In Proceedings of the 3rd International Conference on Image and Video Retrieval (CIVR '04), 2004.
- [2] P. Martins, R. Jesus, M.J. Fonseca, N. Correia, "Clip art retrieval combining raster and vector methods," 11th International Workshop on Content-Based Multimedia Indexing (CBMI), pp.35-40, 2013.
- [3] X. Ding, L. Chen, and T. Wu, "Character Independent Font Recognition on a Single Chinese Character," Pattern Analysis and Machine Intelligence, vol.29, no.2, pp.195-204, 2007.
- [4] Y. Zhu, T. Tan, and Y. Wang, "Font recognition based on global texture analysis," Pattern Analysis and Machine Intelligence, vol.23, no.10, pp.1192-1200, 2001.
- [5] 栗山, "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," 情報処理学会研究報告, no.2013-CG-152(3), pp.1-7, Sep.2013.
- [6] I. Berger, A. Shamir, M. Mahler, E. Carter, and J. Hodgins, "Style and abstraction in portrait sketching," ACM Trans. on Graph., no.32(4), pp.55:1-55:12, 2013.
- [7] W. T. Freeman, J. B. Tenenbaum, and E. C. Pasztor, "Learning Style Translation for the Lines of a Drawing," ACM Trans. on Graph., vol.22, no.1, January 2003.
- [8] T. Lorenzo, S. Martin, and F. Andrew, "Efficient Object Category Recognition Using Classes," European Conference on Computer Vision, 2010.
- [9] L.J.P. van der Maaten and G.E. Hinton, "Visualizing High-Dimensional Data Using t-SNE," Journal of Machine Learning Research 9, pp.2579-2605, 2008.



(a) DEX イラスト素材



(b) Openclipart



(c) Art Explosion

図 9: スタイル・タグデータを用いた検索結果例



(a) DEX イラスト素材



(b) Openclipart



(c) Art Explosion

図 10: t-SNE を用いた可視化例 (100 点抜粋)