

推薦論文

イラスト画像のスタイル識別子生成

栗山 繁^{1,a)}

受付日 2015年1月16日, 採録日 2015年5月9日

概要: 画像検索では各データに付与された関連キーワードをクエリとして用いる方法が一般的であるが、画像認識の分野では類似のシーンや物体を含む画像をクエリとして用い、その特徴量から類似内容の画像を検索する手法が数多く提案されている。イラスト素材となるクリップアート等の画像は、描かれている内容だけでなく作画のスタイルでも絞り込み検索ができれば、画風や画調に統一感のあるコンテンツを制作する際に役立つものと考えられる。本論文では、作画スタイルが類似したクリップアートやイラスト等の素材画像を検索する際に活用できるスタイル識別子の生成方法を提案する。数千枚のイラスト画像群を用いた実験により、50バイト以下の識別子で既存手法よりも高い検索精度が得られることを確認した。

キーワード: イラスト検索, 作画スタイル, LBP 特徴量, HSV カラー空間, スタイル識別子

Generating Style Descriptor of Illustrations

SHIGERU KURIYAMA^{1,a)}

Received: January 16, 2015, Accepted: May 9, 2015

Abstract: In addition to an image retrieval with relevant keywords or queries, many content-based retrievals were proposed by introducing various image features and pattern matching algorithms to select images having similar scenes or objects. Style-based image ranking is helpful in narrowing down illustrative material or clipart among those retrieved with keywords, and such capability can support the authoring environment of visual contents to obtain the consistency of drawing styles. This paper therefore introduces a new type of descriptor for illustrations that are suited to retrieve or rank them with respect to their drawing styles. Using thousands of illustrations, we experimentally found that our style descriptor of less than 50 bytes outperformed an existing method in retrieval accuracy.

Keywords: retrieval of illustrations, drawing style, local binary patterns, HSV color space, style descriptor

1. はじめに

イラスト画像の素材はプレゼンテーション、広告、教育、および絵葉書等、ビジュアルなコンテンツの制作に幅広く利用されているが、このクリップアートとも呼ばれるイラスト画像を無償ダウンロードできる共有型リポジトリ^{*1}やWebサイトの数も年々増加している。

これらサイトの多くは各画像に付与されたキーワードを用いて所望する画像を検索できる。また、通常の検索サイトにおいて「イラスト」や「クリップアート」等の単語と

ともに所望の描画内容（たとえば、関連テーマや物体の名称）を表す単語をクエリに用いれば、関連するイラスト画像が大量に検索されて一覧表示される。しかしながら、検索で得られる画像の数が増加するにともない、その中から意図する作画の様式（以後、これをスタイルと呼ぶ）を有する画像を選択する際の候補も膨大となるので、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。

本論文では簡易な計算で得られるイラスト画像の特徴量から、作画スタイルの属性を表す低次元かつ低容量の識

¹ 豊橋技術科学大学情報・知能工学系
Toyohashi University of Technology, Toyohashi, Aichi, 441-8580, Japan

a) sk@tut.jp

本論文の内容は2014年6月のVisual Computing/グラフィクスとCAD合同シンポジウムにて報告され、グラフィクスとCAD研究会主査により情報処理学会論文誌ジャーナルへの掲載が推薦された論文である。

*1 <http://openclipart.org>

別子を自動生成する手法を提案する。そして、その識別子をクエリ（検索語）の一部として与えることにより、スタイル特徴が類似した画像群を順位付けで検索できる機構を開発する。本論文で提案する手法は、以前に開発した手法 [1], [2] の色特徴量の計算方法を改良し、特徴量の正規化と簡略な重み付けに学習を導入し、さらには特徴量の次元数と容量をラベルデータを用いない計算で大幅に削減する。また、その後には発表された手法 [3] との性能比較を通じて、本識別子の優位性を考察する。

本研究ではラスタデータ形式で保存されているイラスト画像のみを対象として扱うので、特に明記されてなければ以後の章では画像とはイラスト画像を指すものとする。

2. 関連研究

イラスト（またはクリップアート）画像の検索手法として、画像の内容（コンテンツ）に基づく検索手法 [4], [5] がすでに提案されている。しかし、イラストの場合は同じ内容のシーンや物体に対してもそれらを描く際の画調・画風が作者によって異なる場合が多いので、イラストの描画内容に基づく検索の精度は実シーンや実物体を撮影した画像と比較すると、いまだ実用的なレベルには至っていない。

フォント書体の種別を識別する手法は作画スタイル特徴の識別と問題設定が似ているが、それらの手法で用いられている Gabor フィルタ [6] や B-spline ウェーブレット [7] に基づく単体の特徴量だけでは、多様な要因に影響されるイラストのスタイルを特徴付けるには不十分である。

形状の部分的な特徴を解析して 2 次元 [8] および 3 次元物体 [9] のスタイルの類似度を判定する手法が提案されている。これらは描画対象の幾何学的な特徴のみに着目するものであり、本研究で着目するような作画スタイルの特徴をとらえるには適していない。

作画スタイルの生成に関連した研究例として、描画線の特徴をイラスト間で転写する手法 [10] や、熟練の専門家が描いた肖像画のストローク情報から描画のスタイル特徴を学習し、一般人が描く肖像画に転写する手法 [11] が提案されている。これらの手法は線描画の細やかな特徴や順序をとらえることはできるが、配色も含めた画像全体のスタイルを要約して特徴付けるのには適していない。

色のスタイル特徴を画像編集に利用した研究例としては、色の一貫性を得るための手法 [12] や、Factor Graph を用いて新たな配色パターンを生成、評価する手法 [13] 等が提案されている。これらの手法は配色スタイルの類似性の判定に利用できるが、スタイル特徴の識別方法やその精度に関しては言及されていない。

近年、画像のカテゴリ分類の研究分野において、芸術的な絵画作品のスタイルに基づく分類性能を調査した研究 [14], [15] が報告されている。これらの手法では、ロココ、バロック、現代アート等の歴史的な位置付けに基づく

分類をスタイルとして学習させており、分類ラベルを用いた Convolutional Neural Network による教師有りの深層学習で最も高い検索性能が得られている。

一般的なイラスト画像にはそのスタイルを分類する修辭法が存在しないので、学習のための分類ラベルを作成する必要があるが、作画スタイルという概念的な特性を一般的な言葉で定義するのは困難であることが予想される。さらに、アイコンや単一の物体を表すだけのイラスト画像は、芸術的な絵画作品と比較すると内包する情報量が少ないので、スタイル特徴の抽出方法も独自の視点で設計するのが望ましい。すなわち、スタイルを精度良く識別するには、多くのイラスト画像の共通な特性である、疎に分布した特徴にも対応できる手法が必要となる。

最近提案されたイラストスタイルの類似度計算法 [3] は、作画スタイルの特徴をとらえる複数種類の画像特徴量を導入しており、その一部は本研究で提案する特徴量と類似した計算で求められる。この研究では、三つ組画像からスタイルの類似したペアを選択する心理実験の結果を学習データとして用いており、人の判断と同等の精度でスタイルの類似したペアを判定できた結果が報告されている。この手法では、距離尺度の学習で求めた重み値が 0 に近いデータ次元を省くことによる多次元特徴量の低次元化も提案している。本提案手法では、文献 [3] の手法よりも高性能なスタイル識別を種類の少ない基本的な画像特徴量だけで達成し、疎な分布を考慮したパワー正規化とラベルデータを用いない次元削減手法の有効性を示す。

3. スタイル識別の画像特徴量

3.1 描画線の特徴量

線（エッジ）の太さや疎密、および曲率等の幾何学的特徴は作画スタイルに影響すると考えられるので、本手法ではカラー画像をグレースケール変換して得られる濃淡画像 \mathbf{I} にラプラシアン・フィルタを施して得られるエッジ画像 $\hat{\mathbf{I}}$ の強度分布を特徴量として求める。ただし、多重解像度での特徴をとらえるために 5 層の画像ピラミッドでのエッジ画像におけるヒストグラムを計算する。自然画像のエッジ強度はラプラス分布に近い疎な分布となることが報告されている [16] が、本手法でのエッジ強度のイラスト画像群に対する平均分布（ただし、4.1 節の実験で用いた全画像を対象とし、第 1, 3, 5 層に対する分布割合を示す）は、**図 1** のようになる。この結果より、イラスト画像のエッジ強度にも自然画像と同様な疎な分布が認められるので、その対数値を等間隔に分割した 16 段階のヒストグラムを用いる。

ゆえに、 i 層目のエッジ画像 $\hat{\mathbf{I}}^i$ のヒストグラムの k 番目の区間 $[\ell_k, \ell_{k+1})$ での画素数を $b_k(\hat{\mathbf{I}}^i)$ 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。

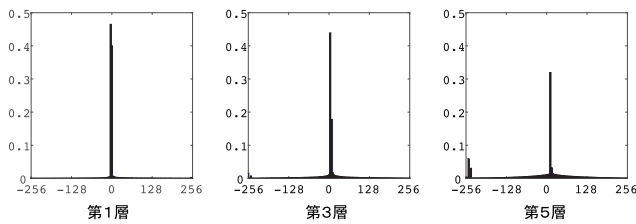


図 1 全画像に対するエッジ強度の平均分布図
Fig. 1 Distribution of average edge intensities.

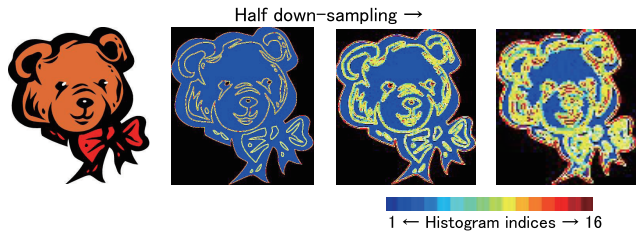


図 2 エッジ画像のヒストグラムを用いたエッジ特徴量
Fig. 2 Edge feature obtained by histogram of edge images.

$$f_{16(i-1)+k} = \frac{b_k(\hat{\mathbf{I}}^i)}{\sum_k b_k(\hat{\mathbf{I}}^i)}, \quad \hat{\mathbf{I}}^i = |\Delta \circledast \mathbf{I}^i|,$$

$$\mathbf{I}^i = \downarrow_{\frac{1}{2}} \mathbf{I}^{i-1}, \mathbf{I}^1 = \mathbf{I}, \quad i = 2, 3, 4, 5,$$

$$\ell_1 = 0, \quad \ell_k = \exp\left(\frac{k \log(256)}{16}\right) - 1, \quad k = 2, 3, \dots, 16$$

ここで、 Δ は 3×3 の非零成分で構成されるラプラシアン演算の行列、 $||$ は絶対値、および \circledast は畳み込み積分を表す。また、濃淡画像 \mathbf{I}^i は \mathbf{I}^{i-1} を双三次補間を用いた $1/2$ 倍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とヒストグラムの区間数の積となるので、この特徴ベクトルは $80 (= 5 \times 16)$ 次元となる。

イラスト画像の背景領域は濃淡変化がなくスタイル特徴が含まれないので、元画像で背景と同色（今回の実験では白色）の画素でのエッジの値（すなわち、ラプラシアン・フィルタの出力）はヒストグラムの計算から除外する。ゆえに、画像ピラミッドの各層で計算されるヒストグラムは、背景と異なる色の画素数で除算される正規化が施される。以後、このエッジ強度に基づいて算出されるヒストグラムのベクトル値をエッジ特徴量と呼ぶ。

エッジ特徴量の算出例を図 2 に示す。ただし、黒色部分は背景色として除外された領域を表す。

3.2 テクスチャパターンの特徴量

テクスチャ解析に適した特徴量は作画スタイルとしての陰影や繰返しパターンの特徴をとらえられる。ゆえに本手法では、テクスチャ解析に用いられる Local Binary Pattern (以後、LBP) を導入し、前節で提案したエッジ特徴量と同様に、このパターンのヒストグラムを 5 層の画像ピラミッドに対して算出した値をテクスチャ特徴量とする。

グレースケール変換された各画素の値 p_c に対して一画

素分だけ離れた 8 近傍画素の値 p_n (ただし、斜め方向の値は線型補間を用いる) との濃淡値の大小関係から以下の式で 8 ビット (= 256) のパターンの LBP を得る。

$$b(n) = \begin{cases} 1 & p_c < p_n, \quad n = 1, 2, \dots, 8, \\ 0 & p_c \geq p_n \end{cases} \quad (1)$$

ただし、 $b(n)$ は 2 値パターンの n ビット目の値を示す。

先行研究 [1] において、方向依存のテクスチャ特徴をとらえられる Gabor フィルタ特徴量の識別性能が低かったため、既存手法 [3] と同様にパターン生成の際の近傍画素の周回の開始位置と方向には無関係な回転不変性のインデクス化を導入する。さらに、LBP では隣接するビット値の変化する回数が多いパターンの出現頻度はきわめて低いので、そのようなパターンはヒストグラムの計算から除外する。実際には、変化回数の上限值を 2 回に設定する一様パターンが広く用いられる [3] が、本手法ではこれに変化回数が 4 回のパターンも加えたインデクス化を用いて、より高い識別精度を得た。この理由として、イラスト画像は自然画像と比較して雑音成分や乱雑なパターンが少ないので、変動回数の多いパターンもスタイル特徴として有効に働いたことが考えられる。回転不変性と独自に設定した一様性の基準により、8 ビットで算出される LBP の値が 28 通りのパターンに分類されてインデクス化されるので、この 28 区間のヒストグラムによって特徴量を算出する。

LBP は式 (1) で表されるように、中心画素と全近傍画素の濃淡値が同じ ($\forall n, p_c == p_n$) 場合と中心画素がある近傍画素よりも高い ($\exists n, p_c > p_n$) 場合の違いを見分けられないという問題点がある。イラスト画像の場合は局部的に濃淡が変化しない平坦な領域が多く存在するので、この曖昧性が識別性能にもたらす悪影響は無視できない。本手法ではこの問題を解決するために、濃淡変化のない領域ではスタイル特徴が存在しないものと見なし、 $\forall n, |p_c - p_n| < \delta$ となる画素 p_c での LBP インデクス値をヒストグラムの計算から除外した。その結果、識別精度の向上を実験的に確認した。ただし、濃淡値の同一性を判定する閾値は、JPEG 等の画像圧縮にともなう微小な変動成分を考慮して $\delta = 5$ に設定した。この問題の別解法として、式 (1) の条件を等号と不等号に完全分離して、3 値の局所パターンでエッジ・テクスチャを識別する手法 [17] が考えられるが、この 3 値に拡張された LBP で実験した結果、識別性能に有意な改善は認められなかった。

以後、この LBP のヒストグラムで構成される $140 (= 5 \times 28)$ 次元のベクトル値をテクスチャ特徴量と呼び、その算出例を図 3 に示す。ただし、黒色部分は計算から除外された平坦領域を表す。

3.3 配色の特徴量

上述したエッジとテクスチャの特徴量はグレースケール

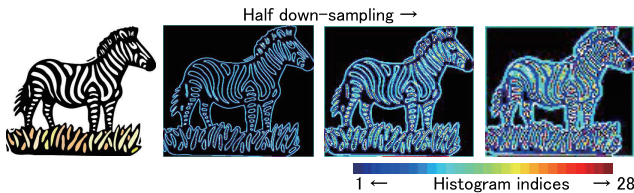


図 3 LBP のヒストグラムを用いたテクスチャ特徴量
 Fig. 3 Texture feature obtained by histogram of LBP.

変換した濃淡画像から算出されるが、配色の特徴も作画のスタイルと深い関わりを有する。今回の実験では認知的な類似度を反映できる量として、RGB 色空間の線形変換で得られる HSV 色空間での値をヒストグラム化したベクトル値で特徴量を構成した。ただし、各要素の値は 0 から 1 の範囲で正規化されている。

色味を決定する要素は H (Hue; 色相) と S (Saturation; 彩度) の組合せであり、これらを 8 段階の値に分割してヒストグラムを計算する。一方、V (Value; 明度) は色の濃淡成分に相当し、この値の微妙な変動により濃淡表現でのスタイルを的確にとらえられると考え、特徴を詳細に計算するためにヒストグラムを 2 倍の 16 段階とする。

多くのイラスト画像においては彩度の高い純色が多用されており、明度の高い色も配色の重要な特徴となる。また、白黒のみで表現されたイラスト画像では、これらの最大/最小値が頻出する。したがって、色のスタイルでは彩度と明度の最高値を個別の特徴量としてとらえるのが望ましい。以上の理由により、S (彩度) と V (明度) の最大 (= 1) と最小 (= 0) の値は個別の区間としてヒストグラム化する。

また、H (色相) の値は彩度と明度の値が低い場所では人の視覚には見分けがつかないことを考慮し、 $S < 0.02$ または $V < 0.02$ の範囲の色の場合は不定値として、便宜的に $[-\infty, 0)$ の区間に分類する。さらに、色相は周期的な値であり $H > 0.8$ の範囲の色の頻度がきわめて低いことを実験的に確認したので、 $0.75 \leq H$ の範囲を 1 つの区間として扱う。

以上の条件により、HSV 色空間での色特徴量は、以下の区間幅 $b_k = [e_k, e_{k+1})$ で計算されるヒストグラムとして与えられる。

$$\begin{aligned}
 H : e_1 &= -\infty, e_{k=2, \dots, 8} = (k-2)/8, e_9 = \infty, \\
 S : e_1 &= 0, e_2 = \delta, e_{k=3, \dots, 10} = (k-2)/8, e_{11} = \infty, \\
 V : e_1 &= 0, e_2 = \delta, e_{k=3, \dots, 18} = (k-2)/16, e_{19} = \infty,
 \end{aligned}$$

ただし、S と V の最初の区間が 0 に近い値のみを含むように、0 値からの微小変動量は δ は、元の RGB 色空間での解像度 $1/256 \approx 4 \times 10^{-3}$ よりも若干小さい値 $\delta = 10^{-3}$ とした。色特徴の場合は画像ピラミッドに基づく多重解像度の計算は不要であり、原スケール画像 I のみでヒストグラムを計算すればよいので、 $36 (= 8 + 10 + 18)$ 次元のベクトルとなる。以後、このベクトル値をカラー特徴量と呼び、

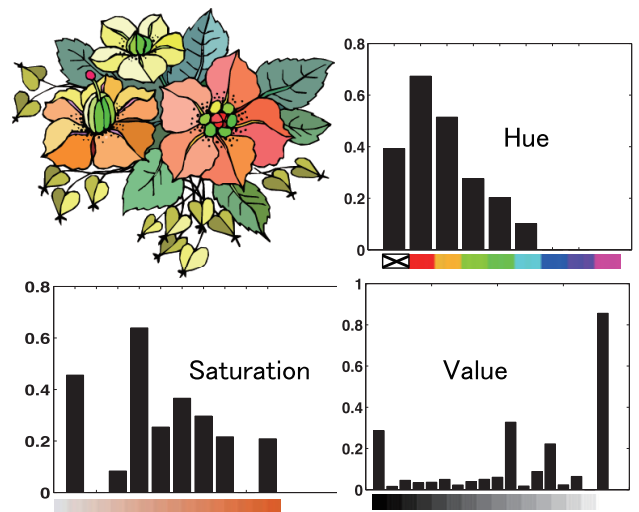


図 4 HSV 色空間でのヒストグラムを用いたカラースタイル特徴量
 Fig. 4 Color style obtained by histogram in HSV color space.

その生成例を図 4 に示す。

3.4 特徴量の正規化と統合

撮影画像や芸術的絵画と比較して、イラスト画像はその表現形態により内包する情報量が少ないものと推察されるが、実際に算出した特徴量ベクトルを観測した結果でも値の偏在が確認された。一般的に疎なベクトル空間での内積を用いた距離計算は、密なベクトル空間での計算と比較して精度が低下する傾向があるので、以下のように $0 < p < 1$ の冪乗でパワー正規化された特徴量 \hat{f} に変換する。

$$\hat{f} = [\hat{f}_1, \dots, \hat{f}_m], \hat{f}_i = \text{sign}(f_i)|f_i|^p,$$

ここで、 m は各特徴量の次元数、 sign は符号関数、および $||$ は絶対値を表す。冪乗 p の値は、特徴量の疎な状態を反映して決定する変数であるが、その値としては $p = 0.5$ が推奨されている [18], [19]。本手法では後述するように、この最適値を教師データを用いて学習させた結果、エッジ、テクスチャ、およびカラーの各特徴量に対して各々 $p = 0.4, 0.6, 0.25$ を設定した際に識別精度が最大となった。この冪乗の値が小さい程、より疎な状態のベクトル値を密な値に変換できるので、カラー特徴量が最も疎な状態であることが示唆された。

エッジとテクスチャの特徴量は画像ピラミッドの各層で求めたヒストグラムを統合して得られるが、距離尺度の計算の際に種類の異なるヒストグラム間の影響を均一化するために、それらの大きさを揃えるのが望ましい。したがって、特徴ベクトルの長さを単位長に揃える L2 正規化; $\hat{f}_i = \hat{f}_i / \|\hat{f}\|$ を施す。ただし、 $\|\cdot\|$ は L2 ノルムを表す。

すべての特徴を統合する際には、この正規化されたエッジ、テクスチャ、およびカラーの特徴量ベクトル $\bar{f}_E, \bar{f}_T, \bar{f}_C$ を重み付きで以下のように \bar{f}_{all} に統合する

$$\bar{f}_{all} = [\bar{f}_E, \bar{f}_T, w \bar{f}_C],$$

ただし、エッジとテクスチャの特徴量はカラー成分と比較すると同様な特性を有すると見なし、カラー特徴に対する重み w のみを設定した。

各特徴量を算出する画像1枚あたりの平均所要時間を計測した結果、エッジ特徴 26ms (ミリ秒)、テクスチャ特徴 73ms、およびカラー特徴 42ms であった。ゆえに、各画像に対して平均して約 141ms で全特徴量を計算できるので、与えられるクエリ画像の特徴量を瞬時に計算して検索に用いることが可能である。ただし、計算時間は MacPro 3.7GHz Quad-Core 上の Matlab (R2014b) で計測した。

提案手法ではすべての特徴を統合すると 256次元となり、既存手法よりも次元数が 87次元多くなっているが、これは5章で述べる次元削減手法を導入することにより、最終的には既存手法の圧縮結果よりも低次元数に変換できる。

4. スタイル特徴量の性能比較

4.1 三つ組画像を用いたスタイル比較実験

本章では、前章で提案した特徴量のスタイルの識別性能を既存手法 [3] と比較する。ただし、使用するデータセットは既存手法の著者が公開している実験用データセット*2を用いた。

このデータセットは、スタイル属性に基づいて分類されているクリップアート素材集 (マイクロソフト社が提供。以後、MS データセットと呼ぶ) と、スタイル属性はないが Mechanical Turk を用いた選択実験により類似度の相対的な比較結果が得られているイラスト群*3 (以後、MTurk データセットと呼ぶ) から構成される。ただし、全画像は長辺が約 400 画素になるように事前にスケーリングされており、MS データセットに関しては透明化されている背景色を MTurk データセットの背景色と同じ白に設定した。

MTurk データセットでは、1,000 枚の画像から無作為に選択した三つ組の画像に対して最初の画像のスタイルが 2, 3 番目の画像のどちらと似ているかを被験者に判定させ、その中から選択一致度の高かった結果のデータを学習用の 25,540 サンプルと試験用の 633 サンプルに分割している。一方、MS データセットでは 3,591 枚の画像の中から同じスタイル属性の 2 枚とそれ以外のスタイル属性の 1 枚を無作為に選び出し、三つ組画像の比較判定結果として学習用の 25,000 サンプルと試験用の 10,000 サンプルを人工的に生成している。以上の設定は、既存研究 [3] での比較実験と同等の条件である。

表 1 に、提案手法と既存手法 [3] の特徴量を用いた場合の各データセットでの三つ組比較の正解率 (すなわち、試験用サンプルと類似判定が一致した割合) を示す。ただし、既存手法における陰影 (Shading) と曲線特徴 (Stroke) の

表 1 三つ組画像のスタイル比較の各特徴量での正解率 (学習なし)
Table 1 Correct rate of style comparison with triplet images for each feature (without learning).

データ	比較手法	エッジ	テクスチャ	カラー	全特徴量
MTurk	既存手法	0.68	0.72	0.68	0.75
	提案手法	0.71	0.75	0.65	0.70
MS	既存手法	0.90	0.89	0.90	0.94
	提案手法	0.89	0.92	0.84	0.88

表 2 三つ組画像のスタイル比較の各特徴量での正解率 (学習あり)
Table 2 Correct rate of style comparison with triplet images for each feature (with learning).

データ	比較手法	エッジ	テクスチャ	カラー	全特徴量
MTurk	既存手法	0.69	0.73	0.70	0.80
	提案手法	0.71	0.75	0.68	0.75
MS	既存手法	0.90	0.90	0.90	0.95
	提案手法	0.89	0.93	0.86	0.95

特徴量は本手法でのエッジ特徴量に対応付け、テクスチャおよびカラーの特徴量は本手法での同名の特徴量と対応付けた。また、特徴量間の距離 (非類似度) の計算には L2 ノルム (ユークリッド距離) を用いた。

ここでは、双方とも学習用データを用いない場合の性能比較として、既存手法では距離尺度学習で求める重み値を使用せずに特徴量間の距離を計算し、提案手法ではパワー正規化の値を全特徴量に対してデフォルト値 $p = 0.5$ に設定してユークリッド距離を計算した。ただし、MTurk のデータセットに関しては複数被験者の多数決に従った値 (文献 [3] 中での Majority) での精度と比較している。この結果が示すように、学習を用いない条件では、提案手法はエッジ特徴量に関して既存手法とほぼ同等の精度が得られている。一方、テクスチャ特徴量では提案手法が若干高い精度を示しており、カラー特徴量と全特徴量に関しては既存手法の方が若干高い精度を示している。

次に、両手法で同じ学習用サンプルを用いた条件下での検索精度の比較結果を表 2 に示す。ただし、既存手法では各次元に対する重み値が、提案手法ではパワー正規化の乗乗値 (エッジ、テクスチャ、カラーの順に各々 $p = 0.35, 0.6, 0.15$ とカラー特徴量を統合する際の重み ($w = 0.35$) が、学習用データに対して最高の正解率になる値として最適化されている。

この結果が示すように、各種特徴量に関しては学習ありとなしの場合で両手法の性能差に同様な傾向が認められる。しかしながら、全特徴量に関しては MS データセットでは同等の精度が得られ、MTurk データセットでは提案手法の精度が若干低くなっている。MTurk データセットはスタイルの類似度が被験者間の判断のみで決定されており、三つ組の画像のスタイルがすべて異なっている場合 (あるいは、すべて類似している場合) の個人ごとの判断の揺らぎ

*2 http://webdiis.unizar.es/~elenag/projects/SIG2014_styleSim/

*3 <http://www.novadevelopment.com>

が含まれると考えられる。さらに、全試験用サンプルに占める割合が6%程度と少ないので、実験結果の信憑性に関しては相対的に低いものと考えられる。

既存手法のエッジに関連する特徴量は、2種類のヒストグラム（一定範囲内のカラー勾配値と黒色に関するもの）と SWT (Stroke Width Transform [20]) を用いているのに対して、提案手法はエッジ抽出された画像の対数輝度ヒストグラムのみから計算されている。また、既存手法は2種類のテクスチャ特徴量 (LBP と Haralick [21]) を用いているのに対して、提案手法では LBP のみを用いている。さらに、カラー特徴量に関しても既存手法が Colorfulness [22] や絵画の品質評価 [8] に用いられた高レベルな特徴量を用いているのに対して、提案手法は HSV 色空間でのヒストグラムという低レベルな特徴量のみを用いている。

以上の比較により、本提案手法の方が既存手法よりも実装が容易であり計算量も少ない。特に、本手法では画素ごとに定義される値のヒストグラムのみで全特徴量が算出されるので、積分画像に基づく計算法 [23] を導入することにより、任意の局所的な矩形領域内でのスタイル特徴量を重複なく効率的に計算できる。すなわち、本手法のスタイル特徴量は三つ組比較の実験において既存手法と同等の精度を得ることができ、スタイル特徴の局所的な計算にも適していると考えられる。

図 5 にパワー正規化の冪乗値に対する学習用データの三つ組比較の正解率の変化を示す。この結果より、冪乗値の一定の範囲内では正解率の変動も小さいことが確認できる。したがって、これらの値はデータセットごとに学習する必要はなく、他のデータセットに対しても有効に働くことが予想される。その実験的な検証に関しては 4.3 節で述べる。

図 5 の丸印の曲線は、カラー特徴の重み値の変化に対する学習用データでの正解率を表す。この値の変化に対する正解率の変動は微小であるが、次節で詳述するように、性

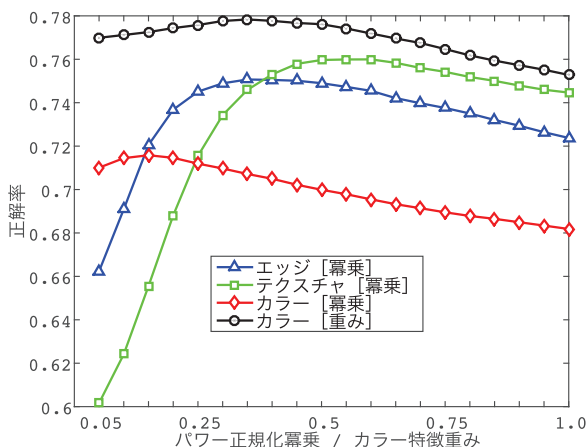


図 5 パワー正規化冪乗とカラー特徴重みに対するスタイル比較精度
Fig. 5 Style comparison accuracy for various conditions of power normalization and color feature.

能評価値には大きな影響を及ぼした。

4.2 検索精度の指標に基づく特徴量の性能評価

特徴量の性能評価として、前節の MS データセットから抜粋した 1 枚のクエリ画像に対して MTurk データセットもあわせた全画像群との特徴量間の距離（非類似度）を算出し、その値を昇順に並べた順位をランキングとする。ここで、MS データセットでの同じスタイル分類の画像を正しい検索結果として性能を評価する。

本実験では、検索の精度を評価するのに各クエリ画像に対する平均適合率を平均した以下の式で与えられる mAP 値 (Mean Average Precision) を用いた。

$$mAP \stackrel{\text{def}}{=} \frac{1}{Q} \sum_{q=1}^Q \left(\frac{1}{S_q} \sum_{i=1}^{S_q} \frac{Rank_C(i)}{Rank_A(i)} \right),$$

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 $Rank_C(i)$, $Rank_A(i)$ は各々、同じスタイル分類での i 番目の画像のクラス内および全画像中でのランキング順位（すなわち、距離の短い順番）を表す。ただし、本実験では MS データセットの全画像をクエリとして用いたので $Q = 3,591$ である。

表 3 に示すように、テクスチャとカラーの特徴量に関しては、本提案手法の方が学習の前後において既存手法よりも高い性能を示している。また、エッジ特徴量に関しては性能差は認められないものの、全特徴量を用いた検索結果においては、提案手法は学習後において既存手法よりも高い性能を示している。ただし、学習後の性能は 4.1 節の実験と同様の方法と学習用サンプルで求めた冪乗値と重み値を用いた。

各特徴量での学習前後での検索精度に大差が生じていないのは、パワー正規化の冪乗値がデフォルト値 ($p = 0.5$) でも十分に機能することを示唆している。一方、すべての特徴量を統合した場合には学習後に性能が大幅に向上しているので、本提案手法においては幾何学的（エッジとテクスチャ）特徴に対する色特徴の重み w の調整が重要であると考えられる。これは、イラストの種類によって配色をスタイル特徴とするものとそうでないもの（たとえば、多様な色を使用した絵画調の作品と白黒成分のみの線画）が存在し、色特徴への重み付けがこれら 2 種類のスタイル特徴に対する分類基準の調整に大きく影響するためと考えられる。

表 3 mAP 値で与えられる検索精度 (学習前/学習後)

Table 3 Retrieval accuracy with mAP (before/after) learning.

比較手法	エッジ	テクスチャ	色	全特徴量
既存手法	0.21/0.19	0.22/0.24	0.17/0.17	0.36/0.38
提案手法	0.22/0.22	0.30/0.31	0.24/0.25	0.32/0.43

4.3 他のデータセットを用いた性能評価

本節では、前節の実験とは異なるデータセットを用いた性能評価の結果を述べる。

クリップアート用の投稿サイトである OpenClipart のサイトから無作為に収集したデータセット 1,000 枚、および市販の素材画像データ集 (DEX 具満タン*4) から生物や人物に関するイラストを収集したデータセット 1,000 枚を MS データセットと組み合わせて試験用サンプルとした。ただし、追加したデータセットの画像に対しても同様に、透明化された背景の白色化と画像サイズを統一化するためのスケージング変換を施した。前節の実験と同様に 4.1 節で求めたパワー正規化の冪乗値とカラー特徴の重み値を用い、MS データセットの全画像をクエリ画像とした場合の検索精度を計算した結果、新たな 2 種類のデータセットと組み合わせた場合の mAP 値は双方共に 0.44 となった。さらに、前節の実験で用いた試験用サンプルにこの 2 種類のデータセットの片方のみ、および両方を加えた場合で計算した mAP 値はすべて 0.43 となった。以上の結果より、異なるデータセットとの組合せや追加に対しても一定の mAP 値が得られているので、本手法の検索精度のデータセットの変化に対する頑健性が示唆された。

5. 次元削減に基づくスタイル識別子の生成

画像アーカイブ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の機能を実装する場合には、各画像のスタイル属性を示す識別子をファイルの保存場所と関連付けたテーブルデータとして保存する必要がある。しかしながら、大規模なデータセットに対しても即応的な検索を保証するためには、特徴量の次元数を削減して登録するのが望ましい。

ここでは、スタイル特徴量の次元数と容量を削減するために、高性能なインデックス生成手法である繰返し量子化法 (Iterative Quantization [24]。以後、ITQ と呼ぶ) を導入する。ITQ は、主成分分析で得られる低次元空間での各軸成分の値の分散量の違いから生じる不効率な量子化を改善するために、その分散量を均等化する回転行列を繰返し計算で求める手法である。原手法は 2 値への量子化を提案しているが、本手法での 256 次元の特徴量をそのまま 2 値化した識別子では、情報量不足からその次元数にかかわらず検索精度が大きく低下してしまった。ゆえに、この ITQ で得られた低次元空間で多段階に量子化した特徴量を識別子とする。

ここでは、量子化の段数を β ビットで表現できる整数値 $z_i^\beta \in [0, 1, \dots, 2^\beta - 1]$ として、以下の式で量子化する。

$$z_i^\beta = \text{round}^\beta \left(\frac{2^\beta (\bar{f}_i - \min_F)}{\max_F - \min_F} \right),$$

ただし、 \min_F と \max_F は学習用サンプルから得られた全特徴量および全次元での最小および最大の値を示す。また、 $\text{round}^\beta(x)$ は最も近い整数 z_i^β への丸め関数であり、 $x < 0$ 、 $x \geq 2^\beta$ の場合には各々 $x = 0$ 、 $x = 2^\beta - 1$ で値を打ち切る。

図 6 に、次元数削減と量子化で生成されたスタイル識別子の容量 (総バイト数) と、その識別子を用いて 4.2 節での実験と同条件で画像検索した場合の mAP 値を示す。ただし、本手法と既存手法の次元削減する前の特徴量に対する mAP 値も比較用に 2 本の横線で示している。

主成分分析による次元削減と回転行列の計算には学習用の全サンプルを用いなくても精度は確保できるので、今回の実験では学習用画像の 1/10 の特徴量を無作為に抽出して計算に用いた。また、回転行列を最適化する繰返し計算の回数はすべて 100 回とした。

図 6 の結果より、各次元の数は 5 ビット以上を確保すれば高い検索性能が安定して得られることが確認できる。また、スタイル識別子に必要な総バイト数は 50 バイト程度 (5 ビット 80 次元, 6 ビット 64 次元, または 8 ビット 48 次元が候補となる) を確保すれば元の 256 次元データに対する精度 (mAP 値) に達していることが確認できる。

次に、このスタイル識別子を 8 ビット 48 次元で生成し、MS データセットの各スタイル分類の画像群のみをクエリとして用いたスタイルごとの mAP 値を、4.2 節での実験と同様に求めた。

図 7 に、mAP 値の最も高い上位の 3 スタイルに対して、その中でも平均適合率が最高となったクエリ (左端) を用いて上位 3 位までに検索された画像と、異なるスタイルの画像で最上位に検索された画像 (右端) を示す。ただし、この 3 スタイルの mAP 値は 1.0, 1.0, 0.9976 であり、同じスタイル分類内の全画像が検索上位を占める理想的な検索結果が得られている。

図 8 に、mAP 値の最も低い下位の 3 スタイルに対して、

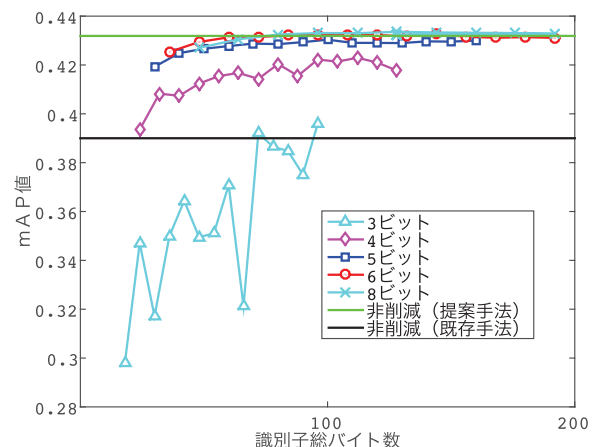


図 6 識別子の容量に対する検索精度の変化

Fig. 6 Change of retrieval accuracy for the capacity of a descriptor.

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



図 7 平均適合率が最高のクエリの検索結果例. 左端はクエリ画像, 右端は最上位の異スタイル画像

Fig. 7 Examples of retrievals with the query of the highest AP.

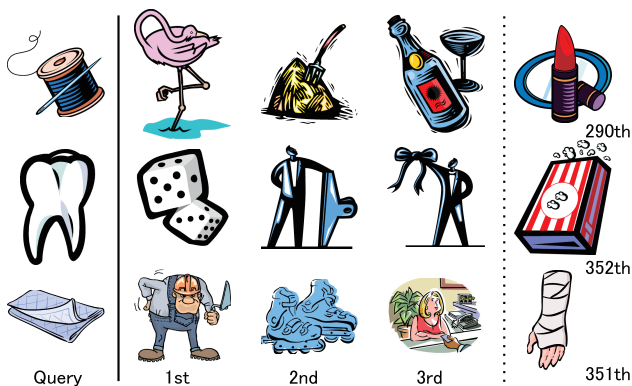


図 8 平均適合率が最低のクエリの検索結果例. 左端はクエリ画像, 右端は最上位の同スタイル画像

Fig. 8 Examples of retrievals with the query of the lowest AP.

その中でも平均適合率が最低となったクエリ (左端) を用いて上位 3 位までに検索された画像と、同じスタイルの画像で最上位に検索された画像とその順位 (右端) を示す。ここで、スタイルごとに求めた mAP 値は上段から 0.0097, 0.0246, 0.0259 であった。この例では上位 3 位までの検索結果はすべて異なるスタイルとして分類された画像であるが、そのスタイルには若干の類似点も認められる。たとえば中段 1 位の画像はその配色と輪郭線のスタイルが同スタイル分類の最上位の画像よりも似ていると判断できる。

図 8 の結果は、作画スタイルによる分類の本質的な曖昧性や、客観的な基準設定の困難性を示唆している。今回の実験で真値として用いた、事前にスタイルで分類された MS データセットの中には、作画スタイルの一貫性に欠ける分類やスタイルの類似した複数の分類が存在する。したがって、性能評価の精度や客観性を向上させるためには、今後はベンチマーク用のデータセットの内容の精査も必要であろう。

本手法で提案したスタイル特徴量は、検索目的以外にもスタイルの分布状態を可視化する際に用いられる。たとえば、各画像の特徴量を次元圧縮法 [25] で 2 次元座標



図 9 肖像のイラストのスタイルマップ生成例
Fig. 9 Example of style map for portrait images.

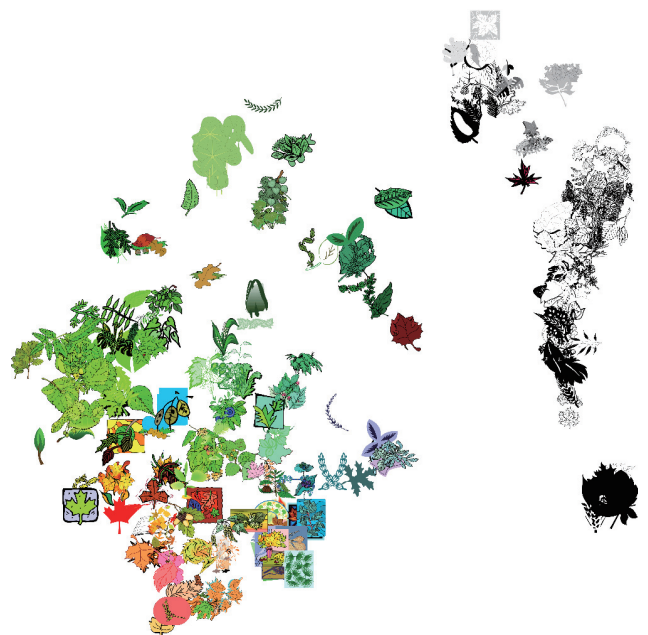


図 10 葉のイラストのスタイルマップ生成例
Fig. 10 Example of style map for leaves' images.

に変換し、その位置に画像を表示することにより、図 9 (OpenClipart から抜粋した 96 枚の肖像のイラスト) や図 10 (ArtExplosion から抜粋した 262 枚の葉のイラスト) に示すように、意味での分類が同じ画像群をスタイルの類似度に基づく距離関係で 2 次元平面上に配置したマップ表示が可能となる。

6. おわりに

本論文ではイラストの作画スタイルの類似度を推定できる識別子の生成方法を提案した。スタイルに特化した識別

子をタグデータやメタデータの一部として画像データと共に保存すれば、イラストの作画スタイルに基づく検索機能以外にも、スタイルに基づく各種の編集機能に役立てられる。

本研究での技術的貢献は以下に要約される：

- エッジ画像, LBP, HSV カラーのヒストグラムのみで、作画スタイルの特徴量を効率的に構成した。
- パワー正規化と色特徴の相対的な重みを調整することにより、平均適合率を用いた検索精度において既存手法よりも高い性能が得られた。
- 繰り返し量子化法により、検索精度を維持しながら 256 次元の特徴量を 50 バイト以下の識別子に変換できた。

本手法の識別子の次元数を固定して量子化のビット数を増加させた場合、8 ビットを超えると精度の向上は認められなかったので最も低い次元数は 8 ビットを用いた量子化での 48 次元となる。これは既存手法の特徴量が距離学習によって削減される 78 次元と比較すると、その次元数の少なさにおいても本手法に優位性が認められる。

他のデータセットを用いた性能実験では、2 種類の変数(パワー正規化の冪乗と色特徴に対する重み)の値を再学習することなく同等の検索精度が得られたので、データセットの更新や追加に対しても、最適化による変数の再調整は必要がないと考えられる。

すべての特徴量は画素ごとに算出される値のヒストグラムにより構成されるので、ヒストグラムの積分画像 [23] により任意の矩形領域での局所特徴量も効率良く計算できる。この性質は、局所特徴集合からビジュアルな単語辞書を構築する Bag-of-Visual-Words の方法にも適しており、単語のヒストグラムやその Fisher Vector を用いた特徴量 [26] を導入すれば、検索精度のさらなる向上が期待できる。

作画スタイルのとらえ方は主観に左右されるので、3 種類の特徴量の影響度を調整しながら対話的に候補を絞り込む機構も実用的には必要となるであろう。特に、本手法では色特徴に対する重み値が検索性能に大きな影響を及ぼしているため、スタイルのとらえ方の分類間での違いを直観的に指定する仕組みが必要となる。その一手段として、検索時の条件指定の操作にクエリ画像のスタイル特徴を要約した図形表示を用いること等が考えられ、これらの機能拡張は今後の課題である。

謝辞 本研究は、人工知能研究振興財団の人工知能研究助成「作画様式を識別する認知的距離空間の学習方法の開発」、および日本学術振興会の基盤研究 (C) 15K00151「作画スタイルに基づくイラスト画像の検索と編集」の支援を受けた。

参考文献

[1] 栗山 繁：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, 情報処理学会グラフィクスと CAD 研究会

- 報告, No.3, pp.1–7 (2013).
- [2] 栗山 繁：クリップアート画像のスタイル分類特徴量—スタイルに基づくランキング, *Visual Computing/グラフィクスと CAD 合同シンポジウム予稿集*, Article No.10 (2014).
- [3] Garces, E., Agarwala, A., Gutierrez, D. and Hertzmann, A.: A Similarity Measure for Illustration Style, *ACM Trans. Graphics (SIGGRAPH 2014)*, Vol.33, No.4, Article No.93 (2014).
- [4] Barroso, M.F., Fonseca, M.J., Barroso, B., Ribeiro, P. and Jorge, J.A.: Retrieving ClipArt Images by Content, *Proc. 3rd International Conference on Image and Video Retrieval (CIVR'04)*, LNCS (2004).
- [5] Martins, P., Jesus, R., Fonseca, M. and Correia, N.: Clip art retrieval combining raster and vector methods, *Content-Based Multimedia Indexing (CBMI)*, pp.35–40 (2013).
- [6] Zhu, Y., Tan, T. and Wang, Y.: Font recognition based on global texture analysis, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.23, No.10, pp.1192–1200 (2001).
- [7] Ding, X., Chen, L. and Wu, T.: Character Independent Font Recognition on a Single Chinese Character, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.29, No.2, pp.195–204 (2007).
- [8] Li, C. and Chen, T.: Aesthetic Visual Quality Assessment of Paintings, *IEEE Journal of Selected Topics in Signal Processing*, Vol.3, No.2, pp.236–252 (2009).
- [9] Xu, K., Li, H., Zhang, H., Cohen-Or, D., Xiong, Y. and Cheng, Z.-Q.: Style-content Separation by Anisotropic Part Scales, *ACM Trans. Graph.*, Vol.29, No.6, pp.184:1–184:10 (2010).
- [10] Freeman, W.T., Tenenbaum, J.B. and Pasztor, E.C.: Learning Style Translation for the Lines of a Drawing, *ACM Trans. Graph.*, Vol.22, No.1, pp.33–46 (2003).
- [11] Berger, I., Shamir, A., Mahler, M., Carter, E. and Hodgins, J.: Style and Abstraction in Portrait Sketching, *ACM Trans. Graph.*, Vol.32, No.4, pp.55:1–55:12 (2013).
- [12] O'Donovan, P., Agarwala, A. and Hertzmann, A.: Color Compatibility from Large Datasets, *ACM Trans. Graph.*, Vol.30, No.4, pp.63:1–63:12 (2011).
- [13] Lin, S., Ritchie, D., Fisher, M. and Hanrahan, P.: Probabilistic Color-by-numbers: Suggesting Pattern Colorizations Using Factor Graphs, *ACM Trans. Graph.*, Vol.32, No.4, pp.37:1–37:12 (2013).
- [14] Karayev, S., Hertzmann, A., Trentacoste, M., Han, H., Winnemoeller, H., Agarwala, A. and Darrell, T.: Recognizing Image Style, *BMVC2014* (2014).
- [15] Bar, Y., Levy, N. and Wolf, L.: Classification of Artistic Styles using Binarized Features Derived from a Deep Neural Network, *ECCV 2014 Workshop VISART* (2014).
- [16] Weiss, Y.: Deriving intrinsic images from image sequences, *Proc. 8th IEEE International Conference on Computer Vision (ICCV 2001)*, Vol.2, pp.68–75 (2001).
- [17] Satpathy, A., Jiang, X. and Eng, H.-L.: LBP-Based Edge-Texture Features for Object Recognition, *IEEE Trans. Image Processing*, Vol.23, No.5, pp.1953–1964 (2014).
- [18] Vedaldi, A. and Zisserman, A.: Efficient additive kernels via explicit feature maps, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2010)*, pp.3539–3546 (2010).
- [19] Perronnin, F., Liu, Y., Sanchez, J. and Poirier, H.:

- Large-scale image retrieval with compressed Fisher vectors, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2010)*, pp.3384–3391 (2010).
- [20] Epshtein, B., Ofek, E. and Wexler, Y.: Detecting text in natural scenes with stroke width transform, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2010)*, pp.2963–2970 (2010).
- [21] Haralick, R., Shanmugam, K. and Dinstein, I.: Textural Features for Image Classification, *IEEE Trans. Systems, Man and Cybernetics*, Vol.SMC-3, No.6, pp.610–621 (1973).
- [22] Hasler, D. and Suesstrunk, S.E.: Measuring colorfulness in natural images, *Proc. SPIE*, Vol.5007, pp.87–95 (2003).
- [23] Porikli, F.: Integral histogram: A fast way to extract histograms in Cartesian spaces, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2005)*, Vol.1, pp.829–836 (2005).
- [24] Gong, Y., Lazebnik, S., Gordo, A. and Perronnin, F.: Iterative Quantization: A Procrustean Approach to Learning Binary Codes for Large-Scale Image Retrieval, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.35, No.12, pp.2916–2929 (2013).
- [25] van der Maaten, L. and Hinton, G.E.: Visualizing High-Dimensional Data Using t-SNE, *Journal of Machine Learning Research*, Vol.9, pp.2579–2605 (2008).
- [26] Perronnin, F. and Dance, C.: Fisher Kernels on Visual Vocabularies for Image Categorization, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR '07)*, pp.1–8 (2007).

推薦文

本研究では、画風や画調が類似したクリップアートやイラストの検索を可能とすることを目的として、イラストスタイルの類似度の定量化手法を提案している。2014年8月のSIGGRAPHにおいて類似の研究発表があったが、シンポジウムでの発表は2014年6月であり、本研究のテーマの設定自体に高い新規性が認められるとともに、実用性も認められる。推薦論文では、数千枚のイラスト画像群を用いた実験を行い、既存研究よりも低容量の識別子でより高い検索精度が得られることを示している。上記の理由により、推薦論文として推薦するに十分に値する。

(グラフィックスとCAD研究会主査 柿本正憲)



栗山 繁 (正会員)

1962年生。1987年大阪大学大学院工学研究科電子工学専攻修了。工学博士。1988年日本IBM東京基礎研究所研究員。1994年広島市立大学情報科学部助教授。1998年豊橋技術科学大学情報工学系助教授。2005年同大学教授。2010年同大学大学院情報・知能工学専攻教授、現在に至る。2005～2009年産総研デジタルヒューマン研究センター・チーム長兼務。コンピュータ・グラフィックスおよび視覚的な情報メディアに関する研究に従事。電子情報通信学会、IEEE、ACM、照明学会各会員。情報科学技術フォーラム (FIT 2003) 船井 Best Paper 賞、本学会山下記念研究賞 (2014)、CBMI 2015 Best Paper 賞等受賞。