

印象情報の付与による類似画像検索性能の評価

小林 匠^{†a)} 森崎 巧一[†] 大津 展之[†]

Evaluation of Image Retrieval Performance by Using Subjective Impression

Takumi KOBAYASHI^{†a)}, Norikazu MORISAKI[†], and Nobuyuki OTSU[†]

あらまし 本論文では、類似画像検索に用いる人間の主観印象情報と検索性能との関係について述べる。一般に検索は画像情報を用いて行われているが、検索結果の評価は行為者の主観的な判断（印象評価）に依っている。従って、検索精度向上のために印象情報の検索システムへの導入が期待されるが、どの程度組み込むか（または不要か）が問題となる。そこで、印象情報と画像情報を異なる割合で扱ういくつかの多変量解析手法を画像検索に適用し、その結果を比較することで印象情報の寄与について論じる。ここでは数量化Ⅳ類などの多変量解析手法、さらにそれらのカーネル法を用いた拡張手法を用いた。家紋画像検索における実験の結果、検索性能向上への印象情報の寄与、さらに印象情報と画像情報をほぼ対等に扱う場合に最良の性能が得られることを確認した。

キーワード 類似画像検索、印象情報、数量化Ⅳ類、正準相関分析、カーネル法

1. ま え が き

近年、情報基盤の発達により人々の扱う情報は質・量ともに飛躍的に増大し、それらを効率的に処理するための技術が必要となっている。特にインターネットにおけるテキスト検索技術は新たな付加価値を創造する程に注目を集めている [1]。さらに、デジタルカメラなどの記録デバイスやデータを保存する記憶デバイスの高精度化に伴って大量の静止画や動画の中からユーザの希望に添うものを検索する技術も望まれている。

画像検索の従来方法としては、画像にその内容を示すキーワードなどの情報を付与し、キーワードを検索キーとして検索する方法（キーワード検索）がある。しかしながら、膨大な画像にキーワードを付与するには非常に大きな労力を必要とし、画像内容を表現するキーワードを的確に選択することも困難である。そのため、近年では画像自体を検索キーとして入力し、それに類似した画像を検索する方法（Content-Based Image Retrieval:CBIR）が盛んに研究されている [12]。CBIRにおいて、検索対象となる画像データベースは大きく分けて2種類存在する [12]。一つは、

比較的広範囲の画像種類（Broad Domain: BD）を扱う類似画像検索である。例えば、インターネットにおける画像検索などは車や家や人などの多種多様な画像が検索対象となるためにこの分類に属する。もう一つは、限定された種類の画像（Narrow Domain: ND）を扱う類似画像検索である。これはある特定の対象、例えばデザインなどに限定した画像検索である。

1.1 従 来 研 究

CBIRにおいては一般に、BDを対象とすることが多い。BDでは、画像内容の変動は非常に多岐に渡るが、検索結果としては入力キー画像と類似した画像群が求められることが多く、ここで問われる類似性は大まかなものとなる。つまり、対象の画像における物理的構造のみから類別可能となる場合が多い。包括的なサーベイは [12] を参照されたい。近年の手法としては、Video Google [14] に代表される bag-of-features によるものが多い。これは画像中の特徴的な領域の局所特徴をクラスタリングにより量子化したものを Visual Word と呼び、その Word 分布に基づいてその画像特徴を構成する手法である。また、CBIRに関連する Workshop もいくつか開催されている [16]。そこでは、high-level feature と呼ばれる対象のカテゴリ情報を抽出すること（アノテーション）が重要なタスクの一つと位置づけられている。BDにおいては対象のカテゴリ（ラベル）情報を重視した手法が多い。

[†] 産業技術総合研究所、茨城県

National Institute of Advanced Industrial Science and Technology, 1-1-1 Umezono, Tsukuba, Japan

a) E-mail: takumi.kobayashi@aist.go.jp

BD における検索は多種多様な画像群にユーザがアクセスできるという点で非常に効果的な手法である。しかしながら、検索の目的によってはこのような大まかな検索よりもさらに詳細な類似性が求められる場合もある。そのような場合には、前処理として検索画像のカテゴリ判別を行い、判別したカテゴリに対応する画像群の中で ND の CBIR を行うことが考えられる。または、ある対象、例えば商標検索などといった、始めから特定の ND が指定された検索も実応用では数多く存在する。そのため、ND 検索も BD と同様に重要となるが、そこでは検索対象が必ずしも物理的構造のみでは類別することのできないことも多く、対象の high-level feature のようなカテゴリ情報を抽出することは困難となる（抽出したとしても同一カテゴリとなる）。特にデザイン検索などでは、対象に明確なラベルを与えることは困難である。そのため、BD の手法を直接的に適用することは難しく、ND の手法では画像の物理的特徴のみならず検索行為者である人間の主観情報を取り入れることが多い。主観情報を導入した画像検索としては、判別分析を用いたもの [7]、正準相関分析を用いたもの [8]、多次元尺度構成法を用いたもの [18] などが挙げられる。

ND での従来研究では主観情報の収集方法が異なるなどの相違点もあるが、手法によってもそれらの扱い方が異なる。ここで問題となるのは、主観情報を検索システムにどの程度導入すべきか（または不要か）ということであるが、この問題がこれまでに明確な研究対象となったことはほとんどない。

1.2 目的

そこで本論文では、上記の問題への一つの解答を得るために、人間の主観的な印象情報と検索性能との関係を調べる。検索結果の類似性の評価は行為者の主観的な判断（印象）によるものであることから、教示情報として人間の印象情報を「適切に」導入することで、客観的には判断しづらい詳細な画像の類似性を再現することが可能となり、特にデザイン検索などにおける検索精度の向上が期待される。

特に本論文では、ND として扱われることの多い「デザイン」を対象とし、その一例として家紋画像検索を扱う。検索システムに反映される印象情報の割合によって検索精度がどのように変動するのか、画像検索における印象情報の寄与を評価することを目的とする。そのため、画像情報と印象情報とを異なる割合で扱ういくつかの多変量解析手法を適用し、その結果を



図 1 家紋の例
Fig. 1 Examples of family crests.

比較することで印象情報の類似画像検索への寄与を評価する。ここで得られた印象情報の寄与率は、今後の類似画像検索システムにおける一つの設計指針となることが期待される。

さらに本論文では、印象情報と画像の物理情報との間の非線形性に注目し、画像の物理的特徴として、単純な画像ベクトルだけでなく非線形な高次局所自己相関特徴 (HLAC) [11] も採用し、検索手法として各種多変量解析手法にカーネル法も導入して比較を行う。非線形な HLAC とカーネル多変量解析手法との組み合わせは従来研究では扱われていなかったものであり、これは手法面での本論文における提案手法となる。実験では従来法と提案手法との性能比較も含まれている。但し、本論文での主張は手法の提案ではなく、あくまで印象情報の寄与の解析である。

本論文の構成は以下の通りである。2 章で検索対象である家紋と扱う情報（印象情報と画像情報）について説明し、3 章では類似画像検索システムの概要と印象の寄与率が異なる多変量解析手法を説明する。4 章で実験を、むすびとして 5 章で結論を述べる。

2. 検索対象

2.1 家紋（画像）

本論文で対象となる家紋は、例えば図 1 のようなものが挙げられ [10] [19]、その大きさ・位置ともに正規化されているものとする。家紋画像は 2 値（白/黒）で表現され、その多くが白色（0）で地を表し、黒色（1）で家紋形状を表している。そのため家紋画像の物理情報を抽出することは比較的容易であり、色などの他要因がないため主観的な印象情報はこの客観的な物理情報と直接的に結びついている。本論文では各家紋に対する主観情報として印象情報、客観情報として画像情報を扱う。

2.2 印象情報

画像検索における印象情報としては、各画像間での比較による類似度 [18] や、画像群をグループ分けすることでのカテゴリ情報 [7] として収集することが多い。し

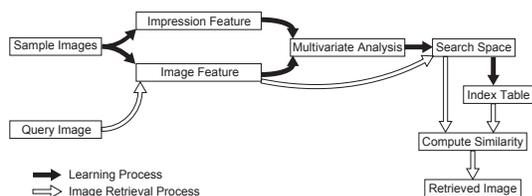


図 2 類似画像検索システムの概要

Fig. 2 Outline of the image retrieval system.

かし、2 対比較などにより各画像間の類似性を人間が評価するには大きな労力を必要とし、画像数が増えるとその情報収集は現実的ではなくなる。そこで、ここでは各々の画像に主観的な印象評価値が与えられている場合を考える（4.1 節参照）。これにより、検索システムに新たに画像を追加する場合でも、その画像に対する印象評価値を求めるだけで良い。

2.3 画像情報

客観的な画像情報として、画像の輝度値をそのまま画像ベクトルとして用いる場合と、高次局所自己相関特徴抽出（HLAC）[11] を行って特徴量を用いる場合とを考える。HLAC 特徴は対象に対する位置不変性や加法性などの性質を有し、特に対象が二値表現の場合には幾何学的特徴（輪郭線の傾きや曲率など）を抽出することができる。

3. 類似画像検索

類似画像検索システムにおける学習と検索の流れは以下ようになる（図 2）。

学習においては、まず学習サンプルから前述の画像情報と印象情報を抽出する（2 章）。次に、各家紋について得られた両者の情報を多変量解析手法によって統合し、両者の情報を反映した検索空間を構築、つまりその検索空間への画像情報（特徴ベクトル）からの写像を獲得する。ここで、多変量解析手法の種類によって両者の情報の統合の割合が異なることに注意されたい。最後に、学習サンプルの検索空間における配置をデータベースとして保持しておく。

検索の場合には、まず入力キーとなる家紋画像から画像特徴を抽出し、検索空間へ射影する。次に、蓄積されたデータベース（学習サンプル）と入力キーの家紋との間で検索空間内での近さを算出し、最終的には入力キーに近い順にデータベースの家紋画像を検索結果としてユーザに提示する。

以下ではここで用いるいくつかの多変量解析手法に

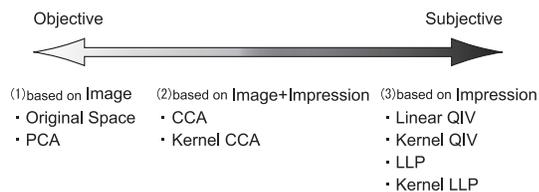


図 3 多変量解析手法の位置づけ

Fig. 3 Overview of the methods of multivariate analysis.

ついて説明する。

3.1 多変量解析手法による検索空間の構築

本論文で適用する多変量解析手法において、統合する画像情報と印象情報の割合は図 3 のようになる。ここでは、画像情報と印象情報との間に非線形関係がある場合を想定し、カーネル法を用いて各種線形手法を非線形手法へと拡張したのも同時に適用する。前述の HLAC 特徴とカーネル法による非線形手法との組み合わせが本論文の提案手法である。以下では画像情報、印象情報をそれぞれベクトル x, y で表す。

(1) 画像情報に基づく検索空間

印象情報を全く用いず、画像情報のみを用いて検索空間を構築する。ここでは画像ベクトル空間をそのまま用いるもの（Original）や、Eigenfaces [17] などと同様に主成分分析を用いて次元を圧縮した低次元空間を検索空間とする手法（PCA）を用いる。後者の PCA 手法では、画像情報における本質的な変動を表現し、ノイズなどの変動を除去するため、未知の入力キー画像に対する汎化性の向上が期待される。

(2) 画像情報と印象情報に基づく検索空間

画像情報と印象情報の両者を対等に統合するために正準相関分析（CCA）を用いる [8] [9]。正準相関分析は画像情報と印象情報との間で次式の相関係数が最大となるような空間表現（写像）を求める手法である。

$$r = \frac{\sum_i \mathbf{a}^T (\mathbf{x}_i - \bar{\mathbf{x}}) (\mathbf{y}_i - \bar{\mathbf{y}})^T \mathbf{b}}{\sqrt{\sum_i \|\mathbf{a}^T (\mathbf{x}_i - \bar{\mathbf{x}})\|^2} \sqrt{\sum_i \|\mathbf{b}^T (\mathbf{y}_i - \bar{\mathbf{y}})\|^2}}$$

ここでは画像空間 (x) のみならず印象空間 (y) も同時に考慮しており、両者は相関係数において対称であることから、CCA はこれらの情報を対等に扱う手法と言える。CCA では画像情報 x の写像 \mathbf{a} と印象情報 y の写像 \mathbf{b} が得られるが、本論文での画像検索では入力キーを画像として与えるため写像 \mathbf{a} のみを用いる。両者を用いることで双方向検索を実現することも可能である [8] [9]。

相関係数は両情報間の線形関係を抽出する評価基準であるため、画像と印象の間に非線形関係がある場合には、その関係性を正確に反映したもとはならない。そこで、カーネル法を用いた非線形正準相関分析 (Kernel CCA: KCCA) [5] も適用する。KCCA は次の固有値問題に帰着する。

$$(JK_x + \kappa I)^{-1} JK_y (JK_y + \kappa I)^{-1} JK_x a = \lambda a$$

ここで κ は正則化パラメータ (本論文では $\kappa = 1$)、 K_x, K_y は対称行列であり、カーネル関数 $k(x_i, x_j)$ を用いると $\{K_x\}_{ij} = k(x_i, x_j)$ となる。検索空間への射影は $\phi(x) = \sum_i a_i k(x_i, x)$ のように表現される。CCA の場合と同様に、ここでも印象情報からの (非線形) 写像も得られるが、本論文では画像情報からの写像のみを用いる。

(3) 印象情報に基づく検索空間

本論文における類似画像検索の入力キーは画像のみであり、印象情報は得られないために印象情報のみを用いた手法、例えば印象ベクトル y の PCA などとは適用できない。そこで、印象空間のメトリックをできる限り反映する空間を画像空間からの射影により近似する。画像からの射影として表現することから、必然的に検索空間には画像情報も含まれることになる。しかし、ここではあくまで印象空間のメトリックを近似するという目的であるため、ここで得られる検索空間には印象情報が大きく考慮されることになる。上記の目的を満たす手法は多次元尺度構成法 [18] など挙げられるが、本論文ではより柔軟な枠組みである数量化 IV 類 [2] を用いて実現する。

まず、各家紋間で類似度 s_{ij} を算出する。本論文での類似度は 2 対比較によるものではなく、各家紋の印象情報から算出される (4.2 節参照)。ここでの類似度とは、 i 番目と j 番目の家紋が (主観的に) 似ていれば s_{ij} は値が大きく、異なっていれば値は小さくなる。数量化 IV 類とはこの類似度に従って家紋サンプルの空間配置を構築する多変量解析手法である。つまり、類似度の大きい (似ている) 家紋対が出来る限り近い位置に配置されるような空間を構築する。得られる空間における座標値を \hat{x} として表現すると、次のように定式化される。

$$\min \sum_i s_{ij} \|\hat{x}_i - \hat{x}_j\|^2, \quad s.t. \sum_i \|\hat{x}_i\|^2 = 1$$

$$\Rightarrow (D - S)\hat{x} = \lambda \hat{x}$$

ここで、対称行列 $\{S\}_{ij} = s_{ij}$ 、対角行列 $\{D\}_{ii} (= D_i) = \sum_j s_{ij}$ 。得られた空間 (\hat{x}) は印象情報から測った類似度を最もよく反映する空間であるために、印象空間のメトリックを保持した検索空間となる。

次に、空間 \hat{x} を“画像情報からの射影”により近似する。 $\hat{x}_i = a^T x_i$ のような数量化 IV 類の線形近似 (Linear QIV: LQIV) は次式の一般化固有値問題に帰着する。

$$X(D - S)X^T a = \lambda X X^T a$$

さらに、カーネル法を用いた数量化 IV 類の非線形近似 (Kernel QIV: KQIV) も KCCA の場合と同様に定式化される。非線形近似では $\hat{x}_i = \sum_j a_j k(x_j, x_i)$ のように表現され、

$$K_x(D - S)K_x^T a = \lambda K_x K_x^T a$$

となる。

また、数量化 IV 類の変形として Normalized Cut [13] がクラスタリング手法として提案されている。

$$\min \sum_i s_{ij} \|\hat{x}_i - \hat{x}_j\|^2, \quad s.t. \sum_i D_i \|\hat{x}_i\|^2 = 1$$

この線形・非線形 (カーネル) 近似も Locality Preserving Projections (LPP), Kernel LPP (KLPP) として提案されている [4]。

以上のように印象情報から測った類似度を用い、印象情報のメトリックを反映させる手法として数量化 IV 類の各種近似手法を適用する。一方で、数量化 IV 類は類似度の定義を変えることで様々なメトリックを反映できる統一的な手法でもある。つまり、類似度を画像情報から、または画像情報と印象情報の双方から測ることで、前述の (1)(2) の手法を実現することも可能となる。これについては 4.5 節にて述べる。

4. 実験

前章で説明した多変量解析手法を用いて 類似家紋画像検索を行う。ここでは 2 通りの実験を行う。実験 1 では印象情報と画像情報の統合の割合が異なる様々な手法を適用する。それらの検索性能を評価し比較することで、類似画像検索への印象情報の寄与を調べることができる。また同時に、画像特徴量や多変量解析手法において非線形性を導入し、これと従来手法との比較も行っている。実験 2 では、実験 1 で得られた結果を用い、類似度の定義を変えた数量化 IV 類 (Kernel QIV) を適用してより詳細な印象情報の寄与を調べる。

表 1 印象語 (26 語)
Table 1 Impression words (26 words).

| | | | |
|--------|-------|--------|--------|
| あてやかな | 上品な | エレガントな | 飾り気のない |
| しゃれた | 清らかな | どっしりした | 洗練された |
| ユーモラスな | 繊細な | 格調のある | 鮮やかな |
| 革新的な | 装飾的な | 簡素な | 伝統的な |
| 気品のある | 風格のある | 古風な | 文化的な |
| 豪華な | 優雅な | 渋い | 力強い |
| 重厚な | 緻密な | | |

4.1 印象情報の収集

まず、予備調査として家紋の印象に適した印象語を厳選した。日本の家紋データ集 [15] から無作為抽出した 12 個の家紋画像に対し、参考文献 [6] [10] [19] 等から選出した約 180 語の印象語 (形容詞) を用いてアンケート調査を行った。このアンケート調査では、提示した家紋に対してどのような印象語が家紋の印象として適切かを二者択一で回答させた。被験者 10 名に評価させた結果、家紋の印象評価で扱えると考えられる上位 26 語を厳選した (表 1)。次に求めた 26 語を用い、日本の家紋データ集 [15] から新たにランダムに取り出した 50 個の家紋画像に対して印象評価アンケートを行った。印象評価は 24 名の被験者に 5 段階の評定尺度によって回答させた。アンケートの結果、各家紋に対して 24 名分の印象情報が得られるが、ここでは各印象語に対して 24 名の中での中央値をその印象値とした。なお、印象情報は個人によって変化することが考えられるが、ここでは家紋に対する印象は個人によって大きく異なることはないかと仮定している。以上の調査により 50 個の各家紋に対する印象情報として 26 次元の印象特徴ベクトル (要素の値は 1~5) y が得られた。

4.2 実験設定

[画像特徴抽出] 家紋画像は解像度を 100×100 に正規化しており、その画像ベクトルは 10000 次元のベクトルとなる。また、HLAC 特徴量としては家紋画像に対して局所相関の空間変位距離を 5 段階 (1, 2, 4, 8, 16 ピクセル) に変えて特徴抽出を行い、それらを連結した $25 \times 5 = 125$ 次元ベクトルを用いた。HLAC 特徴ではグレースケール画像の解像度を変えてスケールピラミッド表現したものに対する有効性も示されているが [3]、前述の通り HLAC は二値画像に対して特に有効であるため、ここでは画像の二値性を保つために HLAC の局所相関変位幅を段階的に変えて特徴抽出を行った。

[カーネル関数] 本論文ではカーネル法を用いた多変量解析手法として KCCA, KQIV, KLPP を適用する。ここでは画像特徴におけるカーネル関数としてガウシアンカーネル $k(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / 2\sigma_x^2)$ を用いた。分散 σ_x^2 は x のサンプル分散を基準に様々な値に変え、実験結果はその内で最も良い結果を表示している。KCCA においては印象情報 y に対してもカーネル関数が適用されるが、そのカーネル関数は次に述べる類似度と同一である。

[類似度] K/LQIV, (K)LPP では家紋間の類似度が必要となり、本論文での類似度は印象情報から算出する。類似度は次式のように定義する。 $s_{ij} = \exp(-\|y_i - y_j\|^2 / 2\sigma_y^2)$ 。ここで分散 σ_y^2 は y のサンプル分散を示している。但し、実験 2 (4.5 節) ではこの類似度の定義を変えることに注意されたい。

[検索方法] 検索空間において、入力キーとユークリッド距離の近いデータベース中のサンプルをその近さに従って検索結果として出力する。また、検索空間の次元は固有値の累積寄与率 (本論文では 99.9%) に従って決定した。PCA, (K)CCA の場合には固有値は降順に並んでいるものとし、累積寄与率 $\eta_k = \sum_{i=1}^k \lambda_i / \sum_{i=1}^n \lambda_i$ となる。K/LQIV, (K)LPP では固有値を昇順に並べ、また固有値 $\lambda = 0$ の場合には全サンプルの座標値が等しいという無意味な空間配置が得られるため、累積寄与率 η_k は

$$\eta_k = \frac{\sum_{i=1 \wedge \lambda_i > 0}^k \lambda_{max} - \lambda_i}{\sum_{i=1 \wedge \lambda_i > 0}^n \lambda_{max} - \lambda_i}$$

とした。ここで n はサンプル数を表す。

4.3 評価方法

Web 検索の場合には、ユーザは表示される検索結果の内上位 4 つまでを注視し、それよりも下位の結果には注意を払わない傾向があるとの調査結果がある [1]。そこで本論文においても注目する検索結果を上位 4 個として性能評価を行った。

評価方法は 50 個の家紋サンプルに Leave-one-out 法を適用し、50 回試行での平均評価値とする。つまり、49 個の家紋を用いて検索空間を構築し、残り 1 個を入力キーとして検索しその検索結果から評価値を算出する。本論文では家紋間の主観的類似度を直接収集していないため、印象評価値が似ていれば家紋自体も (主観的に) 類似しているという仮定の下で以下のように評価値を算出する (図 4)。ここでは入力キー家紋に対する印象情報 (26 印象語による評価値) も得ら

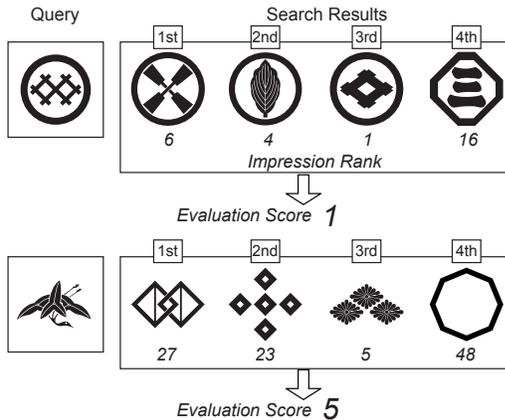


図4 画像検索の例とその評価値 (HLAC+KCCA)
Fig. 4 Example of the image retrieval result and its evaluation score using HLAC and KCCA.

れていることに注意されたい。

入力キーから類似画像検索により検索結果上位4個の家紋を出力する。また、同時に印象空間(26次元)において入力キー家紋の印象情報とユークリッド距離が近いものから順にデータベースの家紋を順位付けし(1~49位)、印象順位と呼ぶ。検索結果に対する評価値は、検索で得られた上位4個の家紋に対する印象順位の中で最も順位の高い家紋の印象順位とする(図4参照)。この評価値が小さい程、印象順位の高いものが検索結果に含まれていることを意味し、主観的な印象からも類似した画像が検索されていることを意味する。つまり、この評価基準の下では評価値が小さい程良好な結果となる。ユーザは一般に検索において最も似ている画像一枚を求めていることが多いため、ここでの評価対象は検索結果内の最高順位の家紋1つとした。

4.4 実験1:各手法による比較

50個の家紋を用いて各手法を上記の評価方法により比較した。その結果を図5に示す。

まずここでは、画像表現と多変量解析手法の組み合わせとして $\{ImageVector, HLAC\} \times \{Linear, Kernel\}$ の4通りを比較する。全体の傾向としては $[ImageVector \cdot Linear]$ の評価値が悪い。これは、輝度表現としての画像(画像ベクトル)と人間の主観的な印象が線形関係では捉えられず、非線形な関係にあることを示唆している。主観的な印象は対象の微妙な形状変化、例えば形状パーツの位置変化やその大きさの若干の変化に対してはほぼ不変であると考えられる。しかしながら画像ベクトル表現ではそれらの変化は大き

な変動となって現れ、全体として線形表現では表すことができなくなっている。一方で、HLAC特徴では位置不変性や加法性などから上述の変動に対してロバストになっており、HLACを用いた場合では線形手法でも良好な結果が得られている。また、非線形手法を用いた場合も同様に結果が改善されており、ここでも主観的な印象と画像とが非線形関係を成していることが示唆されている。非線形表現の中でも、非線形(カーネル)手法よりも非線形特徴(HLAC)を適用することによる改善が大きく、画像表現としての非線形性が重要であることもわかる。全体としては $[HLAC \cdot Kernel]$ の評価値が最も良い。

以上より、非線形性が導入されている $[ImageVector \cdot Kernel]$ 、 $[HLAC \cdot Linear]$ 、 $[HLAC \cdot Kernel]$ において印象情報の寄与について論ずる。ここでは、やはり印象情報を導入することによる検索性能の向上が確認される。しかしながら、印象情報を大きく考慮した場合、つまり(K)LPP、K/LQIVの手法よりも印象情報と画像情報を対等に用いた(K)CCAの手法による結果が良好であることがわかる。また、 $[HLAC \cdot Kernel]$ におけるKCCAが全手法の中で最良の評価値であった。評価が印象基準であることから、印象基準での類似性を用いた(K)LPP、K/LQIVの手法が一見すると最適であるように考えられるが、印象だけでなく画像情報も用いた(K)CCAが最良であったことに注意されたい。これは未知画像に対する汎化性によるものと考えられる。つまり(K)CCAでは画像情報、ここでは家紋画像としての制約を考慮することで、入力キーとなる未知画像に対する汎化性が増したと解釈できる。一方で(K)LPP、K/LQIVではほぼ印象情報のみを用いているために、画像としての制約をほとんど考慮せずに過学習を起こしてしまったとも言える。

本実験では、Leave-one-out法を用いているために評価値の変動はその平均値の変動に比べ若干大きくなっている(4程度)。しかしながら、異なる特徴空間である画像ベクトルとHLACのどちらの場合でも同様の傾向((K)CCAの結果が良好)が見られるため、本実験、並びに上述の議論は印象情報の働きに対してある程度の普遍性を有していると考えられる。

4.5 実験2:類似度による比較

ここでは、印象の寄与についてより詳細な評価を行う。つまり、先の実験では印象・画像情報を対等に扱った(K)CCAによる結果が最良であったが、統一的な手法である数量化IV類において印象情報と画像情報の

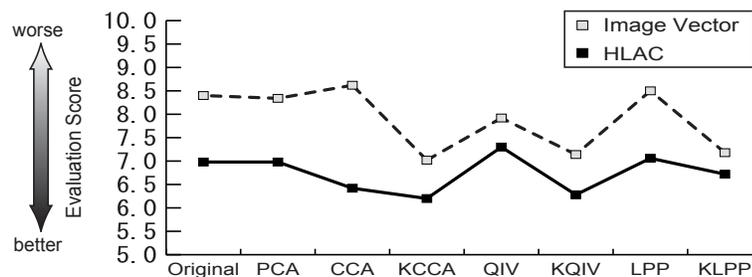


図 5 手法の比較実験の結果

Fig. 5 Experimental results of comparison of the methods.

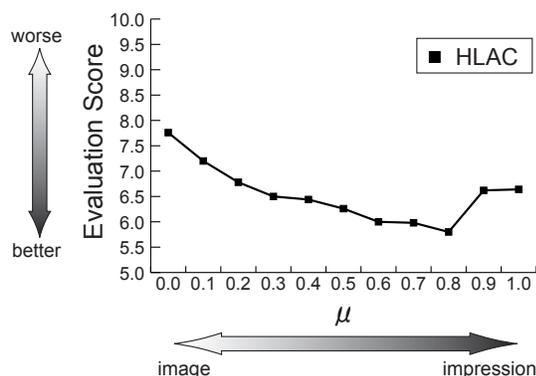


図 6 印象割合の比較実験の結果

Fig. 6 Experimental result of comparison of impression ratio in similarity.

統合の割合をより細かく変えてその評価を行う。他手法では統合の割合が固定であるためにここでは適用することができない。

実験 1 の結果により [HLAC・Kernel] における KQIV が良好な結果を出しているため、ここでの実験では画像特徴として HLAC を用い、手法として KQIV を適用する。実験 1 では、KQIV の類似度を印象情報のみから算出していたために KQIV をほぼ印象情報のみによる手法として扱ったが (4.2 節参照)、印象情報と画像情報を統合した類似度を用いることにより KQIV は両者の情報を含んだ手法となる。

印象情報と画像情報からの類似度を統合すると

$$s_{ij} = \mu \exp\left(-\frac{\|y_i - y_j\|^2}{2\sigma_y^2}\right) + (1 - \mu) \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma_x^2}\right)$$

となる。ここで μ は印象・画像情報の統合の割合を示す ($0 \leq \mu \leq 1$)。 $\mu = 1$ の場合は実験 1 の設定と同一

であり印象情報に大きく偏ったもの、 $\mu = 0$ の場合は類似度として画像情報のみを用いており印象情報を全く用いない手法となる。統合の割合 μ を様々に変えて 4.3 節で述べた評価方法を用いると、 μ と評価値との関係は図 6 のようになる。ここでも先の実験 1 (図 5 参照) と同様の傾向が見られる。まず、印象・画像情報のどちらか一方ではなく、両者を考慮した場合に評価値が最良となっている ($\mu = 0.8$)。このときの評価値は先の実験 1 での最良値 (KCCA において) よりも良い値が得られていることに注意されたい。さらに、画像のみの場合 ($\mu = 0$) よりも印象に大きく偏った場合 ($\mu = 1$) に評価値は良くなっている。

最良評価値は $\mu = 0.8 (> 0.5)$ のときに得られており、 $\mu = 1$ での評価値 $> \mu = 0$ での評価値であることから、ここでの評価値は印象情報寄りに良好となる傾向がある。KQIV は画像からの非線形射影で数量化 IV 類を近似したものであり、既にその写像表現の中に画像情報が若干ではあるが含まれている。そのため、その分の画像情報への偏り (バイアス) を考慮すると、印象寄り ($\mu = 0.8$) で最良の評価値となることは、全体として画像・印象情報を対等に扱っている場合に最良値が得られているとも解釈できる。

5. むすび

本論文では、類似画像検索における印象情報の寄与について、家紋画像検索を対象として調べた。ここでは、各家紋に対して印象情報 (印象評価値) が付与されたものとし、いくつかの多変量解析手法を適用して画像情報と印象情報を統合することで検索空間を構築した。各々の手法によって画像情報と印象情報の統合の割合が異なっており、それらの結果を比較することで印象情報の寄与を評価した。さらに、良好であった数量化 IV 類 (の非線形近似手法) において、画像情報

と印象情報を統合した類似度を用い、その統合割合を変えて結果を比較することでも寄与を調べた。実験結果により、以下のことが確認された。

- 印象情報の寄与により検索性能は向上する。
- 印象情報ばかりを用いるのではなく、印象情報と画像情報の双方を対等に考慮することで検索性能が向上する。

さらに、画像情報として画像ベクトルと非線形な HLAC 特徴ベクトルを、手法としても線形手法とカーネル法を用いた非線形手法を適用し比較を行った。その結果、HLAC や非線形手法を適用することで検索性能の向上が見られた。そのため、検索過程において非線形性を導入することが有効であり、特に非線形な画像特徴としての HLAC が効果的であることも示された。

今後の展望は、評価実験において実際のユーザによる評価を行うことや、印象情報と画像情報の統合のためのより効果的な多変量解析手法の提案が挙げられる。

謝 辞

本研究は経済産業省都市エリアプロジェクトの一環として行われたものである。また、御討論頂いた産業技術総合研究所脳神経情報部門栗田多喜夫副部門長に深く感謝をいたします。

文 献

- [1] AdvertisingAge, "Search Marketing Fact Pack 2006," <http://adage.com/images/random/searchfactpack2006.pdf>.
- [2] 圓川隆夫, "多変量のデータ解析," 朝倉書店, 1988.
- [3] F. Goudail, E. Lange, T. Iwamoto, K. Kyuma and N. Otsu, "Face recognition system using local auto-correlations and multiscale integration", IEEE Trans. Pattern Analysis and Machine Intelligence, vol.18, no.10, 1996.
- [4] X. He, P. Niyogi, "Locality Preserving Projections," Advances in Neural Information Processing Systems, 2003.
- [5] D. R. Hardon, S. Szedmark, J. Shawe-Taylor, "Canonical correlation analysis: an overview with application to learning methods," Technical Report CSD-TR-03-02, Department of Computer Science, Royal Holloway, University of London, 2003.
- [6] 本明寛, "造形心理学," 美術出版社, 1962.
- [7] 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.
- [8] 栗田多喜夫, 加藤俊一, 福田郁美, 坂倉あゆみ, "印象語による絵画データベースの検索," 情報処理学会論文誌, vol.33, no.11, pp.1373-1383, 1992.

- [9] N. Morisaki, T. Kobayashi, N. Murata, T. Kurita, N. Otsu, "Interactive Searching Method between User's Subjective Impressions and Family Crest Images," International Design Congress, 2005.
- [10] 沼田頼輔, "日本紋章学," 明治書院, 1926.
- [11] N. Otsu, T. Kurita, "A new scheme for practical flexible and intelligent vision systems," Proc. IAPR Workshop on Computer Vision, pp. 431-435, 1988.
- [12] A. W.M. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain, "Content-Based Image Retrieval at the End of the Early Years," IEEE Trans. Pattern Analysis and Machine Intelligence, vol.22, no.12, 2000.
- [13] J. Shi, J. Malik, "Normalized cuts and image segmentation," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, no. 8, 2000.
- [14] J. Sivic and A. Zisserman, "Video google: A text retrieval approach to object matching in videos," Proc. International Conference on Computer Vision, 2003.
- [15] System Product Co., LTD., "かもんかもん Ver.2," 2004.
- [16] TREC Video Retrieval Evaluation, <http://www.nipir.nist.gov/projects/t01v/>
- [17] M. Turk, A. Pentland, "Eigenfaces for recognition," Journal of Cognitive Neurosci, vol.3, no.1, pp.71-86, 1991.
- [18] 田邊勝義, 大谷淳, 石井健一郎, "多次元心理空間を用いた類似画像検索," 電子情報通信学会論文誌 D-II, vol.J75, no.11, pp.1856-1865, 1992.
- [19] 辻合喜代太郎, "日本の家紋," 保育社, 1974.

(平成 xx 年 xx 月 xx 日受付)

小林 匠 (正員)

平 15 東京大・工・機械情報工卒。平 17 同大学院修士課程了。同年(株)東芝入社。現在、産業技術総合研究所研究員。在中よりパターン認識・多変量解析の理論と応用の研究に従事。

森崎 巧一

平 16 筑波大・芸術学研究科博士課程修了。現在、産業技術総合研究所研究員。造形から受ける人の印象評価に関する研究に従事。芸術工学会、大学美術教育学会など各会員。

大津 展之

昭 44 東京大・工・計数工卒。昭 46 同大学院修士課程了。同年電子技術総合研究所入所。以来、パターン認識、画像処理、多変量データ解析、人工知能に関する数理的基礎研究に従事。工学博士。現在、産業技術総合研究所フェロー。筑波大学連係大学院教授、日本行動計量学会、情報処理学会など各会員。

Abstract In this paper, we investigate the contribution of subjective impression to the performance in image retrieval systems. Since the retrieved images are evaluated by user's subjective impression, the human impression is needed to be introduced into image retrieval systems in order to improve the performance. A problem is how much the impression information should be incorporated or not. We apply several methods, such as Quantification IV and Canonical Correlation Analysis together with kernel trick, which incorporate both impression and image information at various ratio. Experimental results show that the method equally combining both two information outperforms the others.

Key words Image retrieval, Subjective impression, Quantification IV, Canonical correlation analysis, Kernel trick