画像特徴量及び，tfidf に準じたキーワード評価による画像分類

宇津木 嵩行　　重松 利季　　田崎 幸彦　　加藤 俊一

中央大学理工学部 〒112-8551 東京都文京区春日 1-13-27

E-mail: {utsugi, shigematsu, tazaki, kato}@indsys.chuo-u.ac.jp

あらまし　近年，個人が趣味で製作した画像コンテンツのみならず，広告などに用いられる商業用の画像コンテンツまでもがネットワーク上でやりとりされるようになってきた．そのため，コンテンツ販売会社は，保有する大量の画像コンテンツから,顧客が望む画像を正確かつ迅速に検索できるシステムを試作・開発しつづけている．本研究では，教師データ群をそれぞれの群で自動的に細分化することで学習の効率化を計り,画像特徴量を構造的特徴と全域的特徴の二つの観点から設計し,専門家の感性をどの程度までコンピュータ上で再現できるのかを検証・考察する.また専門家の感性モデルを類似画像検索システムへ応用した．さらに，画像の特徴をキーワードを利用して表現した画像分類も試みた．写真の被写体や被写体の様子，撮影条件や季節などの画像内の特徴をテキスト化し，キーワードの出現頻度を tfidf に準じた重み付けにより評価し，印象とキーワードを結びつけた．

キーワード　画像データベース，自動分類，テキストデータ処理，データマイニング，感性モデル

1. はじめに

近年，個人が趣味で製作した画像コンテンツのみならず，広告などに用いられる商業用の画像コンテンツまでもがネットワーク上でやりとりされるようになってきた．そのため，コンテンツ販売会社は，保有する大量の画像コンテンツから，顧客が望む画像を正確かつ迅速に検索できるシステムを試作・開発しつづけている[2].

現在の検索ポータルサイトで提供されているような画像検索システムは，画像に付加された索引語とユーザーが入力したキーワードによるテキスト型検索を採用している．しかし，芸術性の特に高い広告用写真の場合，同じ被写体を扱っていたとしてもカメラマンの意図や撮影技法によって写真から受ける印象は大きく変化する．このような印象をキーワードで適切に記述することは非常に難しい．そのため，テキスト型検索システムでは，顧客の意図と検索結果から受ける印象との間にしばしばズレが生じる．キーワードの入力のみで手軽に利用できることは大きな利点であるが，以上のような問題が挙げられる．

ところで，我々は「ナチュラルな印象」,「フレッシュな印象」のように，イメージ語を用いて写真から受ける印象を評価することがある．筆者らはこのイメージ語による印象評価に着目し，写真の内容と写真から受ける印象との間の関係を数理的に記述することを目指している[2].

先行研究では，高次自己相関特徴を改良した 3 点間コントラストを画像特徴量とし，写真の専門家が付加したイメージ語で分類された画像を教師データ群として，プロの写真家の感性のモデル化を試みている[3]．しかし，3 点間コントラストをベースとした画像特徴量では，画像の局所的な部分まで特徴としてしまうので，画像全体の印象や構図が重要となる印象や雰囲気を表す特徴量としては不適切であると考えられる．

そこで，本研究では，教師データ群をそれぞれの群で自動的に細分化することで学習の効率化を計り，画像特徴量を構造的特徴と全域的特徴の二つの観点から設計し，専門家の感性をどの程度までコンピュータ上で再現できるのかを検証・考察する．また専門家の感性モデルを類似画像検索システムへ応用した．

さらに，画像の特徴をキーワードを利用して表現した画像分類も試みた．写真の被写体や被写体の様子，撮影条件や季節などの画像内の特徴をテキスト化し，キーワードの出現頻度を tfidf に準じた得点付けにより評価し，印象とキーワードを結びつけた．

2. 本研究のアプローチ

2.1. 画像特徴量による

先行研究では，人間の目の特徴抽出機構を模した 3 点間コントラストを提案しその有効性を検証してきた[3]．また，画像データベースの階層的分類と画像領域分割とを組み合わせ，人が対象を主観的に分類する際に，対象のどの部位のどういった特徴に着目しているのかを推定する手法を開発，内容型画像検索システムに応用してきた[3].

内容型画像検索システムとは，キーワードを廃し，画像そのものを検索キーとしてデータベースから類似画像を検索するシステムの事を指す[4]．画像には被写体の情報のみならず，カメラマンの意図や撮影技法など鑑賞者の印象を左右する様々な情報が内包されている．そのため，適切なモデルに基づいて構築された内容型画像検索システムには顧客の意図と検索結果から受ける印象との間にズレが生じにくいという利点がある．

しかしながら,内容型画像検索システムには以下のような問題がある．

1．複雑な構図や背景を持つ画像への対応　：　認識した

いオブジェクトだけが写っているような,簡単な構図や背景を持つ画像を検索することはできるが,複雑な構図や背景を持つ画像を検索することはできない.

２．検索キーとなる画像の用意 ： 画像をキーとする内容型検索では,全てのユーザーが検索キーとなる画像を用意しなくてはならない.

３．明確なイメージへの対応 ： 画像に対する明確なイメージを持っているのに,それを表す画像を持っていないユーザーに,画像をキーとした内容型検索では応えることができない.

このように,不特定多数の顧客を対象とするネットワーク上でのコンテンツ販売への応用を視野に入れた場合，内容型画像検索システムでは,ユーザーのニーズに応えることが出来ない.

そこで，本研究では印象を評価する際に用いられるイメージ語に着目し，画像中に含まれる鑑賞者の印象を左右する情報とイメージ語による評価との間の関係を,教師データを自動細分化した上で,マハラノビス距離を用いた判別分析によって学習・モデル化する手法を提案する．提案手法では,教師画像群をその印象に基づき（イメージ語でラベル付けされた）印象グループに分類する.これを親グループと呼ぶ.親グループをそれぞれ k 平均法でさらに細かく分類する.これを子グループとする.こうして細分化された各印象グループから画像の構造的特徴と全域的特徴を抽出する.構造的特徴には Gabor フィルタを,全域的特徴にはエントロピーとカラーヒストグラムを利用する.そうして得られた特徴量に,変数増加法による変数選択と異常値の除去を行う.そして,細分化された各印象グループごとに,マハラノビス距離を用いた判別分析を用いてモデルを構築する.このモデルを画像検索システムへ応用することで,ユーザーはイメージ語を指定するだけで,そのイメージを喚起するであろうという画像を検索することができるようになる.さらに,本研究では画像間の類似度を定義し,これを類似画像検索システムへ応用する.

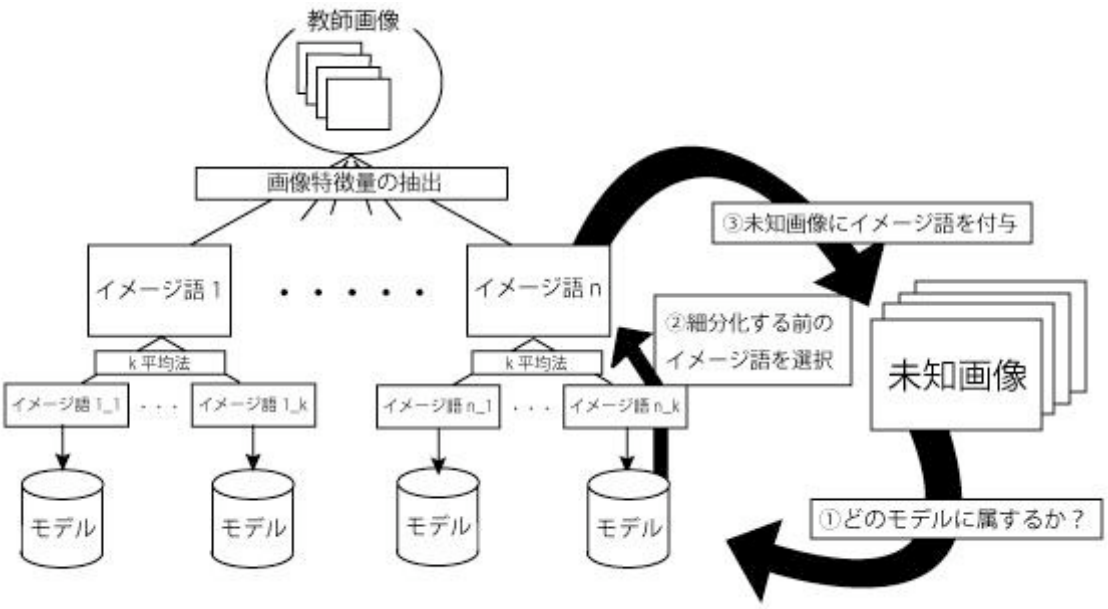


図 1. 本システムの流れ

2.2. キーワード評価による

先行研究では，Multiple Instance Learning (MIL)を用いて，画像自動分類システムが試作されており[13]，オブジェクトの認識に利用されつつある画像領域分割の手法が用いられている．しかし，画像領域分割の手法は，2,1 に挙げられた内容型検索システムの問題点である複雑な構図や背景を持つ画像への対応が未解決である.

そこで，本研究では，まず，特定の人間がオブジェクトや撮影手法などの画像の特徴を判定しキーワード化した.これは，画像に写るオブジェクトを正確に記述するだけではなく，オブジェクトの様子や雰囲気，カメラマンの撮影意図なども記述出来るため，広告など商業用に使われる芸術性の高い画像の特徴を記述出来ると考える.

そして，2.1 と同様に，印象を評価する際に用いられるイメージ語に着目し，画像中に含まれる鑑賞者の印象を左右するキーワード情報とイメージ語による評価との関係を，tfidf に準じたキーワード評価により得点付けする．tfidf とは，テキストマイニングの分野でよく利用されるアルゴリズムで,文章中の単語の出現頻度を利用して,X（x=1,2,･･,m）の文章間で相対的に特徴的に扱われる単語を抽出する手法である．これを画像に付与されたキーワー

ドに利用した．特定の印象グループのキーワードの出現頻度を利用して，y(y=1, 2, ･･, n)の印象間で相対的に特徴的に扱われる単語を抽出し，特徴的に扱われていればいるほど高くなるような得点付けをする．得点は，キーワードごとにyの印象グループに対して算出され，特定のキーワードがどの印象グループ内での得点が高いかを元に，画像に付与されたキーワードの得点の合計を求めることにより，画像がどのような印象に最も近いかを算出した．

3. 画像特徴量

人間の視覚の知覚過程では，局所的から全域的な明暗や色彩や彩度の特徴を抽出する神経回路が存在する．この神経回路によって抽出された特徴を統合することで人間はモノの形状やテクスチャを知覚している．そこで本研究では，画像特徴量を構造的特徴と全域的特徴の二つの観点から設計する．構想的特徴としては，ガボールフィルタを利用し，全域的特徴としてはカラーヒストグラムとエントロピーを利用する．人間の脳の初期視覚野ではガボール変換を行うことで特徴抽出を行っていることが知られている[5]．このことから我々は，ガボールフィルタから得られる特徴は人間が画像の印象を判断するのに適切な特徴だと考える．また，カラーヒストグラムやエントロピーは画像の大まかな特徴を表すのに適している．我々は，画像の印象は，局所的な部分よりもむしろ，画像のおおまかな色彩の分布によって大きく左右されると考えているため，全域的特徴としてこれらを利用した．本研究では，画像特徴を画像平面全体から抽出したものと，画像を 4×4 の領域に分割しそれぞれの領域から抽出したものを組み合わせ，これを特徴量ベクトルとする．

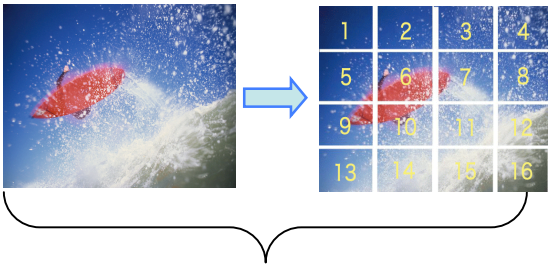
3.1. 構造的特徴

ガボールフィルタは，ガボール関数のパラメータを変化させることで，画像の特定の方位のエッジを抽出することができる．また，ガボールフィルタは人間の脳の初期視覚系のモデルとして知られている[5]．2 次元正弦関数型ガボール関数 G(x; y)は次のように定義される．

$$G(x,y) = Ke^{-\frac{1}{2}\left(\frac{(x-\mu x)^2}{\sigma^2 x}+\frac{(y-\mu y)^2}{\sigma^2 y}\right)} \times \sin(2\pi f_x \cos\theta + 2\pi f_y \sin\theta + \phi)$$

ただし，K は振幅，(μ_x, μ_y) はガボール関数の中心，σ_x と σ_y は標準偏差，f_x と f_y は周波数を表す．

本研究では，θ を 45°，90°，135°，180°の 4 パターン，周期を 10px，20px，40px の 3 パターンで設定し，全ての組み合わせから得られる一枚の画像全体の画素の平均値と，4×4 に分割した領域での平均値を特徴量からなる，計



構造的特徴	全域的特徴
Gabor filter	Color histogram Entropy

図 2. 画像特徴量の抽出

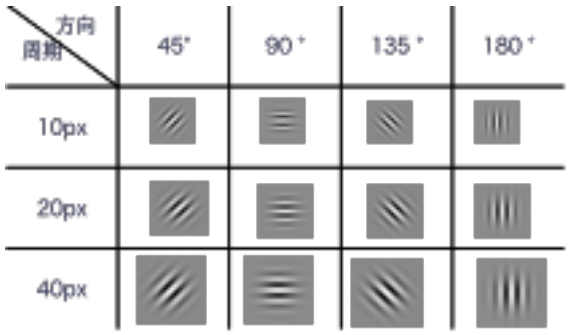


図 3. 用した Gabor filter のカーネル

204 の特徴量を構造的特徴量とする．

3.2. 全域的特徴

本研究では，画像の色空間を RGB 色空間から Lab 色空間に変換してから，以下の特徴抽出の処理を行っている．Lab 色空間の利点として，人間の視覚過程に近いと言われており，また RGB 色空間と違い，L 軸，a 軸，b 軸の全ての軸が独立していることから数学的にも扱いやすいということが挙げられる

3.2.1. カラーヒストグラム

画像特徴量として最も代表的なものがカラーヒストグラムである．これは画像全域にわたる色彩の分布をヒストグラム化したもので，画像の全域的な特徴を表現できるため，画像検索などでよく利用される[4]．しかしヒストグラムは画像中の色の頻度を表しているため，色の位置やテクスチャ等の局所的な特徴を得ることができない．そのため，カラーヒストグラムだけを画像特徴量として用いた場合，人間とは異なった評価をする場合も多い．

本研究では，Lab の 3 軸について，一枚の画像全体のカラーヒストグラムの平均値とエントロピー，4×4 に分割した領域でのカラーヒストグラムの平均値からなる，計 53 の特徴量を全域的特徴量とし，構造的特徴量と併せて 257 個の特徴を特徴ベクトル X と定義する．

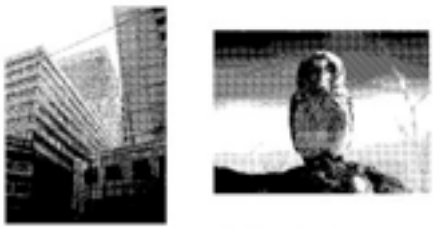
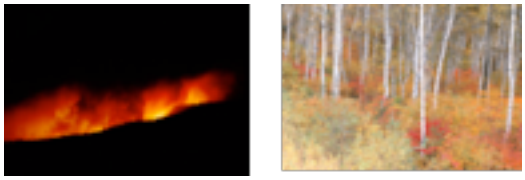


図 4. カラーヒストグラムが類似している画像例



H = 2.4　　H = 5.6

図 5. 色の頻度によるエントロピーの変化

3.2.2. エントロピー

エントロピーは以下の式で定義される.

$$H = -\sum_{1}^{n} P(i)\log_2 P(i), (i = 1,\ldots,256)$$

i は画像の各構成色, P(i)は画像中の各色の割合, H は情報エントロピーである. 本研究では, 画像の色数をベクトル量子化により 256 色まで減色した後にエントロピーを計算している. 画像が単色のみで構成されているような画像の H は最小値の 0 になり, 逆に 256 色が同じ頻度で現れる場合は最大値の 8 になる.

4. 感性のモデル化

4.1. Wilks の Λ 統計量による変数選択

本研究で用いる特徴量の次元は 257 次元である. 特徴量の次元数が増えるにつれ, 標準誤差が増加するため, モデルの説明変数の数は適切な説明変数だけにするのが望ましい. よって本研究では Wilks の Λ 統計量を用い, 変数増加法により変数選択を行うことで, 特徴量の次元数を削減する. Wilks の Λ 統計量は以下の式で定義される.

$$\Lambda = \frac{|S_w|}{|S_T|}$$

ここで, S_w は郡内の平方和・積和行列, S_T は全体の平方和・積和行列である. p 個の変量 x^* が q 群の判別に用いられてるとき, x^* に含まれない, 変量 x_i を追加したときの判別力の増加は

$$\Lambda(x_i \mid x^*) = \frac{\Lambda(x^*, x_i)}{\Lambda(x^*)}$$

と定義される. 本研究では, $\Lambda(x_i \mid x^*)$ が > 0.01 の時のみ, その変数を用いることにし, それ以外のときは棄却した.

4.2. 教示データの自動分類

イメージ語のようなあいまいさな要素を多く含む基準で, 教師データ群を作成し学習させる場合, それぞれの郡内から抽出される画像特徴量で分散が大きくなってしまうことがある. そのような, 分散が大きいデータ群からモデルを作成すると, 信頼性の低いモデルができてしまう. 一つでも信頼性の低いモデルが存在すると, そのモデルが要因となって, 画像検索システム全体の精度が大きく低下する問題がある.

そこで本研究では, イメージ語で分類されたデータ親グループと呼び, 親グループを k 平均法を用いて子グループへと細分化する. k 平均法により画像を自動分類することで, 短いサイクルで学習を繰り返すことができるため, システムの検証とモデルの学習を効率的に行える. 全ての子グループの組み合わせで, 画像がどのイメージに属するかマハラノビス距離を用いた判別分析で判定する. 画像は判別分析の際に, 勝率が 90%以上の子グループの親グループへと重複して分類される. 図 5 は, 未知画像が sexy_2 に属すると判定された場合に, その親グループである sexy へ分類している例である.

4.3. k 平均法

k 平均法の計算方法は以下の通りである.

1. 各データ x_i (i=1, ･･･, n)に対してランダムにクラスタを割り振る.
2. 割り振ったデータをもとに各クラスタの中心 V_j (j =1, ･･･, k)を計算する. 計算は割り当てられたデータの各要素の平均（重心）を用いる.
3. 各 x_i と各 V_j との距離を求め, x_i を最も近い中心のクラスタに割り当て直す.
4. 全ての x_i でクラスタの割り当てが変化しなかった場合は処理を終了する. それ以外の場合は新しく割り振られたクラスタから V_j を再計算して上記の処理を繰り返す.

k 平均法ではクラスター数をあらかじめ決定する必要がある, 本研究では, クラスター数を相関比が最大となる値

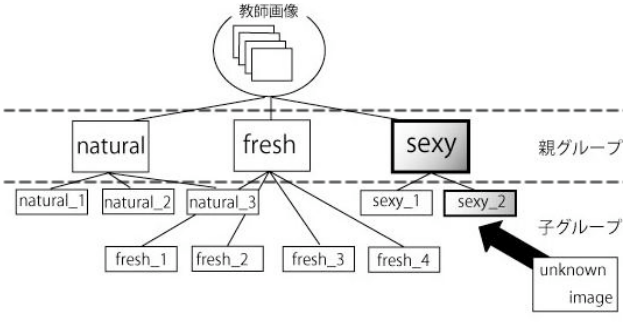


図 6. 教示データの細分化と未知画像の判定

に設定している. 相関比 η は以下の式で定義される.

$$\eta = 1 - \frac{|S_w|}{|S_T|}$$

ここで, S_w は郡内の平方和・積和行列, S_T は全体の平方和・積和行列である.

4.4. マハラノビス距離を用いた判別分析

今，教示用画像の画像特徴量ベクトル集合 $X = \{x_1, \cdots, x_n\}$ は，あらかじめユーザの主観的判断基準に基づいて分類された後に, k 平均法で細分化された K 個の群

$$X^{(i)} = \{x\}(i = 1, \cdots, K), \bigcup_{i=1}^{K} X^{(i)} = X$$

に分類されている. 各画像特徴量ベクトルには，どの群に属しているかの情報が与えられているものとする．この時，画像特徴空間上の X_i と例示画像との間のマハラノビス汎距離 D_i^2 は,

$$D_i^2 = (x_0 - \bar{x}_i)^T \sum_i^{-1} (x_0 - \bar{x}_i)$$

で定義される． $x_0, \bar{x}_i, \sum_i^{-1}$ はそれぞれ，例示画像の画像特徴量ベクトル， $X^{(i)}$ の重心ベクトル， $X^{(i)}$ の群内分散・共分散行列の逆行列である.

マハラノビス汎距離による判別では，画像特徴空間上で例示画像と各群までのマハラノビス汎距離 D_i^2 (i=1, … , N)を計算し，$\min\{ D_i^2 \mid i = 1, \dots, N \} = D_k^2$ となる群 k に例示画像が属すると判定する.

マハラノビス汎距離を用いた判別法には，良く知られている線形判別法に比して逐次学習が容易であるという特徴がある．また，SVM(Support Vector Machine) に代表されるような非線形判別法に比べて判別結果に自分の主観的判断基準が十分に反映されていないと感じた際の追加学習や，教示用データの入れ替えによる再学習が容易である. そのため，教示用データの追加・入れ替えの容易性とあいまって, ユーザの視覚感性の経時的変化に対応した感性モデルの再構築を少ない労力，コストで実現することができる.

4.5. 類似画像検索システムへの応用

i 番目の画像の特徴量ベクトルを X_i としたとき, i 番目の画像と最も類似した画像との距離を

$$D_i = \min\{ (| X_i - X_j |) \mid i = 1, \dots, N, j = 1, \dots, N, i \neq j \}$$

と定義する.本研究では全ての画像で類似度を計算し,これをイメージ語による画像検索システムに組み込むことで, より直感的な画像検索システムを実現している.

5. tfidf に準じたキーワード評価と画像分類

画像の特徴を十分に表現出来るようにキーワードを付与し，印象グループごとに tfidf に準じた評価により得点付けを行う.

まず，画像の特徴を適切に表現するために，様々な観点からキーワードを付与した．そして，印象グループごとに特徴的であるキーワードが算出され，画像ごとに付与されたキーワードの tfidf の合計得点を出す．この合計得点を利用して，画像分類を行う.

まだ印象が決定されていない画像に対して，画像の tfidf の合計得点が，印象グループごとに算出される. 画像は，最も tfidf 得点の高い印象グループが，画像の最も近い印象と判定し，分類する.

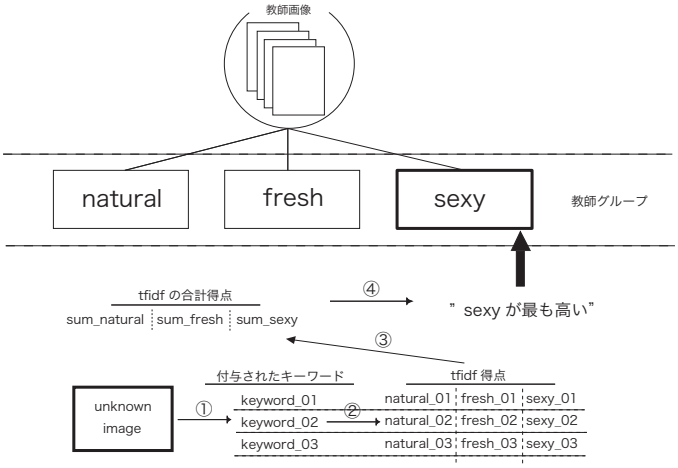


図 7. 未知画像の印象判定

画像がどの印象に最も近いか判定，画像ごとにキーワードを付与した．それらを利用し，tfidf に準じた，キーワードの出現頻度を算出する．図 6 は, 未知画像に付与されたキーワードから tfidf を求め, sexy に属すると判定された場合のその流れの例である.

5.1 キーワードの付与

付与するキーワードの内容は,

- 画像に写るオブジェクト
- オブジェクトの様子，雰囲気,
- 画像の様子や雰囲気を反映した概念

などである．図 7 は，画像と，その画像に付与したキーワードの例である.

5.2 キーワード

特定の印象 Ii (i=1, …, x) におけるキーワード Kj (j=1, …, y) の tfidf 値を算出する方法は，以下の通りである.

1. 印象 Ii に判定された全ての画像に振られた全てのキーワードの絶対数 Si，印象 Ii に判定された画像のうちキーワード Kj が付与された回数 Pk(k=1, …, z)，全ての印象の数 x，キーワード Kj が付与された印象の数 q，を求める.
2. 1．を利用して，tf，idf を以下のように求める.

tf = *Pk* / *Si*

idf = log(q / x) ［底は２］

tfidf = tf * idf

tfidf の値が高いほど，キーワード *Kj* は印象 *Ii* にて相対的に多く付与され，他の印象では付与されていないと言える．表１は，本研究で行った実験で算出された，印象「cyber」に属する，キーワード tfidf 得点の上位１０件である．

ビジネス	プッシュボタン
スティルライフ	ブレ
携帯電話	ソフトフォーカス
電話	通信
CG	情報
1 台	コミュニケーション
	無機質

図 8. 画像と付与されたキーワード例

表 1. 印象「cyber」の tfidf 値上位 10 件

cyber	tfidf
研究	0.00768225
テクノロジー	0.006108871
画面	0.004888705
通信	0.00437936
化学	0.00427621
開発	0.003866727
試験管	0.003665323
落ちつき	0.003491932
地球	0.002900013
世界視野	0.002748992

6. 実験

本研究では，画像特徴量を用いた視覚的印象のモデル化の実験と，tfidf に準じたキーワード評価を利用した画像分類の実験を行った．

6.1. 画像特徴量を利用した視覚的印象のモデル化

画像特徴量を用いた実験では，表 1 で示した 8 種のイメージ語を用いて視覚的印象のモデル化を試みる．教師データは，コンテンツ業務に関わる写真の専門家 2 人の合議により決定，作成された．実験用の画像コンテンツとしては，プロの写真家達による風景写真 7028 枚を使用した．表 1 で示されている各イメージ語で表現される印象を代表する写真を使用した．提案手法によって教師データがどの程度学習できているのかを確かめるため，Leave-One-Out（LOO）法を用いて検証を行った（表 1，表 2）．Leave-One-Out 法とは N 個の教師データのうち 1 個をとりのぞいてテストデータとし，残り N-1 個を使って学習するという手順を全データに対して繰り返して判別精度を評価する手法である．

表 2. 教示データのイメージ語による分類と loo 法による判別精度

親グループ	画像枚数	精度 (%)
Classic	1340	78.9
Cyber	163	77.5
Elegant	589	77.0
Fresh	1612	87.8
Modern	359	77.5
Natural	1446	80.5
Sexy	301	80.3
Wild	769	82.1

表 3. 教示データの自動分類を行った場合の loo 法による判別精度

親グループ	子グループ	画像枚数	精度 (%)	平均精度 (%)
Classic	Classic_1	297	94.4	
	Classic_2	308	94.1	
	Classic_3	130	92.4	93.6
	Classic_4	335	95.0	
	Classic_5	270	92.1	
Cyber	Cyber	163	77.5	77.5
Elegant	Elegant_1	82	88.8	
	Elegant_2	79	91.3	
	Elegant_3	143	93.6	91.2
	Elegant_4	182	91.8	
	Elegant_5	103	90.6	
Fresh	Fresh_1	395	95.3	
	Fresh_2	251	93.1	
	Fresh_3	364	98.3	96.3
	Fresh_4	227	98.0	
	Fresh_5	375	96.8	
Modern	Modren_1	71	89.1	
	Modren_2	154	87.3	87.8
	Modren_3	134	86.9	
Natural	Natural_1	255	94.0	
	Natural_2	408	95.6	
	Natural_3	178	91.4	94.6
	Natural_4	377	95.6	
	Natural_5	228	96.5	
Sexy	Sexy_1	127	88.6	89.6
	Sexy_2	170	90.5	
Wild	Wild_1	248	84.2	
	Wild_2	236	94.2	92.9
	Wild_3	97	89.8	
	Wild_4	188	93.4	

表2の精度は, 教示データの自動分類を行わないで, LOO 法を用いて判別精度の検証を行った場合, 表2は, 教示データの自動分類を行った後に, LOO 法を用いて判別精度の検証を行った場合である.

Cyber のイメージ語は教示データが 163 枚と少なかったため, 自動分類を行うことができなかったため, 表 1, 表 2 で同じ精度になっている.それ以外のイメージ語では,平均で約 10%以上の判別精度の向上に成功した.また,教示データが多い Classic, Elegant, Fresh, Natural, Wild ではどれも 90%以上の判別精度を実現しているが,教示データの少ない Cyber, Modern, Sexy では判別精度が低くなっていることもわかる.このことから,教示データを増やせば,現状では精度の低い Cyber, Modern, Sexy などのイメージ語でも 90%以上の判別精度を実現できるのではないかと考えられる.

6.2. tfidf に準じたキーワード評価による画像分類

tfidf に準じたキーワード評価を利用した画像分類の実験では，表 4 で示した 11 種のイメージごを用いて，画像の印象と画像の tfidf 得点を結びづけ，画像分類を行った．教師データは，6.1 と同様に印象を決定し，作成された．コンテンツは，風景写真の他に，画像特徴量では分類の難しい人間が多く写る写真など，計 30328 枚を利用した．教師データがどの程度学習出来ているかを確かめるために，6.1 と同様に L00 法を用いて検証を行った.

平均約 89.3%の判別精度が実現出来た. イメージによって差が観られるが，最低でも 84.4%の精度が認められる.

表 4. tfidf による画像分類と loo 法による判別精度

イメージグループ	画像枚数	精度 (%)
Active	2661	92.1
Classic	3743	89.2
Cyber	591	93.2
Elegant	2838	86.1
Fresh	4437	84.4
Modern	2779	84.8
Natural	8990	90.4
Pop	1238	85.0
Pretty	3923	92.5
Sexy	824	89.8
Wild	1304	94.4

7. まとめと今後の展望

本研究では，画像コンテンツから受ける視覚的印象のモデル化手法について論じた．提案手法では，教師画像群をその印象に基づき印象グループに分類することで主観評価をコンピュータに教示する．その際に，k 平均法を用いて自動的に細分化することで, より正確なコンピュータへの教示を可能にした. 画像の印象を主観的に判断する際, 人は画像の全域的特徴と局所的特徴を組み合わせて評価していると考えられる．そこで本研究では，Gabor フィルタを利用した構造的特徴とカラーヒストグラム, エントロピーを利用した全域的特徴を併用した. さらに, 変数選択法により適切な変数だけを残した後, 最終的な評価を下すモデルを, マハラノビス距離を用いた判別分析により構築した.

以上の工夫により，教師画像に対する印象に基づく画像自動分類の精度評価において, 平均して 90%以上の精度を実現した．さらに, これを類似検索システムへ応用し, 現在このシステムを評価運用中である.

解析結果からもわかるように, 教示データが少ない場合は, 判別精度が低くなってしまう場合が多い. さらに, 教示データを増やすことは, 最もシンプルで効果的な対策であるが, 簡単に教示データを追加することはできない. そこで, 今後は少ない教示データからでも, 正確なモデルを構築できるようなアルゴリズムを考えていきたい. また, 今回は 8 種類のイメージ語しか用いることが出来なかったため，今後はより多くのイメージ語を用いて同様の実験を行いたい.

また，キーワードを用いた画像分類では，画像特徴量で判別の難しい種類の画像での判別精度が高かったこと，また，より多くのイメージ語で評価出来たことは，高精度・高品質の画像分類を実現していくことに有用である．また，キーワードデータの特徴として，完全に客観的なデータでないのにも関わらず，特徴量として機能していたことも評価出来る．しかし，人的・時間的にコストのかかる手法なので，今後は継続的なシステムの制作のためにオートマティックにキーワードを付与する手法を必要とすることが課題となる.

謝辞

本研究は，一部，科学研究費補助金・基盤研究（S）「実空間における複合感性と状況理解の多様性のロボティクス的モデル化とその応用」(課題番号 19100004), 中央大学理工学研究所・共同研究「感性ロボティクス環境による共生的生活空間の構築と感性サービスへの応用」などの支援を受けて実施した.

本研究を進めるにあたり, 貴重なアドバイスやデータの提供を戴く（株）アマナホールディングス, (株) アマナイメージズの皆様，統計的な分析法や学習アルゴリズムに関してアドバイスを戴くＡＴＲ知識科学研究所の多田昌裕博士，日頃より，熱心な研究討論や実験への協力を戴く中央大学理工学部ヒューマンメディア研究室の皆様，感性ロボティクス研究センターの皆様に感謝します.

参考文献

[1] 加藤俊一ほか，:“ヒューマンメディア情報環境の展望と技術的課題” 電子技術総合研究所彙報，第 60 巻，第 8 号，pp.475, 509，1996.

[2] 多田昌裕，加藤俊一 ： “SVM を用いた視覚的印象の分析･学習と画像自動分類への応用” 電子情報通信学会技術研究報告，Vol.104, No.573(20050114), pp. 45-50，2004.

[3] 多田昌裕，加藤俊一 ：“類似する画像領域の特徴解析と視覚感性のモデル化”信学論，Vol.J87-D-II，No.10 ，pp.1983□ 1995，2004.

［4］栗田，加藤，福田，坂倉：“印象語による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.

[5] 中村享平，是角圭祐，森江隆：“V1 単純細胞の機能を模倣するガボールフィルタ LSI の評価” 電子情報通信学会技術研究報告，Vol.105, No.131(20050617) pp. 17-22

[6] 車妍，望月茂徳，蔡東生：“後期印象派絵画の色彩情報について解析”，情報処理学会研究報告．グラフィクスと CAD 研究会報告, Vol.2007, No.13(20070219) pp.109-113

[7] L. Spillmann and J.S. Werner, : Visual Perception, AcademicPress,San Diego, 1990.

[8] B. Sch¨olkopf, et al : “Estimating the support of a highdimensional distribution，Technical Report 99-87, MicrosoftResearch, 1999.

[9] 麻生英樹，津田宏治，村田昇 ： 統計科学のフロンティア 6 パターン認識と学習の統計学，岩波書店，東京，2003.

[10] CM ビショップ，元田浩，栗田多喜夫，樋口知之，松本祐治，村田昇 ： パターン認識と機械学習 上，シュプリンガージャパン株式会社，2007.

[11] 進藤博信:感性に伝わるフォトニケーション，英治出版，2004.

[12] http://www.amana.jp

[13] 多田昌裕，ZHANG Zhongfei(MARK)，加藤俊一 ：“MIL を用いた視覚的印象の分析・学習と画像自動分類への応用”電子情報通信学会技術研究報告 告,Vol.105,No.673(20060309) pp.13-18，2005