社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2006－CVIM－153 (3)
2006／3／16

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

多田 昌裕† Zhongfei(Mark) Zhang†† 加藤 俊一†††

† ATR メディア情報科学研究所 〒619-0288 「けいはんな学研都市」光台 2-2-2
†† ニューヨーク州立大学 ビンガムトン校 Binghamton, NY 13902-6000
††† 中央大学理工学部 〒112-8551 東京都文京区春日 1-13-27
E-mail: †mtada@atr.jp, ††zhongfei@cs.binghamton.edu, †††kato@indsys.chuo-u.ac.jp

あらまし コンテンツを販売する業界では従来は，例えば写真中の事物を説明するキーワードにより写真を分類し，検索に供してきた．しかし，「ナチュラル」「フレッシュ」のような主観的な印象に基づいて写真を分類したり，検索することが出来れば，より直観的なコンテンツの提供が可能になる．本研究ではMultiple Instance Learning (MIL) を用いて，専門家が写真の印象を評価する際に重視した構図や特徴を推定し，SVMによりそれをモデル化する手法を提案する．筆者らは構築したモデルを多種多様な商業用の写真に適用し，印象に基づく画像自動分類システムを試作した．
キーワード 視覚的印象，視覚感性，MIL, SVM，画像自動分類

Visual Impression Modeling Using Multiple Instance Learning

Masahiro TADA†, Zhongfei (MARK) ZHANG††, and Toshikazu KATO†††

† ATR Media Information Science Laboratories Hikaridai 2-2-2, Keihanna Science City, Kyoto, 619-0288 Japan
†† Dept. of Computer Science, Watson School of Engineering and Applied Sciences, Binghamton University Binghamton, NY 13902-6000, USA
††† Fac. of Science and Engineering, Chuo University Kasuga 1-13-27, Bunkyo-ku, Tokyo, 112-8551 Japan
E-mail: †mtada@atr.jp, ††zhongfei@cs.binghamton.edu, †††kato@indsys.chuo-u.ac.jp

Abstract Most of digital contents distributers use key words which correspond to objects in images to index various kinds of photo images. But these key words not always match with visual impression of images. In this paper, we propose a method to evaluate visual impression of images by using image key words. By statistically analysing typical photo examples of each image key word given by professional photographers, we have modeled their image evaluation process based on visual impressions (KANSEI Model). By using the KANSEI model, we have developed automatic image classification system for various kinds of photo images based on visual impressions.
Key words Visual Impression, Visual KANSEI, MIL, SVM, Automatic Image Classification

1. はじめに

従来のコンテンツ検索サービスの多くは，全てのコンテンツにメタデータを付与し，顧客が入力したキーワードとのマッチングを行うことで検索を実現している．しかし，特に芸術性の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．このような視覚的印象をメタデータで適切に記述することは非常に難しい．そのため，メタデータとのキーワードマッチングに基づく検索システムでは，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少なからぬズレが生じる．

ところで，我々は「ナチュラルな印象」，「フレッシュな印象」という具合に，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．本研究では，写真の専門家による印象の評価例をもとに，Multiple Instance Learning (MIL) を用いて専門家が写真の印象を評価する際に重視した構図や特徴を推定し，そのモデル化を試みる．筆者らは構築した感性モデルをプロの写真家達による多種多様な商業用の写真に適用し，印象に基づく画像検索システムを試作した．本稿では，試作したシステムの評価を通じ，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．

2. 本研究のアプローチ

筆者らはこれまで，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 [2]．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 [2]．

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．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，全ての顧客に検索キーとなる画像を用意させるというのは現実的ではない．例えば，作品に対する確固たるイメージや要求を持っているにもかかわらず，それを表現するのにふさわしいコンテンツが手元にない顧客のニーズに対して，内容型画像検索システムは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を MIL により学習，イメージ語による評価との間の関係をサポートベクターマシン (Support Vector Machine: SVM) および 1 クラス SVM によって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とで主観評価をコンピュータに教示する．その際，同一画像から複数の印象を受ける場合もあることを考慮し，一枚の画像を複数の印象グループに重複して分類することを許可した．しかしながら，画像は様々な情報を内包するため，そのままでは画像中のどの部位のどの情報が主観的な印象評価の際に影響を与えたのかを推定することは難しい．そこで本研究では，まず画像領域分割手法を用いて各画像からオブジェクト・背景を分離・抽出し，それらのうち鑑賞者の印象を左右する情報を持つものを MIL により推定，SVM および 1 クラス SVM によりモデル化する．

3. 3 点間コントラストベースの画像特徴量

筆者らは，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ている [2]．3 点間コントラストは次式で定義される．

$$\mathrm{Cont}^{(i)}(\mathbf{a}_1^{(i)}, \mathbf{a}_2^{(i)}, \mathbf{r}) = \frac{\{f(\mathbf{r}+\mathbf{a}_1^{(i)})-f(\mathbf{r})\}+\{f(\mathbf{r}+\mathbf{a}_2^{(i)})-f(\mathbf{r})\}}{|f(\mathbf{r}+\mathbf{a}_1^{(i)})|+|f(\mathbf{r}+\mathbf{a}_2^{(i)})|+2|f(\mathbf{r})|}. \quad (1)$$

ここで，$\mathbf{r}, (\mathbf{a}_1^{(i)}, \mathbf{a}_2^{(i)}), f(\mathbf{r})$ はそれぞれ参照点，変位，参照点 $\mathbf{r}$ の色彩であり，(1) 式の分母は視神経への刺激の強度，分子は刺激の差分である．3 点間コントラストは刺激強度で正規化している為，刺激強度に対してスケール不変であり，またノイズに強いという特性をもつ．

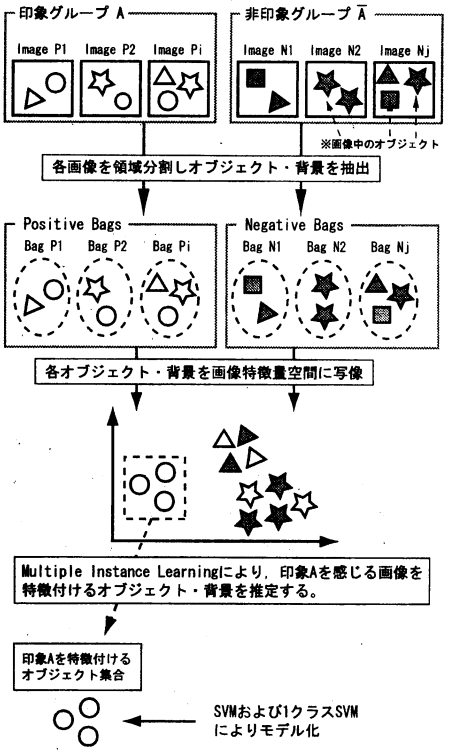


図 1　視覚的印象のモデル化手法概要

図 2 に本研究で採用したコントラストを測定するパターン（全 28 種）を示す．図中の “+” は参照点 $\mathbf{r}$，“*” は変位 $\mathbf{a}_1^{(i)}, \mathbf{a}_2^{(i)} \in \mathbf{R}^2$ $(i = 1, \cdots, 28)$ を示す．

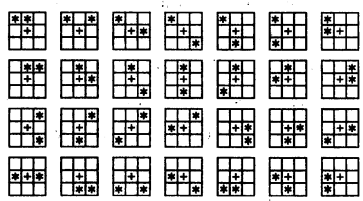


図 2　3 点間コントラストを測定するパターン

ところで，視覚の知覚過程には，ある点が刺激を受けて興奮作用を起こすと，その周辺の点が逆に抑制作用を起こす機構（側抑制と呼ぶ）があることが知られている [3]．側抑制は様々な明るさの背景の下で，注目点の近傍のコントラストを局所並列的に強調するメカニズムであると考えられる．また，興奮点が n 個ある場合には，各抑制効果の加算性が成り立つことが示されている．筆者らはこの側抑制の仕組みを画像特徴量に導入することを考え，(2) 式を定義した．

$$\gamma(i,\ \mathbf{r}) = \frac{\mathrm{Cont}^{(i)}(\mathbf{a}_1^{(i)}, \mathbf{a}_2^{(i)}, \mathbf{r})}{\sum_{i=1}^{n(I)} |\mathrm{Cont}^{(i)}(\mathbf{a}_1^{(i)}, \mathbf{a}_2^{(i)}, \mathbf{r})|}. \qquad (2)$$

ここで，$n(I)$ は参照点 $\mathbf{r} \in P$ で計測される 3 点間コントラスト $\mathrm{Cont}^{(i)}(\mathbf{a}_1^{(i)}, \mathbf{a}_2^{(i)}, \mathbf{r})$ のパターン数（本研究では 28 種）である．ある $\mathrm{Cont}^{(i)}(\mathbf{a}_1^{(i)}, \mathbf{a}_2^{(i)}, \mathbf{r})$ の値が大きくても，参照点 $\mathbf{r}$ 周りで大きな値をとる 3 点間コントラストのパターンが多ければ $\gamma(i,\ \mathbf{r})$ の値は相対的に小さく，大きな値をとるパターンが少なければ相対的に大きくなる．

本研究では，画像平面 P 上の画像領域 P_k における $\gamma(i,\ \mathbf{r})$ の平均および分散を局所特徴量，平均色および色の分散を全域的特徴量とする．この全域的特徴量と局所的特徴量を併せたものを画像特徴量ベクトル $\mathbf{x}$ として採用する．

4. 視覚的印象のモデル化

4.1 イメージ語による視覚的印象の評価

イメージ語と画像の特徴とを関係付けようという試みは，今までにもいくつかなされてきた [4]．それら先行研究の多くは，画像に対する印象をイメージ語ごとに 5 段階（とても感じる，やや感じる，あまり感じない，まったく感じない，どちらともいえない）で評価し，この結果と画像の特徴とを正準相関分析を用いて対応付けている．

しかし，筆者らが複数の写真の専門家にヒアリング調査を行ったところ，彼らは画像にある印象を「どの程度」感じるかにはあまり頓着せず，「感じるか否か」のみ考慮することが多いという結果を得た．この結果に基づき，本研究では画像群を，イメージ語ごとに画像からその印象を受けるのか否かで 2 分し（印象グループと非印象グループ），教師データとした．ところで，一枚の画像から受ける印象は必ずしも 1 つに限定されるわけではなく，一枚の画像から同時に複数の印象を受けるということも考えられる．本研究ではこの問題に対処するため，印象に基づくグループ分けをするにあたり，同一画像を異なる印象グループに重複して分類することを許可した．

4.2 画像領域分割

多くの写真画像は画像平面全域が同じテクスチャで構成されているわけではないため，画像部位によって画像特徴量の分布は異なる．画像データは本質的に多義性を有し，たとえ同一画像平面上であっても，画像特徴量の分布が異なる部位は各々異なる意味・情報を表現していることが多い．そこで本研究では，まず画像平面中で画像特徴量の値の分布が同じであると考えられる領域を，情報量基準の一種である MDL を用いたクラスタリング手法により抽出する．

4.2.1 MDL 基準

MDL 基準は Rissanen により，符号化における記述長最小化 (Minimal Discription Length) 原理として導出されたものであり，モデルのパラメータの記述長とモデルを用いてデータを記述したときの記述長の和が最小になるモデルを最良とみなす [5]．

画像 I_i の画像平面 P は $M \times M$ の小領域 P_k $(k = 1, \ldots, M^2)$ に分割されており，さらに各小領域は $N \times N$ の極小領域 $P_{k,\alpha}$ $(\alpha = 1, \ldots, N^2)$ に分割されているとする．また，極小領域 $P_{k,\alpha}$ から抽出した画像特徴量ベクトルを $\mathbf{x}_{k,\alpha}$ とし，その集合を $X_k = \{\mathbf{x}_{k,\alpha}\}$ とする．

いま，小領域 P_k 及び P_l を統合し，画像特徴量ベクトルの集合 $X_{kl} = \{\mathbf{x}_{kl,\alpha'} \mid \mathbf{x}_{kl,\alpha'} \in X_k \bigcup X_l\}$ を生成することを考える．$\mathbf{x}_{kl,\alpha'}$ にパラメータ $\theta_{kl} = (\mu_{kl}, \Sigma_{kl})$ の n 変量正規分布 $p(\mathbf{x}; \theta_{kl})$ を仮定すると，小領域を統合した場合の MDL は，

$$\begin{aligned}\mathrm{MDL}_{(uni)} &= -\log L(\hat{\theta}_{kl}) + \frac{J_{(uni)}}{2} \log \frac{2N}{2\pi} \\ &\quad + \log \int \sqrt{|I(\theta_{kl})|}\, d\theta_{kl}, \end{aligned} \qquad (3)$$

で算出することができる．ここで，$\hat{\theta}_{kl}$，$J_{(uni)}$，$I(\theta_{kl})$，$|\cdot|$ はそれぞれ，θ_{kl} の最ゆう推定量，パラメータの自由度，Fisher 情報行列，行列式である．また，$L(\cdot)$ はゆう度関数であり，$L(\cdot) = \prod p(\cdot)$ である．

一方，小領域 P_k と P_l を統合しない場合，$\mathbf{x}_{k,\alpha} \in X_k$，$\mathbf{x}_{l,\alpha} \in X_l$ にパラメータ $\theta = (\theta_k, \theta_l)$ を持つ確率分布

$$p(\mathbf{x}; \theta) = \frac{p(\mathbf{x}; \theta_k)^{\delta_k}\, p(\mathbf{x}; \theta_l)^{\delta_l}}{\int p(\mathbf{x}; \theta_k)^{\delta_k}\, p(\mathbf{x}; \theta_l)^{\delta_l} d\mathbf{x}}, \qquad (4)$$

$$\delta_i = \begin{cases} 1 & \text{if } \mathbf{x} \in X_i \\ 0 & \text{otherwise} \end{cases}, \qquad (5)$$

を仮定する．ここで，$p(\mathbf{x}; \theta_k)$，$p(\mathbf{x}; \theta_l)$ はそれぞれパラメータ θ_k，θ_l を持つ n 変量正規分布の確率密度関数である．

このときの MDL は，

$$\begin{aligned}\mathrm{MDL}_{(div)} &= -\log L(\hat{\theta}) + \frac{J_{(div)}}{2} \log \frac{2N}{2\pi} \\ &\quad + \log \int \sqrt{|I(\theta)|}\, d\theta, \end{aligned} \qquad (6)$$

で算出することができる．ここで，$\hat{\theta} = (\hat{\theta}_k, \hat{\theta}_l)$ は $\theta = (\theta_k, \theta_l)$ の最ゆう推定量，パラメータの自由度 $J_{(div)} = 2\, J_{(uni)}$ である．

4.2.2 画像領域分割アルゴリズム

画像 I_i の領域分割手順は以下のとおりである．

（1） 画像 I_i の画像平面 P を $M \times M$ のメッシュに分割し，各々を小領域 P_k $(k = 1, \ldots, M^2)$ とする．また，小領域 P_k をさらに $N \times N$ のメッシュに分割し，各々を極小領域 $P_{k,\alpha}$ $(\alpha = 1, \ldots, N^2)$ とする．

（2） $P_k \subset P$ $(k = 1, \ldots, M^2)$ から画像特徴量ベクトル集合 $X_k = \{\mathbf{x}_{k,\alpha} \mid \alpha = 1, \ldots, N^2\}$ を抽出する．

（3） 次元圧縮のため，$X = \bigcup_{k=1}^{M^2} X_k$ に主成分分析を適用し，$Y = \bigcup_{k=1}^{M^2} Y_k$，$Y_k = \{\mathbf{y}_{k,\alpha}\}$ を得る．

（4） 特徴空間 Y における小領域の重心間距離を測定し，距離最小となる小領域の組を統合対象とする．

（5） 小領域を統合した場合としなかった場合の MDL をそれぞれ算出し，$\mathrm{MDL}_{(uni)} \leq \mathrm{MDL}_{(div)}$ ならば，小領域を統合する．そうでなければ，重心間距離が 2 番目に小さい小領域の組を統合対象とする．

（6） (4)–(5) の処理を繰り返す．

こうして抽出された領域は，画像 I_i を構成するオブジェクト

や背景に相当すると考えられる．なお，本稿では $M=25, N=4$ として領域分割を行った．

4.3 Multiple Instance Learning

画像領域分割アルゴリズムにより画像から抽出されたオブジェクト・背景には，印象評価に影響を与えるものと与えないものとが混在している．従来の教師あり学習では，全てのデータにラベル付けがなされていることが前提となるため，これら抽出したオブジェクト単位でさらに教示を行う必要があるが，ユーザの負担を考えると現実的ではない．全てのオブジェクト・背景が印象評価に影響を与えると仮定してモデル化を行うことも可能ではあるが，実際には影響を与えるものと与えないものとが混在しているため，高い精度のモデルを構築することは難しい．

このようなあいまいさを含む教師データを扱うため，本研究では Multiple Instance Learning (MIL) [6] を採用する．従来の教師あり学習では，全てのデータに対してラベル付けがなされている必要があったのに対し，MIL ではいくつかのデータ（インスタンス）の集合であるバッグと呼ばれる単位ごとにラベル付けを行う．バッグ中のインスタンスがどれか一つでも "Positive" であれば，そのバッグは "Positive" とラベル付けされる．逆にバッグ中のインスタンスが全て "Negative" であれば，そのバッグは "Negative" とラベル付けされる．

本稿では，画像中の各オブジェクトに対応する画像特徴量ベクトルをインスタンス，画像単位でのインスタンスの集合（一枚の画像を構成するオブジェクトに対応するインスタンス集合）をバッグと呼ぶ．たとえば，印象 A を受ける画像 I_{A_i} 中のオブジェクトの少なくとも一つには，印象 A を喚起させるものがあるはずである．この場合，I_{A_i} に対応するバッグ B_{A_i} には印象 A に関して "Positive" とラベル付けする．逆に，印象 A を受けない画像 $I_{not\_A_j}$ 中のどのオブジェクトにも，印象 A を喚起させるものはないはずなので，対応するバッグ $B_{not\_A_j}$ には "Negative" とラベル付けする．

MIL では，あいまいさを含んだ教師データから，"Positive" なインスタンス（印象評価に影響を与えるオブジェクト・背景）を推定する手段として，Diverse Density (DD) [6] と呼ばれる値を用いる．DD の基本的な考え方は，各々異なる Positive Bag に所属するインスタンスが近辺に多く存在し，かつ Negative Bag のインスタンスは近辺に存在しない場所に，求めるべき "Positive" なインスタンスが存在するというものである．

いま，Positive Bag を $B_1^+, B_2^+, \ldots, B_m^+$，Negative Bag を $B_1^-, B_2^-, \ldots, B_n^-$ とする．また，B_i^+，B_i^- の j 番目のインスタンス（オブジェクトから抽出した画像特徴量ベクトル）を各々 B_{ij}^+，B_{ij}^- とする．このとき，あるオブジェクト $\mathbf{x}$ の DD の値は次式で表される．

$$\mathrm{DD} = \prod_i \Pr(\mathbf{x}|B_i^+) \prod_i \Pr(\mathbf{x}|B_i^-). \tag{7}$$

ただし，

$$\Pr(\mathbf{x}|B_i^+) = 1 - \prod_j (1 - \Pr(B_{ij}^+ = \mathbf{x}),) \tag{8}$$

$$\Pr(\mathbf{x}|B_i^-) = \prod_j (1 - \Pr(B_{ij}^- = \mathbf{x})), \tag{9}$$

$$\Pr(B_{ij} = \mathbf{x}) = \exp(-||B_{ij} - \mathbf{x}||^2). \tag{10}$$

ここで，$||B_{ij} - \mathbf{x}||^2$ は特徴量空間における B_{ij} と $\mathbf{x}$ の距離である．本研究では，印象グループに属する全ての画像から抽出したオブジェクト・背景に関して DD の値を計算し，閾値を超えたもののみを印象評価に影響を与えるオブジェクト・背景とみなし，SVM によるモデル化を行う．

4.4 SVM

サポートベクターマシン（SVM）は線形識別器の一種であるが，カーネル関数を用いることにより非線形へと拡張することが出来る [7]．線形識別器は 2 クラスが線形分離可能であるときには高い認識率を期待できるが，非線形で複雑な問題に対してはその限りではない．そこで，非線形な写像 Φ を用いて線形分離性を高めることが考えられる．ただし，Φ で写像される先での内積 $(\Phi(\mathbf{x}) \cdot \Phi(\mathbf{x}'))$ は，元の空間で定義されるカーネル関数 $K(\mathbf{x}, \mathbf{x}')$ の値と一致するものとする．カーネル関数としては，次式で定義されるガウシアンカーネルなどが用いられる．

$$K(\mathbf{x}, \mathbf{x}') = \exp(-\frac{||\mathbf{x} - \mathbf{x}'||^2}{\sigma^2}) \tag{11}$$

いま，印象グループ A と印象グループ B の識別関数を SVM で構築することを考える．まず，印象グループ A, B に属する画像からオブジェクト・背景 Obj_i $(i = 1, \ldots, N_{AB})$ を抽出し，その画像特徴量を $\mathbf{x}_i$ $(i = 1, \ldots, N_{AB})$ とする．また，それぞれのクラスラベルを y_i $(i = 1, \ldots, N_{AB})$ と表記する．Obj_i が印象グループ A に属し，かつ DD の値が閾値を超えるのであれば $y=1$，Obj_i が印象グループ B に属し，かつ DD の値が閾値を超えるのであれば $y=-1$ とし，DD が閾値を超えないものは無視する．ここで N_{AB} は印象グループ A, B に属する画像から抽出したオブジェクト・背景の総数である．

このとき印象グループ A と B の識別関数は

$$f_{AB}(\Phi(\mathbf{x})) = \mathrm{sgn}(\mathbf{w}^T \Phi(\mathbf{x}) + b) \tag{12}$$

となる．関数 $\mathrm{sgn}(u)$ は $u > 0$ のときには 1，$u \leqq 0$ のときには -1 となる符号関数である．また，$\mathbf{w}$ は重みベクトルである．識別平面と教示用データとの間のマージンを最大化するには，以下の問題を解けばよい．

目的関数：

$$\frac{1}{2}||\mathbf{w}||^2 + C \sum_{i=1}^{N_{DD}} \xi_i \rightarrow \text{最小化}$$

制約条件：

$$y_i(\mathbf{w}^T \Phi(\mathbf{x}_i) + b) \geqq 1 - \xi_i, \quad \xi_i \geqq 0$$

ここで，C は目的関数の第一項と第二項のバランスをとるためのパラメータであり，その設定は実験的に行う．また，N_{DD} は DD の値が閾値を超えた Obj_i の個数である．本研究では，印

象グループの全組合せについて各々識別関数 $f_{AB}(\Phi(\mathbf{x}))$ を求め，それらの出力を総合することで当該オブジェクト・背景ではどのような印象が喚起されるのかを判定する．

4.5 1クラスSVM

SVMは2クラスの識別器のため，どのようなデータであっても必ずどちらかのクラスに識別される．そのため，SVMのみを用いた場合，外れ点（どの印象も喚起されないような画像）が誤って印象グループに識別される恐れがある．そこで本研究では，外れ点検出のため，1クラスSVM[8]を導入し，SVMと併用することを提案する．

1クラスSVMは，ガウシアンカーネルを用いて写像を行うと，元の空間の外れ点は原点近くに写像されるという性質を利用し，原点と印象グループとを分けるような超平面を求める[9]．

いま，印象グループ A に属する画像から抽出したオブジェクト・背景を Obj_i $(i=1,\ldots,N_A)$，その画像特徴量ベクトルを $\mathbf{x}_i$ $(i=1,\ldots,N_A)$ とする．ただし，Obj_i のDDの値は閾値を超えているものとする．また，N_A は印象グループ A に属する画像から抽出したオブジェクト・背景の総数である．

このとき識別関数は

$$f_{(1svm),A}(\Phi(\mathbf{x})) = \mathrm{sgn}(\mathbf{w}^T\Phi(\mathbf{x}) - h) \qquad (13)$$

となる．あらかじめ決められた割合 $\nu \in (0,1]$ の印象グループが原点側に残る（外れ点とされる）ような超平面を求めるには，以下の問題を解けばよい．

目的関数：

$$\frac{1}{2}||\mathbf{w}||^2 + \frac{1}{\nu N_{DD}}\sum_{i=1}^{N_{DD}}\xi_i - h \rightarrow \text{最小化}$$

制約条件：

$$\mathbf{w}^T\Phi(\mathbf{x}_i) \geq h - \xi_i, \quad \xi_i \geq 0$$

ここで，N_{DD} はDDの値が閾値を超えた Obj_i の個数である．

本研究では，印象グループごとに独立に識別関数 $f_{(1svm),A}(\Phi(\mathbf{x}))$ を求め，たとえSVMによって印象グループに識別されたとしても，1クラスSVMによって外れ点と判断された場合には無視するものとする．

なお，本稿ではLIBSVM[10]のソースコードをもとにしてSVMおよび1クラスSVMを実装した．

5. 実　　験

本稿では，表1に示す9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．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人物写真を使用した．

教師データとしては，コンテンツ業務に関わる写真の専門家10人の合議により決定された，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，提案手法の有効性を検証した。

まず，各教師画像から画像領域分割手法を用いてオブジェクト・背景を分離抽出した．次に，印象グループごとにMILを適用し，DDの値によって印象を喚起するオブジェクト・背景を推定，それらをSVMおよび1クラスSVMによりモデル化した．なお，本研究では印象グループの全組合せに対して各々SVMによる2クラス識別器を生成し，それらの出力を総合してどの印象群に属するべきかを判定する．画像平面の一定の割合を占めるオブジェクト・背景が印象グループ A に分類されたとき，そのオブジェクトを含む画像からは印象 A を感じると判定する．

表1　使用したイメージ語と教師画像枚数

イメージ語	教師画像枚数
Active/Sporty	663
Classic/Authentic	213
Cute/Pretty	618
Elegant/Romantic	537
Fresh/Clean	1379
Modern/Urban	1138
Natural/Relax	1525
Pop	284
Sexy/Gorgeous	306

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，交差確認法（CV）法を用いて検証を行った（表2）．交差確認法とは 教師データを m 個のグループ（本稿では $m=5$）に分け，$(m-1)$ 個のグループで学習，残り1個のグループでテストを行うという手順を繰り返すテスト手法である．また比較のため，MILを用いずにモデル化を行った場合（全てのオブジェクト・背景は印象評価に何らかの影響を与えると仮定した場合）についても同様に検証した．

表2　CV法による精度評価

イメージ語	MILあり	MILなし
Active/Sporty	82.5%	69.4%
Classic/Authentic	79.7%	69.3%
Cute/Pretty	72.8%	56.6%
Elegant/Romantic	80.8%	68.2%
Fresh/Clean	90.5%	69.5%
Modern/Urban	85.2%	68.3%
Natural/Relax	87.2%	67.5%
Pop	79.9%	69.9%
Sexy/Gorgeous	71.9%	58.8%

表2に示すように，提案手法ではいずれのイメージ語に関しても70%以上の精度で専門家の評価を再現できていることが分かる．一方，MILを用いなかった場合の精度は，用いた場合と比べ10%程度低くなっている．これは，画像領域分割アルゴリズムにより画像から抽出されたオブジェクト・背景には，印象評価に影響を与えるものと与えないものとが混在しており，このあいまい性がモデルの精度に悪影響を及ぼしたためだと考えられる．それに対し，MILを用いた場合，DDの値を手がかりにして印象評価に影響を与えるオブジェクト・背景のみを抽出，モデル化を行ったため，教師データからあいまいさを排除し，より高い精度のモデル化を実現できたと考えられる．

6. ま と め

本稿では，画像コンテンツから受ける視覚的印象のモデル化

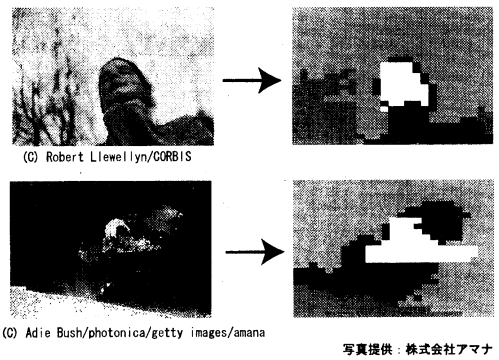
手法について論じた．

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，同一画像から複数の印象を受ける場合もあることを考慮し，一枚の画像を複数の印象グループに重複して分類することを許可した．

画像の印象を主観的に判断する際，人は画像中の一部のオブジェクトや背景を重点的に評価していると考えられる．しかしながら，画像は様々な情報を内包するため，そのままでは画像中のどの部位のどの情報が主観的な印象評価の際に影響を与えたのかを推定することは難しい．そこで本研究では，まず画像領域分割手法を用いて各画像からオブジェクト・背景を分離・抽出し，それらのうち鑑賞者の印象を左右する情報を持つものをMILにより推定する手法を提案した．

こうして抽出した鑑賞者の印象を左右する情報のみを用い，どの印象グループに属するのかを判定する識別器群をSVMを用いて構築した．ただし，SVMは2クラス識別器であるため，単独で使用した場合，どのようなデータも必ずどちらかのクラスに分類されてしまい，外れ点が印象グループと誤認される恐れがある．そこで，本研究では外れ点検出のために1クラスSVMを導入し，SVMと相互補完的に運用した．

以上の工夫により，専門家による写真の印象判断の過程を高い精度でモデル化することができた．



(C) Robert Llewellyn/CORBIS

(C) Adie Bush/photonica/getty images/amana

写真提供：株式会社アマナ

図3　画像領域分割の結果例

謝辞　本研究を進めるにあたり，大量の写真のコンテンツと専門家による印象の解釈の事例をご提供いただいた（株）アマナ[12]に感謝します．本研究の一部は，NICTの委託研究「超高速知能ネットワーク社会に向けた新しいインタラクション・メディアの研究開発」，科学研究費補助金 基盤研究(S) 15100004，および中央大学理工学研究所・共同研究「感性ロボティックス環境による共生的生活空間の創造」の資金により実施されました．

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，"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.

[3] L. Spillmann and J.S. Werner, Visual Perception, Academic Press, San Diego, 1990.

[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] J. Rissanen. "Fisher Information and Stochastic Complexity," IEEE Trans. Inf. Theory, vol.42, pp.40-47, 1996.

[6] O. Maron and T. Lozano-Pérez, "A framework for Multiple Instance Learning." In advances in Neural Information Processing Systems, 10, MIT press, 1998.

[7] 赤穂昭太郎，津田宏治，"サポートベクターマシン -基本的仕組みと最近の発展-，" 数理科学，No.444，pp.52-58，2000.

[8] B. Schölkopf, et al., "Estimating the support of a high-dimensional distribution," Technical Report 99-87, Microsoft Research, 1999.

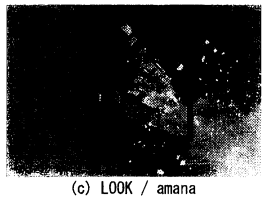
[9] 麻生英樹，津田宏治，村田 昇，統計科学のフロンティア6 パターン認識と学習の統計学，岩波書店，東京，2003.

[10] http://www.csie.ntu.edu.tw/~cjlin/libsvm/

[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.

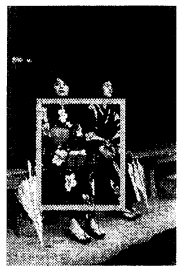
[12] http://www.amana.jp

Active / Sportyな印象を受ける画像例



(c) LOOK / amana

Classic/Authenticな印象を受ける画像例



(c) YASUNO/A.collection/amana

Natural/Relaxな印象を受ける画像例

(c) Kareem Black/photonica/getty images/amana

写真提供：株式会社アマナ

四角枠内は，MILにより印象評価に影響を与えると判断されたオブジェクトである．

図4　教師画像の一例