

機関番号：13601

研究種目：基盤研究(C)

研究期間：2008～2010

課題番号：20500129

研究課題名（和文）階層型確率モデルに基づく画像表現とその応用に関する研究

研究課題名（英文）A study on image representation by hierarchical probabilistic models and its applications

研究代表者

丸山 稔 (MARUYAMA MINORU)

信州大学・工学部・准教授

研究者番号：80283232

研究成果の概要（和文）：

本研究では、画像特徴の生成過程を階層型の確率モデルを用いてモデル化し、これに基づいて画像認識を行う手法を考案した。本手法においては画像特徴発生確率を決定する“トピック”を考え、このトピックの分布により画像特徴生成過程を記述する階層型の確率モデルである確率的トピックモデルを用い、これに基づくパラメータ推論方式を示すと共に、このモデルを文書画像の領域属性認識等に用いるための基本手法を示した。さらに提案手法を情報量の少ない小領域認識に適用するための拡張法も示している。さらに本研究では、確率的トピックモデルを用いず、識別関数の階層構造により高速・高精度な識別を行うための手法についても検討した。なお、本研究の一部は文書画像解析に関する国際ワークショップ（IAPR International Workshop on Document Analysis Systems 2008）において論文賞（best paper award honorable mention）を受賞した。

研究成果の概要（英文）：

We have mainly studied about hierarchical probabilistic model for image representation and its application to category recognition. In our method, we used probabilistic topic model. We proposed a method for classifying regions in a document image based on pLSA (probabilistic latent semantic analysis). Classification is carried out based on topic proportion of the region that is estimated via MAP-based EM algorithm. The model is further extended to classify smaller image regions. We also presented the method to detect characters in a natural scene images by using decision stumps. A part of this research work received best paper award honorable mention at the IAPR workshop on Document Analysis Systems 2008.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2008年度	1,200,000	360,000	1,560,000
2009年度	1,000,000	300,000	1,300,000
2010年度	600,000	180,000	780,000
年度			
年度			
総計	2,800,000	840,000	3,640,000

研究分野：総合領域

科研費の分科・細目：情報学・知能情報学

キーワード：画像認識、確率的トピックモデル、パターン識別、識別関数、pLSA

### 1. 研究開始当初の背景

画像や動画中に含まれるオブジェクトを抽出したり、画像全体の内容を理解・認識したりする能力の実現は情報検索やヒューマンコンピュータインタラクションなど広範な分野の進展にとって重要な課題である。従来このような画像内容理解については例題からの多クラス識別関数の学習問題として研究が進められてきているが、学習の対象となるオブジェクトカテゴリ等は事前に確定されて、学習データも一括して与えられる条件の下で各カテゴリに対応する識別関数を構築する研究が多い。しかしながら、多様なオブジェクトに関して柔軟な識別能力を発揮する系を構築するためには、特定のカテゴリに特化した識別関数の構築に留まらず、一般的なオブジェクトについて、画像特徴の生成過程を含めたモデル化を行い、これに基づく認識手法を検討する必要があると考えられる。

### 2. 研究の目的

本研究の目的は、柔軟性、拡張性の高い画像認識系を構築することである。このためには特定のオブジェクト（カテゴリ）に特化した識別関数構築ではなく、一般的な画像モデル化手法に基づいて画像認識手法を検討する必要があると考えられる。そこで、本研究では画像のモデルとして、観測できない変数を導入し、この変数に基づいて画像中の特徴量生成が制御されるような階層型の一般画像特徴記述モデルを考える。本研究ではこの確率モデルのパラメータを、与えられた学習データから学習・推論するためのアルゴリズムを開発すると共に、この確率モデルの画像分類、画像中のオブジェクト認識などに応用し、その有効性を検証することを目的とする。

### 3. 研究の方法

本研究では階層型確率モデルとして主として確率的トピックモデルを使用した。これは画像中の特徴量の生成過程を記述するもので、画像は観測できない変数（潜在変数）である複数のトピックにより構成され、各トピックに付随した特徴量発生確率に従って画像特徴が生成されると考える階層型のモデルである。確率的トピックモデルは元々は文書（テキスト）解析のために考案されたものであるが、多次元ベクトルで表される画像特徴ベクトルをベクトル量子化によって“単語”として扱うことにより、画像表現にも適用することが可能になる。

本研究ではこの階層型確率モデルを中心とした画像モデル化・推論アルゴリズムを考案すると共に、得られた方式の実装による検

証計算機実験、及びその評価によって手法の改良に努めた。本研究の認識・識別対象としては、主として文書画像を用い、文書画像各領域の属性を認識することを試みた。さらに、情景画像中に文字領域を含む画像領域の抽出、一般の多カテゴリ画像の分類などへの展開も試みた。

### 4. 研究成果

(1) 確率的トピックモデルに基づく画像表現と画像分類への応用

画像の認識・識別のためには、画像から特徴量を抽出し、これに基づいた処理が行われるのが通常である。画像特徴は多次元ベクトルの形で得られることが多いが、これらをベクトル量子化（クラスタリング）して“単語”として扱うことにより、画像を単語の集合（BoW=Bag of Words）として表現でき、文書分類等で用いられる確率的トピックモデルを適用することができる（図 1-1 参照）。この場合画像は“文書”（単語集合）とみなして認識・識別を行うことになる。

確率的トピックモデルにおいて、各トピック  $z$  は観測されない変数（潜在変数）であり、各単語  $w$  の発生確率  $p(w|z)$  を与える。本研究では、各文書  $d$  毎に固有のトピック発生確率  $p(z|d)$  を有する pLSA モデルを用い、これを画像認識、特に文書画像中の小領域のタイプ（属性）認識に用いる手法を提案し、その有効性を検証した。いま、トピック  $z=j$  の下での単語発生確率を  $p(w|z=j)=\phi_w^{(j)}$ 、文書  $d$  におけるトピック  $j$  の発生確率を  $p(z=j|d)=\theta_j^{(d)}$  とおく。pLSA モデルのパラメータ  $\{\phi_w^{(j)}\}, \{\theta_j^{(d)}\}$  は画像データから得られる単語集合に基づいて EM アルゴリズムを用いて推論することができる。また、このとき、E-step の計算より、 $p(z|w, d)$  も得られる。本研究では、このとき Dirichlet prior  $\phi_w^{(j)} \sim \text{Dir}(\phi_w^{(j)} | \alpha, \dots, \alpha)$ ,  $\theta_j^{(d)} \sim \text{Dir}(\theta_j^{(d)} | \beta, \dots, \beta)$  を用いて、MAP 推定に基づく EM アルゴリズムを適用した。これらのパラメータのうち  $\{\theta_j^{(d)}\}$  は文書（画像データ）毎に決定する必要があるが、新規データが与えられたときには fold-in アルゴリズムを用いてパラメータ推定を行った。

本研究では、パラメータ  $\{\phi_w^{(j)}\}, \{\theta_j^{(d)}\}$  の推定後、これらの値に基づいてカテゴリ識別を行う。例えば、文書画像の場合は、画像を小領域に分割し、得られた領域を“文書”  $d$  として扱って、特徴量のベクトル量子化（単語抽出）を行い、これらに対して上記の推論方式を適用して得られたパラメータから、カテゴリ  $c$ （日本語領域、英語領域、手書き領域、数式領域）を確率  $p(c|d)$  に基づいて決定する。 $p(c|d)$  は以下の方式で計算を行った。

$$p(c|d) = \sum_{j=1}^T p(c|z=j)p(z=j|d)$$

ここに、Bayes の定理より  $p(c|z) \propto p(z|c)p(c)$  である。 $p(c)$ ,  $p(z|c)$  については以下で求めた

$$p(c) \approx N_c / N, \quad p(z|c) \approx \sum_{category(d_i)=c} p(z|d_i) / N_c$$

$N, N_c$  は学習データの総数及びカテゴリ  $c$  の総数である。

図 1-2 に提案手法による識別性能を他の識別方式と比較した結果を示す。より複雑な構造を有する確率的トピックモデルである LDA や、SVM 等と比較して提案手法は同等以上の識別能力を持つことを示すことができた。なお、この実験の対象となった画像中の領域のサイズは  $240 \times 240$  ピクセルであった。また、ベクトル量子化の元となる特徴抽出のためには SIFT descriptor/detector を用いた。本研究においては SIFT 以外にも Harris-affine、wavelet などの特徴抽出手法も実装し比較実験を行っているが、文書画像の回転やズームなどの幾何変換を加えた際の耐性実験より、SIFT が良好な性質を持つことを確認している。

なお、本研究のうち確率的トピックモデル pLSA を文書画像中の領域属性識別に適用した研究成果は、文書画像解析に関する代表的な国際ワークショップの一つである 2008 年度 IAPR International Workshop on Document Analysis Systems (DAS2008) において論文賞 (Best paper award honorable mention) を受賞した。

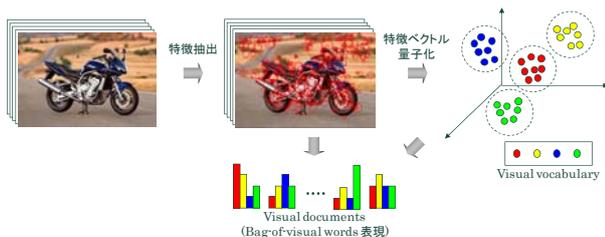


図 1-1 : BoW 表現に基づく画像認識手法

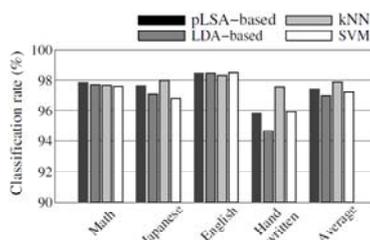


図 1-2 : 種々の識別関数との比較

(2) 階層型の画像モデルに基づく文書画像のカテゴリ詳細識別方式

画像が単一カテゴリのオブジェクトだけでなく複数カテゴリのオブジェクトを含む場合、領域分割と認識・識別を並行して行う必要がある。(1) で述べた本研究の成果は、分割された領域についてその属性を認識する手法に関するものであった。この手法により、信頼度が高い認識結果を得るためには、対象領域中に認識のためのデータであるベクトル量子化された特徴量が十分多く含まれている必要がある。このため対象領域を小さく取るとは、内部に含まれる特徴量の数の減少を招くために、望ましくない。実際

(1) で用いた領域サイズ  $240 \times 240$  を細分し、 $60 \times 60$  領域に対して確率的トピックモデル pLSA に基づく手法を適用したところ、図 2-1 に示すような結果を得た。図 1-2 の結果と比較すると明らかなように、小領域になったことで識別性能の低下を招いていることが分かる。一方、画像オブジェクトの精密な抽出、分類のためには、識別対象はできるだけ小さく取れることが望ましい。そこで、本研究では、この問題点を解決するために、階層型の推論モデルにより小領域の属性識別を行う手法を考案し、文書画像を対象として有効性の検証を行った。

本研究においては、(1) で考えたような十分な特徴量を含む比較的大きな画像領域に関する推論結果を、この領域を細分した小領域の推論に利用する階層型、粗密 (coarse to fine) 型の推論方式を用いる。十分な特徴量を含む画像領域から生成される“文書”を  $D$ 、これを細分して得られる小領域を  $d$  とおく。このとき  $d$  に(1)で述べた手法と同じ手法を適用した場合、識別能力の低下を招くことが示されているため、以下のような経験分布に基づく推論手法で  $p(c|d)$  を求めることをまず試みた。

$$p(c|d) \propto \sum_w n(w,d) p(c|w,d)$$

ここで  $p(c|w,d)$  は親領域  $D$  を利用して

$$p(c|w,d) = \sum_z p(c|z) p(z|w,d) \approx \sum_z p(c|z) p(z|w,D)$$

により算出する。 $p(c|z)$  は(1)で述べた手法を用いる。この手法による小領域識別性能実験の結果を図 2-2 に示す。図 2-1 に示した結果と比較すると、特徴量の少ない小領域であっても、提案手法により識別能力が向上することが分かる。

本研究では、上記の手法をさらに改善するために、(1)で述べた MAP 推定に基づく EM アルゴリズムを拡張する手法に関する検討も

行った。小領域で問題になるのは含まれている特徴量が少ないときの推論精度である。(1)で示した手法では、このための対策としてDirichlet prior 導入によるMAP型の推論を行っているが、 $\{\theta_j^{(d)}\}$  の prior として用いているのは  $\text{Dir}(\theta_j^{(d)} | \beta, \dots, \beta)$  なる単純な設定を行っているDirichlet分布であった。小領域dに関する推論を行う場合、領域の階層構造を利用して、十分な特徴量を含む親領域Dに関する推論結果をpriorに用いることが考えられる。そこで、本研究においては、領域Dに関して(1)で述べた手法を適用して $\{\theta_j^{(D)}\}$ を求めた後、Dの細分であるdのパラメータ $\{\theta_j^{(d)}\}$ をEMアルゴリズムによって推定する際に、以下のようなDirichlet priorを用いる。

$$\theta_j^{(d)} \sim \text{Dir}(\theta_j^{(d)} | \beta_1, \dots, \beta_r), \quad \beta_j = \beta + \gamma \theta_j^{(D)}$$

ここに $\gamma$ は親領域の影響度を制御するためのパラメータである。これを用いた場合のEM (fold-in) アルゴリズムは以下のようになる。

E-step

$$\hat{z}_{wj}^{(d)} \leftarrow \frac{\phi_w^{(j)} \theta_j^{(d)}}{\sum_j \phi_w^{(j)} \theta_j^{(d)}}$$

M-step

$$\theta_j^{(d)} \propto \sum_w \hat{z}_{wj}^{(d)} n(w, d) + \beta + \gamma \theta_j^{(D)} - 1$$

EMアルゴリズム適用後は(1)で示した手法と同様に  $p(c|d)$  を求め識別を実行することができる。

本手法の識別能力とパラメータ $\gamma$ の影響を調べるために、文書画像の領域分割を対象として検証実験を行った。まず各カテゴリ(日本語、英語、手書き、数式)からサンプリングした領域のうち2種類を混合した合成データを生成し、これに対して本手法を適用した。図2-3に種々の混合比率に対して $\gamma$ (親領域の影響度を表すパラメータ)を変化させて小領域識別能力を計測した結果を示す。 $\gamma=5$  周辺の値を選択することにより、種々の混合比率を持つ領域に対して良好な識別能力を示すことが分かる。また本手法と他手法(SVM, kNN, 経験分布に基づく手法)と比較した結果を図2-4に示す。他手法と比較しても本手法、特に適切な $\gamma$ 値設定を行った場合は良好な識別能力を有することが分かる。

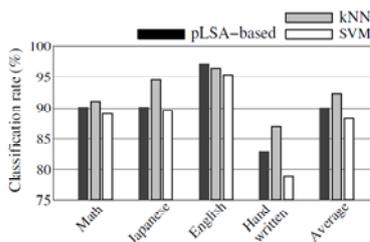


図 2-1：細分割領域に関する識別性能比較

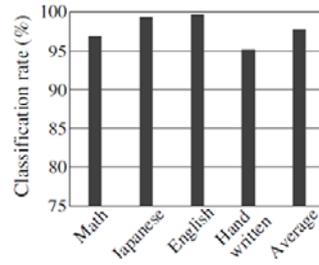


図 2-2：経験分布に基づく手法による手法

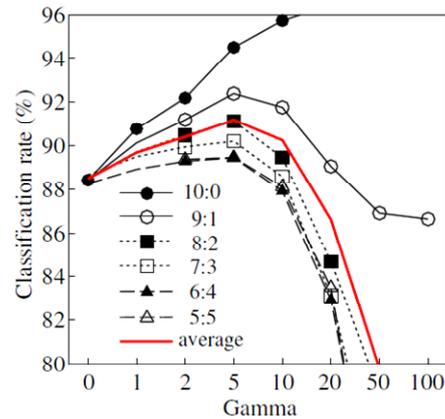


図 2-3：細分割パラメータによる識別性能比較

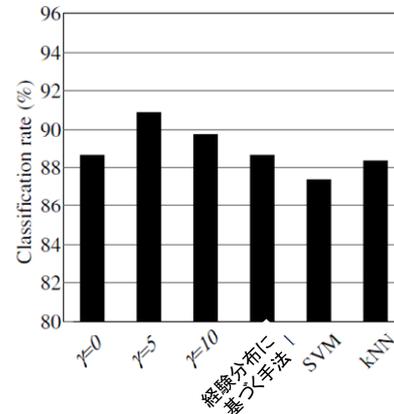


図 2-4：細分割認識能力の比較

### (3) Decision Stumps に基づく画像中のオブジェクト検出法

これまで述べて来た手法は各カテゴリのデータ生成過程を階層型の確率モデルで記述するものであった。これらの手法はその記述能力や柔軟性に特長があるものの、計算効率、計算速度を要求されるような応用例の場合には必ずしも十分な性能を有しているとは言えない。本研究では、認識・識別の対象カテゴリが確定している教師付学習の問題設定の下で、階層型の識別関数を利用して、識別を高速化するための手法の検討も行った。

本研究では情景画像中の文字部分の抽出

(図 3-1 参照) を対象として、識別能力と計算速度の向上のための手法を検討した。

高い識別能力を得るためには、適切な特徴量の選択が必要になる。本研究ではオブジェクト検出手法によく用いられる Haar wavelet, HOG(Histogram of Oriented Gradiend) などを用いた。情景画像中では文字領域部分の濃淡値ヒストグラムの形状は、(地の部分と文字部分の) 多峰状態を成し、典型的な単峰状分布である正規分布 (gaussian) からは大きく外れているものと考えられる。そこで、本研究では正規分布からのズレを計るために、正規分布の場合は 0 となるモーメント特徴量の尖度、歪度 (kurtosis, skew) を併せて用いた。また、識別関数評価の高速性を実現するために、本研究では識別関数構築のための基本要素として以下に示すように decision stump を用い、これらの線形結合により最終的な識別関数を構築した。decision stump は高次元特徴ベクトルのただ一つの要素にのみ着目し閾値計算をするだけであり、RBF をカーネルとする SVM 等と比較して計算コストが低いものと考えられる。

$$h(x) = \text{sgn}[x_i - \theta_i] = \begin{cases} 1 & x_i > \theta_i \\ -1 & \text{otherwise} \end{cases}$$

$$f(x) = \text{sgn}\left[\sum_{i=1}^T \alpha_i h_i(x)\right]$$

本研究では、使用できる複数特徴量種類を指定した上で、boosting アルゴリズム (AdaBoost) を用いて decision stump に基づく識別関数の学習を行うことで、識別に使用する特徴ベクトル種類と成分の選択を行った。表 3-1 には、単一種類の特徴量を用いて RBF をカーネルとする SVM によって学習を行った場合の識別能力評価実験の結果を示す。表中 Sparse Haar (5%) は Haar wavelet 係数のうち、値が大きいもの 5% のみを識別に使用した場合を示している。これに対し、decision stump を boosting アルゴリズムによって学習し識別を行った結果を図 3-2、3-3 に示す。図 3-2 は単一種類の特徴量を使用した場合、図 3-3 は複数種類の特徴量を用いた場合の結果である。以上の結果より、特に複数の特徴量に基づく識別関数を boosting を用いて組み合わせることにより、単純な SVM に比べて高精度の識別が可能になったことが分かる。サイズ 640×480 の画像 10 枚を処理した場合の処理時間の比較及び処理時間の内訳を表 3-2、3-3 に示す。なおこの実験においては Intel Core 2 Quad 2.83GHz (1 コアのみ) を使用した。これらの結果より decision stump に基づく複数特徴種類を用いた階層型識別関数は高速・高精度な情景中の文字領域抽出が可能であることが分かる。



図 3-1: 情景画像中からの文字抽出例

表 3-1: 各特徴量を用いた SVM の識別能力比較

Feature	Precision	Recall
Haar	0.856	0.887
Sparse Haar (5%)	0.898	0.860
HOG	0.825	0.895
Moment	0.825	0.964

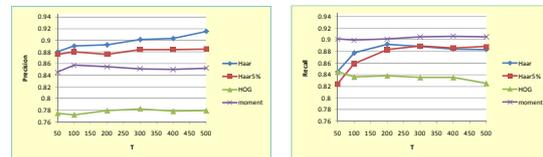


図 3-2: 単一特徴量の場合の性能比較

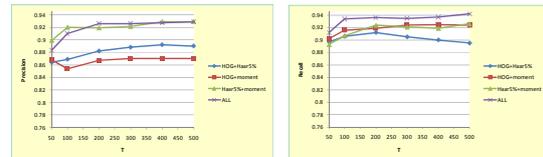


図 3-3: 複数特徴量の場合の性能比較

表 3-2: 処理時間の比較

特徴抽出			識別	
Haar5%	moment	HOG	RBF-SVM	stump
344[ms]	32[ms]	176[ms]	1879[ms]	11[ms]

表 3-3: 処理時間の内訳

Haar 5% + RBF-SVM	Haar 5% + stump	Haar5% + moment + stump	HOG + moment + stump
2221[ms]	354[ms]	386[ms]	219[ms]

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

- [雑誌論文] (計 1 件)
- ① 山口拓真, 丸山稔, “確率的トピックモデルによる文書画像の領域分割”, 電子情報通信学会論文誌 D, Vol.J92-D, No.6, pp.876-887 (2009). (査読有)  
<http://hdl.handle.net/10091/12047>

〔学会発表〕（計 5 件）

- ① 殿岡大輔, 丸山稔, “単語出現頻度と画像特徴に基づく確率的トピックモデルの論文検索への応用”, 電子情報通信学会信越支部大会論文集, 8C-4, pp.136, (2010). (査読無し)
- ② 殿岡大輔, 丸山稔, “単語出現頻度に基づく論文集合の分類”, 電子情報通信学会信越支部大会論文集, 8A-2, pp.137, (2009). (査読無し)
- ③ M.Maruyama and T.Yamaguchi, “Extraction of characters on signboards in natural scene images by stump classifiers”, Proc. 10th International Conference on Document Analysis and Recognition (ICDAR2009), pp. 1365--1369 (2009). (査読有)  
<http://hdl.handle.net/10091/12016>
- ④ T.Yamaguchi and M.Maruyama, “Feature extraction for document image segmentation by pLSA model”, Proc. 8th IAPR International Workshop on Document Analysis Systems (DAS2008), pp. 53-60 (2008). (査読有)  
<http://hdl.handle.net/10091/12236>
- ⑤ T.Yamaguchi and M.Maruyama, ”Image categorization by a classifier based on probabilistic topic model”, Proc. 19th International Conference on Pattern Recognition (ICPR2008), pp. 1 -- 4 (2008). (査読有)  
<http://hdl.handle.net/10091/12019>

## 6. 研究組織

### (1) 研究代表者

丸山 稔 (MARUYAMA MINORU)

信州大学・工学部・准教授

研究者番号：80283232