釧路工業高等専門学校紀要第48号（平成27年）

MDL 基準を用いた特徴選択と Bag of Visual Words モデルへの応用の試み

天元 宏[*1]

An Approach to Feature Selection by Using MDL Criterion and Application to the Bag of Visual Words Model

Hiroshi TENMOTO

Abstract — We are proposing a histogram approach for feature selection and classification. The axes are divided into equally-spaced intervals, while the division numbers differ among the axes. The main difference from the similar approach is that feature selection is embedded in the model selection criterion. As a result, this criterion brings feature selection for a small number of training samples and convergence to the optimal Bayes error for a large number of training samples. In this paper, we applied the proposed method to an image classification problem in which local image features and bag of visual words model are used for obtaining feature vectors. Through experiments, we confirmed that the proposed method can successfully select important visual words with additional small errors.

Keywords: Pattern Recognition, Image Classification, Unsupervised Learning, Bag of Visual Words Model, Randomized Algorithm

1. はじめに

パターン認識の研究分野は, 画像等の観測データを直接扱う信号処理寄りの特徴抽出と, 特徴抽出後の特徴空間における識別機の構成や誤差評価などを扱う学習理論に大きく分けられる. 一般に, 特徴抽出で測定された特徴は大規模となる傾向があるが, それはまた, 学習理論において次元の呪いと呼ばれる性能低下を引き起こす. そのため, 特徴抽出で測定された大規模な特徴集合から, 識別に貢献する少数の有効な特徴を抜き出す特徴選択に関する研究が国内外にて盛んである (例えば文献 [1]). 特徴選択の主要な手法は, 元の大規模な特徴集合から一部の特徴を仮に選択し, それによる識別性能が向上するように, 選択した部分特徴集合を修正して行くものである. しかし, それらの手法では, 識別性能を向上させるために各特徴を使うべきか否かという指標を得ることはできるものの, 特徴そのものをどの程度の精度で記述 (表現) するべきかという指標を得ることができない.

そこで著者らは, 近年ラフ集合 [2, 3] の分野で研究が盛んな粒度 (測定精度) の考え方を学習理論へ応用し, 識別対象となるデータを識別に最も適した粒度で離散化して識別を行う手法を提案している [4, 5]. 提案手法における粒度の調整は, 任意の粒度における訓練サンプルの識別状況を情報量基準により評価することで実行する. 提案手法により最適に調整された各特徴の粒度はまた, その特徴の識別への貢献度評価を与える事から, 提案手法は特徴選択を一般化した手法と見なすことができ, 著者らはこの手法をソフト特徴選択と呼んでいる. この事実は, 提案手法が特徴抽出と学習理論を統一的な見地から捉え, システム全体のコスト削減, 高速化, 高性能化を一体として統合的に遂行できる枠組みとなる可能性がある事を示している.

さらに, システムを利用するユーザーの観点より評価すると, 生成される識別規則の可読性も重要となる. 近年隆盛なサポートベクトルマシンやニューラルネットワークといった手法ではユーザーには直感的に理解不能な規則を生成してしまう. これに対し, 提案手法の規則は各特徴の区間 (高い・低い等の程度) を条件とした IF-THEN 列となり, 極めて可読性が高い識別規則となる.

本報告では, 著者らが提案しているこのソフト特徴

[*1] 釧路高専情報工学科

選択の手法の概略を紹介すると共に, 画像認識への応用を試みた結果を報告する. 本報告の画像認識では局所特徴量と Bag of Visual Words モデルを用いた.

2. ヒストグラムに基づく識別機

識別対象となるデータを離散化して識別を行うということは, ヒストグラムに基づく識別機を構成することを意味する. ヒストグラムに基づく識別機は, 訓練サンプル数が無限大へと向かう際にベイズの意味で最適な識別機への収束が保証されている強力な識別機の一つである. これまで多数のヒストグラムに基づく識別機へのアプローチが提案されてきた (例えば文献 [6, 7, 8, 9, 10]).

それらの研究の主な目的は, ベイズ誤差への収束を高速化することであった. その意味で, 主要な注意は大規模なデータの場合へ向けて払われていた. しかし, 現実には小規模や中規模のデータに直面しなければならない場合の方が多い. その場合, 良く取られるアプローチの一つは, 特徴選択による次元の削減である. 特徴選択段階を経ることで, 現実に対処しなければならない難しい問題を, 高い予測性能を備えた, 言い換えると識別機のパラメータのより正確な推定が可能な識別機を構成できる様な, 別の小さな問題へと変換することがきでる.

この目的のため, 特徴選択手法として様々なアルゴリズムが提案されてきた (様々な特徴選択手法の比較検討としては文献 [11] を参照されたい). 提案法では, これらの両方を達成することを考える. つまり, 小規模や中規模なデータに対しては特徴選択を, 大規模なデータに対してはベイズの意味での最適性を追求したいと考える. 提案法では, これを各特徴軸上で異なる大きさの分割とすることで達成する. もし一切分割が必要なければ, その特徴軸は識別に貢献しない. その一方で, 目の細かい分割が必要な特徴軸に対しては, その特徴軸は識別にとって重要である. 提案法では, その様な最適な分割を MDL 基準に従って探求する.

3. ヒストグラムに基づく識別機の MDL 符号化

m 次元ユークリッド空間 $U = R^m$ 中に n 個の訓練サンプルが与えられたという条件下で, ヒストグラムに基づく識別機を構成する問題を考える. その場合, 一つの訓練サンプルは

$$x = (x_1, x_2, \ldots, x_m) \in R^m$$

で表現され, 訓練サンプル列 z^n はクラス集合 $Y = \{1, 2, \ldots, c\}$ を伴って

$$\begin{aligned} z^n &= (x, y)^n \\ &= \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\} \\ &\in (U \times Y)^n \end{aligned}$$

と記述できる.

MDL 基準 [12] に従い, 受信者が x^n および c, m を知っているとの仮定の下で, クラスラベル列 y^n を送信するコストを測る. $L(\phi \,|\, x^n)$ を識別機 (識別規則) ϕ を送信するために必要なビット長とする. 加えて, ϕ が与えられたとの条件の下で, クラスラベル列を送信するコストを $L(y^n \,|\, \phi, x^n)$ とする. その場合, 全体のコストは

$$L(y^n, \phi \,|\, x^n) = L(y^n \,|\, \phi, x^n) + L(\phi \,|\, x^n)$$

と書くことができる. もちろん, $L(y^n \,|\, \phi, x^n)$ は $n \log_2 c$ (y^n の素朴なビット長) よりも小さくしたい. 本手法の場合, $L(\phi \,|\, x^n)$ はヒストグラムの情報を送信するためのビット長, つまり, 各軸の分割数であり, $L(y^n \,|\, \phi, x^n)$ はそのヒストグラムが形成するセルにおけるクラスラベル列のビット長の総和である. 本手法ではセルと同様に $x_1, x_2, \ldots, x_n$ も数値の辞書順などの適当な方法で並んでいるものと仮定する.

本手法で識別機として考えるのはヒストグラムであり, 第 i 番目の軸を 2^{d_i} 個の等間隔の区間に分割する. 各軸の両端は訓練サンプル上での最小および最大で決定する. 従って, 分割は m タプル

$$\boldsymbol{d} = (d_1, d_2, \ldots, d_m)$$

で表現する. また, $d = \sum_{i=1}^{m} d_i$ として, d により分割数の総和を記述する. そうすると, $\prod_{i=1}^{m} 2^{d_i} = 2^d$ 個のセルが存在することになる.

本手法では MDL 基準の意味で最適な分割 $\boldsymbol{d}$ を発見する. MDL 基準では, 短いビット長ほど識別のためにより良い分割であることを意味する. 以下, 符号長 $L(y^n, \phi \,|\, x^n)$ を順に記述する.

手順 1. まず簡単のため, 一つのセルにおけるサンプルのラベル列の符号化を行う. n をそのセルに落ちるサンプルの個数とする. クラスラベル列を

$$y_1, y_2, \ldots, y_n$$

と記述する．そうすると，この列の符号化には

$$\log_2 \left(\begin{array}{c} n+c-1 \\ c-1 \end{array} \right) + \log_2 \left(\begin{array}{c} n \\ n_1, n_2, \ldots, n_c \end{array} \right)$$

ビットを必要とする．ここで，受信者に n の値を知らせる必要はない．これは一旦分割が知られると，各セルに落ちるサンプルの個数はカウントできるからである．n を第 r 番目のセルに対する n^r で書き換えると，

$$\begin{aligned} & L(y^n \,|\, \phi, x^n) \\ &= \sum_{r=1}^{R} \left\{ \log_2 \left(\begin{array}{c} n^r+c-1 \\ c-1 \end{array} \right) \right. \\ &\qquad \left. + \log_2 \left(\begin{array}{c} n^r \\ n_1^r, n_2^r, \ldots, n_c^r \end{array} \right) \right\} \end{aligned}$$

を得る．ここで，R は空でないセルの個数である．

また，幾つかのセルでは単一のクラスから来るサンプルのみとなることが起こりうる．その様な場合を考慮するため，全ての非空な R 個のセルを単一クラスのセル R^P と，混合クラスのセル R^M に分割する．従って，R^M の値および対応する R^M 個の位置を送信するために，

$$\log_2 R + \log_2 \left(\begin{array}{c} R \\ R^M \end{array} \right)$$

ビットを必要とする．加えて，各単一クラスのセルのクラスラベルを送信する必要があるため，

$$R^P \log_2 c$$

ビットを必要とする．結果として

$$\begin{aligned} & L(y^n \,|\, \phi, x^n) \\ &= \log_2 R + \log_2 \left(\begin{array}{c} R \\ R^M \end{array} \right) + R^P \log_2 c \\ &\quad + \sum_{r=1}^{R^M} \left\{ \log_2 \left(\begin{array}{c} n^r+c-1 \\ c-1 \end{array} \right) \right. \\ &\qquad \left. + \log_2 \left(\begin{array}{c} n^r \\ n_1^r, n_2^r, \ldots, n_c^r \end{array} \right) \right\} \end{aligned} \tag{1}$$

を得る．

手順 2．次に分割の符号化を行う．本手法では受信者へ $\boldsymbol{d}$ を知らせなければならない．そのため最初に d，そして $d_1, \ldots, d_m (d = \sum d_i)$ を符号化する．従って，

$$L(\phi \,|\, x^n) = \log_2 \left(\begin{array}{c} d+m-1 \\ m-1 \end{array} \right) + \log_2^* d$$

ビットが必要となる．しかし，多くのパターン認識問題では，少数の特徴のみが識別のために重要である．その様な場合では，このビット長を

$$\begin{aligned} & L(\phi \,|\, x^n) \\ &= \log_2 m + \log_2 \left(\begin{array}{c} m \\ m-m^+ \end{array} \right) + \log_2^* d \\ &\quad + \log_2 \left(\begin{array}{c} d \\ d_1, d_2, \ldots, d_m \end{array} \right) + \log_2 d \\ &\quad + \sum_{i=1}^{m^+-2} \log_2 d_i \end{aligned} \tag{2}$$

として圧縮することが期待できる．ここで，m 個の要素中の m^+ 個は非ゼロの d_i である．

トータルで，式 (1) および式 (2) より，ビット長は

$$\begin{aligned} & L(y^n, \phi \,|\, x^n) \\ &= L(y^n \,|\, \phi, x^n) + L(\phi \,|\, x^n) \\ &= \log_2 R + \log_2 \left(\begin{array}{c} R \\ R^M \end{array} \right) + R^P \log_2 c \\ &\quad + \sum_{r=1}^{R^M} \left\{ \log_2 \left(\begin{array}{c} n^r+c-1 \\ c-1 \end{array} \right) \right. \\ &\qquad \left. + \log_2 \left(\begin{array}{c} n^r \\ n_1^r, n_2^r, \ldots, n_c^r \end{array} \right) \right\} \\ &\quad + \log_2 m + \log_2 \left(\begin{array}{c} m \\ m-m^+ \end{array} \right) \\ &\quad + \log_*^2 d + \log_2 \left(\begin{array}{c} d \\ d_1, d_2, \ldots, d_m \end{array} \right) \\ &\quad + \log_2 d + \sum_{i=1}^{m^+-2} \log_2 d_i \end{aligned}$$

で評価することとなる．

大きな n に対してはこの式のエントロピー評価を利用することが可能である．その目的のため，次の二つの近似を用いる．

$$\log_2 \left(\begin{array}{c} n+c-1 \\ c-1 \end{array} \right) \simeq (c-1) \log_2 n$$

$$\begin{aligned} & \log_2 \left(\begin{array}{c} n \\ n_1, n_2, \ldots, n_c \end{array} \right) \\ &\simeq nH \left(\frac{n_1}{n}, \frac{n_2}{n}, \ldots, \frac{n_c}{n} \right) - \frac{c-1}{2} \log_2 n \end{aligned}$$

ここで

$$H \left(\frac{n_1}{n}, \frac{n_2}{n}, \ldots, \frac{n_c}{n} \right) = -\sum_{i=1}^{c} \frac{n_i}{n} \log_2 \frac{n_i}{n}$$

である．すると，$L(y^n, \phi \,|\, x^n)$ は

$$
\begin{aligned}
& L(y^n, \phi \,|\, x^n) \\
& \simeq \left(RH\left(\frac{R^P}{R}, \frac{R^M}{R}\right) + \frac{1}{2}\log_2 R + R^P \log_2 c \right) \\
& \quad + \left(n \sum_{r=1}^{R^M} \frac{n^r}{n} H\left(\frac{n_1^r}{n^r}, \frac{n_2^r}{n^r}, \ldots, \frac{n_c^r}{n^r}\right) \right. \\
& \qquad \left. + \frac{c-1}{2} \sum_{r=1}^{R^M} \log_2 n^r \right) \\
& \quad + \left(mH\left(\frac{m^+}{m}, \frac{m-m^+}{m}\right) + \frac{1}{2}\log_2 m \right) \\
& \quad + \left(\log_2^* d + dH\left(\frac{d_1}{d}, \frac{d_2}{d}, \cdots, \frac{d_m}{d}\right) \right. \\
& \qquad \left. + \frac{m-1}{2}\log_2 d \right) \\
& = \mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{IV} \qquad (3)
\end{aligned}
$$

と書き換えることができる．

4. ヒストグラムに基づく識別機の MDL 符号化の評価

本手法の基準である式 (3) がどの様に働くかを確認してみたい．この基準を用いることで，n が無限大に向かうに従い本手法の識別機はベイズの意味で最適な識別機に近づく．しかし，基準式 (3) ではそれ以上のことも考察できる．最初に，最も優勢な項は項 I および，項 II，項 IV である．ある特定の $\boldsymbol{d}$ により訓練サンプル上で完全な識別が行われるとき，$R^M = 0$ かつ $R^P = R$ であるため，項 II は消える．その場合，問題は項 I および項 IV の最小化となる．従って，本手法が成し遂げたいことはまず最初に d の値を最小化し，その後，エントロピー $\{d_i/d\}$ を最小化することである．つまり，この方法は，エントロピーを減少させるためにいくつか d_i をゼロ化すること，すなわち特徴選択を推進する．これはこれまでの MDL 研究 [6, 7, 8, 9, 10] との大きな違いである．訓練サンプルの個数 n が小さい場合，例え訓練サンプル上の識別が完全でなくとも，いくつかの d_i をゼロ化する働きが起こりうる．その様な場合は項 I および項 IV が優勢であるためである．以上より，本手法の基準はサンプル数が少ない場合に特徴選択として働くことが分かる．

5. 部分クラス法を用いた未知サンプルの識別

本手法のアプローチにおいて未知サンプルを識別する方法を記述する．高次元特徴空間では，訓練サンプルを含むセルは容易に疎となる．従って，サンプルを含むセルを内挿するためにヒストグラムによるアプローチと共に，部分クラス法 [13] を用いる．この方法はあるクラスの中でそのクラス領域として超矩形の集合を見つける．各超矩形は次の二つの条件を満たす「部分クラス」と呼ばれる．

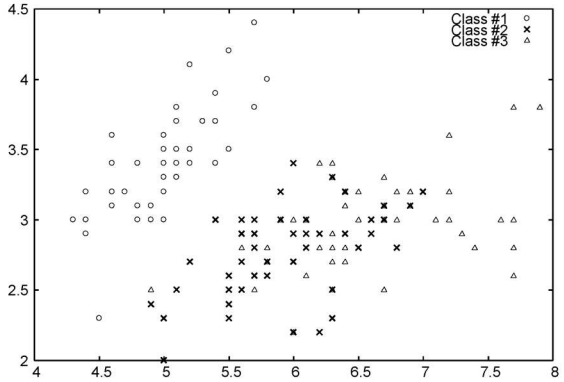
1. 排他性: その超矩形は負のクラスからのいかなるサンプルも含まない．
2. 極大性: その超矩形は上記の排他性を満たす全ての部分集合の族において，包含関係に関して極大である．

この問題はサンプル数に関して組合せ全検査を必要とするため，本手法では準最適解を得るために確率的アルゴリズム [13] を利用する．

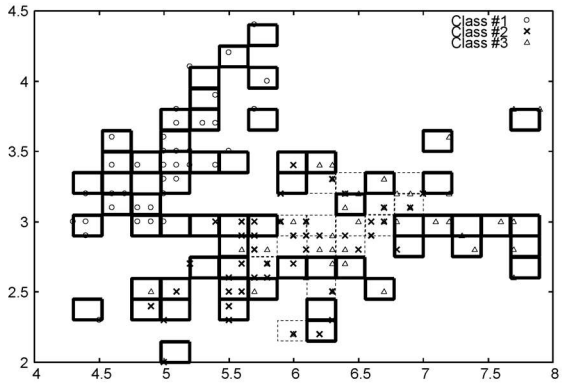
識別のための具体的な手続きは以下の通りである．

1. 未知サンプルを含むセルを見つける．
2. そのセルが訓練サンプルを含むならば，多数決規則を適用することによりクラスラベルを決定する．
3. そのセルが訓練サンプルを一つも含まないならば，部分クラス集合を用いて次の様にクラスラベルを決定する．
   (a) サンプルが落ちるセルを含む部分クラスを見つける．
   (b) 一つ以上の部分クラスが見つかった場合，最も内部にサンプルを位置させる部分クラスを採用し，その部分クラスでクラスラベルを決定する．
   (c) 一つも部分クラスが見つからない場合，最も近い部分クラスを採用し，その部分クラスのクラスラベルで決定する．

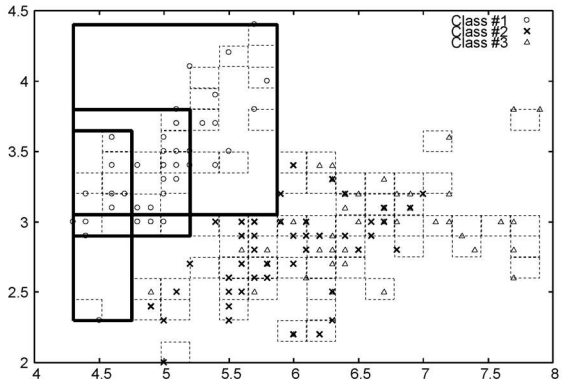
統計的パターン認識の例で良く用いられるアヤメデータの最初の 2 特徴上の実行例を図 1 に示す．訓練サンプルはセルで粗く分割され，部分クラスは単一のクラスのサンプルから成る「清潔な」セルの上のみに形成されている．結果的に得られる識別境界は，セルと部分クラスに基づいて構成されている．



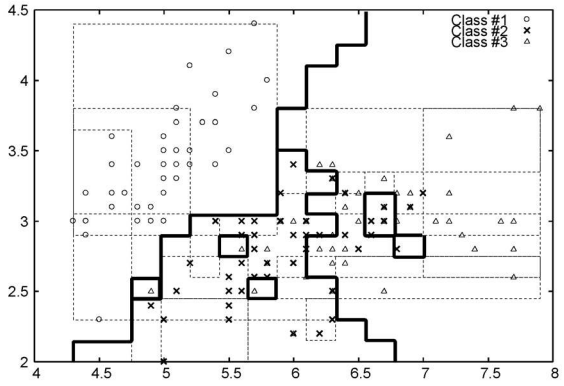
(a) 訓練サンプル集合



(b) 単一クラスのセルと混合クラスのセル



(c) クラス ω_1 のセル上の部分クラス集合



(d) 識別境界

図 1　アヤメデータセットの最初の 2 特徴に対する識別境界構成の例

6. 遺伝的アルゴリズムを用いた MDL 符号長の最小化

本手法では式 (3) を最小化する解を探索するために, GA(遺伝的アルゴリズム) を用いている. 各 d_i を遺伝子, m タプル $(d_1, d_2, \ldots, d_m)$ を染色体として符号化した. それにより各染色体は分割 $\boldsymbol{d}$ を表すこととなる.

以下の実験では, 母集団の大きさ, つまり染色体の数を 100 とし, 交叉確率を 0.6, 突然変異確率を $1/m$ とした. また, ルーレット選択およびエリート戦略を用い, 100 世代の反復で終了とした.

7. 10 次元人工データセット Friedman を用いた基礎実験

提案法の高次元データでの有効性を確認するため, 2 クラス 10 次元の人工データセット”Friedman” を用いて実験を行った. このデータセットでは, 識別に有効な特徴は先頭の 4 特徴であり, 残りの 6 特徴は識別には有効ではない. クラス ω_1 のサンプルは平均 0 および単位共分散行列の 10 次元多変量正規分布に従って分布している. また, クラス ω_2 のサンプルの分布は, 最初の 4 特徴に関しては半径 3.5 と半径 4 の超球の間で一様分布し, 残りの 6 特徴に関してはクラス ω_1 のそれと同じ分布である.

実験結果を表 1 および図 2 に示す. サンプル数が少ない場合, 訓練誤差が高く, それ故, 提案法は多くの特徴を削除してしまう. それに対し, 中規模から大規模の場合には, 必要な四つの特徴のみ正しく選択されていること, つまり, 特徴選択に成功していることを確認できる.

表 1　Friedman データセットに対する分割結果

n	2^{d_1}	2^{d_2}	2^{d_3}	2^{d_4}	2^{d_5}	$\cdots$	$2^{d_{10}}$
100	1	1	4	1	1	1	1
500	4	4	4	4	1	1	1
1000	4	4	4	4	1	1	1
5000	4	4	4	4	1	1	1
10000	4	4	8	8	1	1	1
真の有効性	有効				非有効		

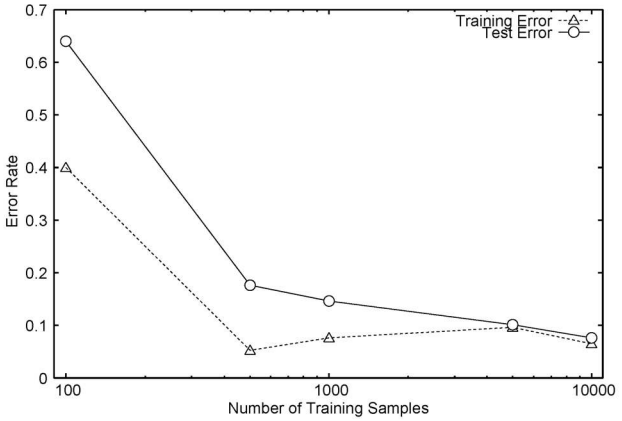


図 2　Friedman データセットに対する訓練および検査誤差

8. 局所特徴量と Bag of Visual Words モデルを用いた画像認識

画像分類の研究分野では, 画像中の特徴的な箇所をうまく取得できるような特徴量の研究がなされており, 局所特徴量によるアプローチが一般物体認識において大きな効果を示している. そこで本実験では, 局所特徴量を用いて画像から抽出した特徴を対象として, 本手法による特徴選択処理を行った.

局所特徴量とは, 画像に写っている対象全体ではなく, 対象の局所領域の情報を取得するのに用いられる特徴量である [14]. 従って, 対象全体の情報を取得する場合と比べ, 対象物体の情報を多く含まない特徴量であるといえる.

局所特徴量は, 画像の局所領域の輝度やエッジなどの情報をもとに構成され, 例えば顔画像であれば, 目や鼻, 口といった, 顔の特徴的な箇所に着目したものを自動的に選択できるという利点がある. また, 一つ一つは単純な情報であるが, それを組み合わせることによって, 対象を高次な情報で表現ができ, 画像全体に着目した特徴量と比較して, 部分的な特徴の変化, たとえば対象物が一部遮蔽されていたり, 照明条件が異なる場合などに対し頑健な特徴が得られる.

画像分類の分野では, 数ある局所特徴量のうち, 特に SIFT 特徴量 [15] が広く用いられている. 本実験では後段の特徴選択の有効性の確認を目的とすることから, 特徴量に関しては, 今回は, 関連研究と同様に SIFT 特徴量に限定して用いることとした.

SIFT(Scale-Invariant Feature Transform)[15] とは, 局所特徴量のひとつであり, 回転やスケーリングに対して不変であるという特長を持つ. SIFT の処理の流れは, キーポイント (特徴点) の検出と特徴量の記述の二段階に大きく分けられる.

キーポイントの検出では, ガウス関数を用いた平滑化画像の差分 (Difference-of-Gaussian 画像, 以下 DoG 画像) を用いて, キーポイントの候補点の検出と, そのスケールの探索を行う.

次に, 特徴量の記述を行う. まず, 特徴量を記述する領域を, キーポイントのオリエンテーション方向に回転する. 記述する領域は, キーポイントのスケールが内接する領域を用い, 図 3 の様に領域を 4×4 のブロックに均等に分割する. ブロックごとに, 前述のオリエンテーションを求める方法と同様にして, 8 方向の勾配方向ヒストグラムを算出する. 従って, 特徴は $4 \times 4 \times 8 = 128$ 次元の特徴ベクトルとなる. また, 特徴ベクトルは, 正規化するため照明条件に対して頑健性が得られるほか, キーポイントのオリエンテーション方向に領域を回転するため, 回転に不変な特徴量となる.

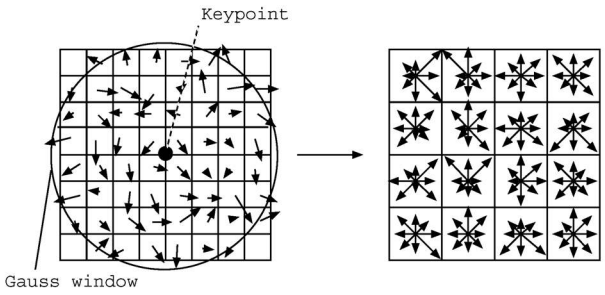


図 3　特徴量の記述

SIFT 特徴量は 128 次元の特徴ベクトルであるが, 画像中で検出されたキーポイントの数だけ記述されるため, ひとつの画像からは複数の特徴ベクトルが抽出されることになる. 従って, このままでは画像を一本のベクトルで記述できず, 各種の識別機を使用することが困難である.

そこで, 画像を局所特徴量の分布で表現したひとつのヒストグラムとして扱う Bag of Visual Words(以下, BoVW) [16] モデルを用いる. BoVW は, 画像から抽出した特徴量に対して以下の処理を行うことで作成する.

ベクトル量子化　図 4 の様に, 複数の画像から抽出された全ての特徴ベクトルを k 個のクラスタにクラスタリングする. このとき, 各クラスタの中心を Visual Word と呼ぶ. k は適当な定数をあらかじめ決めておくものとする.

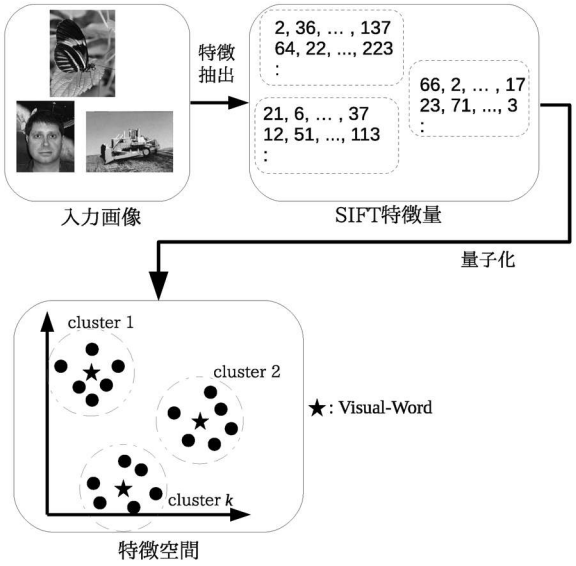


図 4　ベクトル量子化の概念図 (入力画像は Caltech256[17] より引用)

Visual Words の特定　各特徴ベクトルに対して, それが属する Visual Word を特定する. 各特徴ベクトルには, 属している Visual Word の ID が割り当てられることになる.

ヒストグラムの作成　画像ごとに, 各特徴ベクトルが属する Visual Word に投票し, ヒストグラムを作成する. 結果として, 画像をひとつのヒストグラムで表現することができる (図 5).

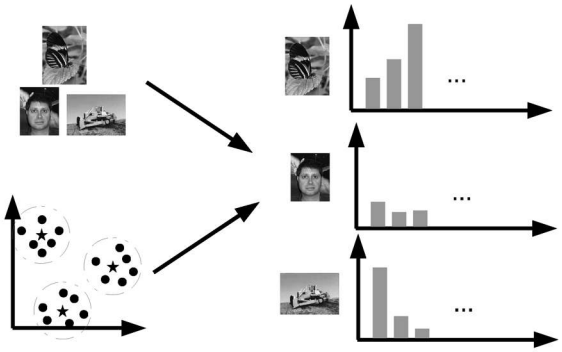


図 5　ヒストグラムの作成

BoVW による表現は, 画像中に写っている物体によく見られる特徴量をヒストグラムの頻度で表現するため, 異なる画像で同じ物体が写っている場合, 頻度の分布が似ると考えられる. また, BoVW は画像を局所特徴量の集合として扱うモデルであるため, 画像の特徴量の位置情報を扱わないという性質がある.

9. Bag of Visual Words モデル上での特徴選択実験

本手法を, 前節で述べた局所特徴量と Bag of Visual Words モデルによる画像認識問題に適用した. ここでクラス $\omega_1, \omega_2, \omega_3$ はそれぞれハサミ, USB メモリ, ステイプラーに対応する.

各クラスに著者らが撮影した 200 枚の画像を用い, その全体の半数を特徴選択および訓練に, 残りの半数を誤差の推定に用いた. 局所特徴量の算出には SIFT を用い, 生の特徴を Bag of Visual Words モデルを用いて特徴ベクトルに変換した. また, 準最適解な分割を見つけるためには GA を用いた. Visual Word の個数を 10 から 1000 まで変化させ, 各場合での誤差および選択された特徴の個数, つまり, $d_i \geq 1$ となる特徴の個数を調査した. その結果を表 (2) に示す.

結果として, 本手法では特徴選択より誤差が概ね 10% 程度上昇するが, 特徴の個数はそれにより 40% に減らすことができることを確認できた. これは本手法が Visual Word を重要なものとそうでないもの (ゴミ特徴) に分類できていることを意味する.

表 2　画像認識データセットに対する結果. m は Visual Word 数すなわち元の特徴数, ε は特徴選択前の誤差, $\widehat{\varepsilon}$ は特徴選択後の誤差, $\Delta\varepsilon$ は誤差の上昇量, $\widehat{m}$ は選択された特徴数, $\widehat{m}/m$ はその割合を示す.)

m	ε	$\widehat{\varepsilon}$	$\Delta\varepsilon$	$\widehat{m}$	$\widehat{m}/m$
10	0.143	0.263	0.120	4	0.400
50	0.117	0.203	0.086	10	0.200
100	0.120	0.227	0.107	20	0.200
500	0.060	0.123	0.063	241	0.482
1000	0.010	0.067	0.057	585	0.585
平均	0.090	0.177	0.087	—	0.373

10. おわりに

著者らが提案しているソフト特徴選択の手法の概略を紹介すると共に, 局所特徴量と Bag of Visual Words モデルを用いた画像認識への応用を試みた結果を報告した. 実験より, 本手法の特徴選択処理により誤差がある程度上昇することは避けられないものの, それにより特徴数を大きく削減できることを確認できた. これは本手法が Visual Word を重要なものとそうでないものに分類できていることを意味する.

今後は, 提案手法の特徴選択の原理面としては, 訓

練サンプル数が無限大へ向かう際の, ベイズの意味での最適な識別機への収束性の解析や, 本手法の MDL 基準のより詳細な検討を行いたい. また, 画像認識を含めた応用面の立場からは, 局所特徴量の種別との関連や Bag of Visual Words の様々な生成方法との関連など, 幅広い手法との関連性の調査を行いたい.

謝辞

本研究の一部は, 文部科学省科学研究費 (課題番号:24500228) の助成により行われた.

参考文献

[1] P. Somol, J. Novovicova and P. Pudil, Efficient Feature Subset Selection and Subset Size Optimization. *Pattern Recognition Recent Advances*, INTECH, 2010, 75–97.

[2] M. Kudo and T. Murai, Extended DNF Expression and Variable Granularity in Information Tables. *IEEE Transactions on Fuzzy Systems*, **16**, 2(2008), 285–298.

[3] Z. Pawlak, Decision Rules, Bayes' Rule and Rough Sets. New Directions in Rough Sets, Data Mining and Granular-Soft Computing. *Lecture Notes in Computer Science*, **1711**, 1999, 1–9.

[4] H. Tenmoto and M. Kudo, Soft Feature Selection by Using a Histogram-Based Classifier. *Lecture Notes in Computer Science, Advances in Pattern Recognition*, Springer, **5342**(2008), 572–581.

[5] M. Kudo and H. Tenmoto, Optimal Division for Feature Selection and Classification. *Proceedings of the International Workshop on Feature Selection for Data Mining: Interfacing Machine Learning and Statistics (FSDM'05)*, Newport Beach, California, USA, 2005, April, 106–107.

[6] J. Rissanen and B. Yu, MDL Learning. In D. W. Kueker and C. H. Smith, editors, *Learning and Geometry: Computational Approaches*, Birkhauser, 1998, 3–19.

[7] K. Yamanishi, A Learning Criterion for Stochastic Rules. *Proceedings of the Third Workshop on Computational Learning Theory*, 1990, 67–81.

[8] K. Yamanishi, Learning Non-Parametric Densities in Terms of Finite-Dimensional Parametric Hypotheses. *IEICE Transactions on Information and Systems*, E75-D, **4**(1992), 459–469.

[9] K. Yamanishi, Learning Non-Parametric Smooth Rules Using Stochastic Rules with Finite Partitioning. *Proceedings of the Computational Learning Theory: EuroCOLT'93*, 1993, 217–227.

[10] H. Tsuchiya, S. Itoh, and T. Mashimoto, An Algorithm for Designing a Pattern Classifier by Using MDL Criterion. *IEICE Transactions on Fundamentals*, E79-A, **6**(1996), 910–920.

[11] M. Kudo and J. Sklansky, Comparison of Algorithms that Select Features for Pattern Classifiers. *Pattern Recognition*, 33, **1**(2000), 25–41.

[12] J. Rissanen, Stochastic Complexity in Statistical Inquiry, *Volume 15 Series in Computer Science*, World Scientific, 1989.

[13] M. Kudo, S. Yanagi and M. Shimbo, Construction of Class Regions by a Randomized Algorithm: A Randomized Subclass Method. *Pattern Recognition*, **29**(1996), 581–588.

[14] 藤吉 弘亘, 山下 隆義, 岡田 和典, 前田 英作, V. Nozick, 石川 尋代, F. de Sorbier, コンピュータビジョン最先端ガイド 2. アドコム・メディア, 2010.

[15] D. G. Lowe, Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, Vol. 60, **No. 2**, 2004, 91–110.

[16] G. Csurka, C. R. Dance, L. Fan, J. Willamowski and C. Bray, Visual Categorization with Bags of Keypoints. *Proceedings of ECCV Workshop on Statistical Learning in Computer Vision*, 2004, 1–22.

[17] G. Griffin, Caltech 256 Image Dataset. http://www.vision.caltech.edu/Image_Datasets/Caltech256/.