Vol.2015-CVIM-196 No.13
2015/3/7

画像分類のための非一様マハラノビス符号化の提案

松澤　知己[1,a)]　武井 渉[1]　大町 真一郎[2]　加藤 毅[1]

概要：画像識別問題はコンピュータビジョン分野において中心的な課題の一つとして，多くの研究がなされてきた．画像識別においては SIFT などの局所記述子群から一つの特徴ベクトルをつくって統計的パターン認識の枠組みにあてはめる方法が定石となっている．中でも Bag-of-Visual-Words とフィッシャーベクトルが最も成功している．それらの符号化法に対し武井らは局所記述子の分布形状を考慮してマハラノビス計量を導入することが有効であることを示した．彼らの方法は記述子の空間全体で同一の計量を用いていたが，本発表では武井らの符号化を一般化して局所ごとに異なる非一様な計量を許すような符号化法を提案する．

キーワード：一般物体認識，Bag-of-Visual-Words 特徴，マハラノビス符号化，フィッシャーベクトル．

1. はじめに

画像のカテゴリ分類問題において，識別性能を左右する最も大きな因子は，学習機械の選択ではなく，画像表現の設計であるといっても過言ではない．画像の表現方法，すなわち特徴ベクトルの抽出方法として，すでに大まかなパイプラインは定まりつつあるといってもよく，パイプライン上の各ステップを如何に改善していくかが，現在のコンピュータビジョン分野において中心的な課題となっている．その定まりつつあるパイプラインは，符号化ステップとプーリングステップの２ステップからなる．このアプローチは，画像を局所パッチの集合とみて，その集合をＳＶＭなどの統計的学習機械に入力できるよう，固定次元の画像表現に変換する．このパイプラインは BoVW 法 [2] に端を発しており， 現在に至っても BoVW 法は最も有名な画像表現である．

この十年間で，符号化ステップとプーリングステップのパイプラインにおける各ステップに多くの改良手法が生まれた．符号化ステップに用いる局所パッチは，近年，画像から Dense に取得されることが多い．すなわち空間的かつ規則的に並べたパッチの集合を採用するアプローチが主流となりつつある [3]．一方，最近の文献においては Interest Point を使うアプローチ [4] も提案されている．符号化の方法には，これまで，BoVW 法の符号化ステップを担う**ヒストグラム符号化** *(Histogram Encoding)* [2], [5], [6]，ソフト割当 [7], [8], フィッシャー符号化 (Fisher Encoding) [3] やその亜種の VLAD [9]，Super Vector 符号化 [10]，Locality Constrained Linear Encoding [11] を含むスパース符号化 [11], [12], [13] といった多くのアプローチが提案されてきた．プーリングステップには，**平均プーリング** *(Average Pooling)* と最大プーリング (Max Pooling) の２種類がある．最大プーリングは主にスパース符号化 [11], [12], [13] に使われる．平均プーリングは，主に，ヒストグラム符号化，ソフト割当法，およびフィッシャー符号化で用いられる．プーリングの際に，空間情報を最終表現に混入するアプローチも検討されている [5], [13]．プーリング法は符号化法と組み合わせて提案されているが，その組み合わせは固定されておらず交換可能である [12]．Chatfield ら [14] は，様々な符号化法を細かく比較検討し，フィッシャー符号化が最も高性能であることを報告している．この事実から，本研究では最も有名な符号化であるヒストグラム符号化に加えて，フィッシャー符号化にも注目する．

正規混合分布によるフィッシャー符号化 [3] はヒストグラム符号化 [6] を一般化し，よりリッチな表現を許すようにしたものである．事実，ヒストグラム符号化は，近似的に，共通の等方性共分散行列を使った正規混合分布によるフィッシャー符号化の部分ベクトルになっていることを示すことができる．ヒストグラム符号化は大きな情報損失を生むことが知られている [15] が，フィッシャー符号化は，パッチ記述子の発生頻度のみならず，１次統計量と２次統計量をも符号化に組み入れることで，情報損失の問題をある程度解決している．フィッシャー符号化は，BoVW 法のコード

[1] 群馬大学理工学部，〒 376–8515 群馬県桐生市天神町 1–5–1
School of Science and Technology, Gunma University Tenjin-cho 1–5–1, Kiryu-shi, Gunma, 376–8515 Japan
[2] 東北大学工学研究科，〒 980–8579 仙台市青葉区荒牧字青葉 6–6–05
Graduate School of Engineering, Tohoku University, 6–6–05, Aramaki Aza Aoba, Aoba-ku, Sendai, 980–8579, Japan
[a)] matsuzawa-tomoki@kato-lab.cs.gunma-u.ac.jp

ブックに当たるものを任意の確率モデルで表すことができる拡張性の高いツールでもある．これまで画像認識の分野では，ほとんどの研究において，その確率モデルに対角共分散正規混合分布が採用されているが，それ以外の選択肢をとることもできる [16], [17].

従来の符号化法は，筆者らの知る限り，いずれの手法も，ユークリッド計量でパッチ記述子を符号化してきた．またコードブックの構築においても，同様にユークリッド計量が用いられている．BoVW 法でコードブックを構築する際は，K-平均法を用いるが，これはセントロイドと局所特徴との二乗ユークリッド距離の和が最小になるように Visual Word が求められる．フィッシャーベクトル法では，コードブックに相当する確率モデルのパラメータを推定するとき，ユークリッド計量空間上でのパッチ記述子の生起確率をモデル化している．これに対して，武井ら [18] は局所記述子の分布形状を考慮してマハラノビス計量を導入することが有効であることを示した．彼らの方法は記述子の空間全体で同一の計量を用いていたが，本論文では武井らの符号化を一般化して，局所ごとに異なる非一様な計量を許すような符号化法を提案する．さらに，２つのデータセットを用いた実験から，非一様なマハラノビス計量を用いることによって更なる識別性能の向上が得られることを示す．

2. 既存手法: ユークリッド符号化法

これまで物体認識や情景認識のために数多くの手法が提案されてきたが，本研究ではそれらの基礎になっており最も有名な BoVW 法と，それらの手法の中でも識別性能が最も高いとの報告があるフィッシャーベクトル (FV) 法に注目する．本節では，その２つの方法を紹介する．

2.1 Bag-of-Visual-Words (BoVW) 法

近年の物体認識や情景認識の多くの手法では，画像を画像パッチの集合とみて，これらを符号化プーリングを経て，固定次元の画像表現に変換し，高度に発展したＳＶＭなどの統計的学習機械の恩恵を受けられるようにしている．BoVW 法 [2] は，その礎を提供し，現在においてもよく用いられている手法である．BoVW 法における符号化とプーリングは次のように実装される：

- **符号化ステップ．** 画像パッチから**局所特徴** (e.g. *SIFT* [19]) を取り出し, その各局所特徴を**コードブック**中で最も近い一つの *Visual Word* に割り当てる．このような符号化は**ヒストグラム符号化**と呼ばれている [14].
- **プーリングステップ．**それぞれの Visual Word に割り当てられた局所特徴の頻度を数え，ヒストグラムを生成する．このようなプーリングは**平均プーリング**と呼ばれている [12].

符号化ステップでは，最も一般的な距離であるユークリッド距離を用いて一番近い Visual Word を選ぶ．ユークリッド距離計量は訓練段階においてコードブックを構築するときにも用いられる．すべての訓練用画像の局所特徴の集合，もしくは部分集合を K 平均法などを用いてクラスタリングすることで K 個の Visual Word を得る．

2.2 フィッシャーベクトル (FV) 法

フィッシャーベクトル (FV) 法はモデルパラメータ $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_m]^\top$ を持つ確率モデル $p(\boldsymbol{X}\,|\boldsymbol{\theta})$ から特徴抽出を行うフィッシャーカーネル [20] を基礎としていて，データ $\boldsymbol{X}$ 間の内積で定義されている．FV 法は 1 画像から得られる局所特徴の分布を反映しながら符号化される． 1 画像から得られる T 個の局所記述子集合 $\boldsymbol{X} := [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T]$ の FV は正規化勾配ベクトル

$$\boldsymbol{f}(\boldsymbol{X}) := \begin{bmatrix} \frac{\nabla_{\theta_1} \log p(\boldsymbol{X}|\boldsymbol{\theta})}{\sqrt{\mathcal{E}_{\boldsymbol{X}}((\nabla_{\theta_1} \log p(\boldsymbol{X}|\boldsymbol{\theta}))^2)}} \\ \vdots \\ \frac{\nabla_{\theta_m} \log p(\boldsymbol{X}|\boldsymbol{\theta})}{\sqrt{\mathcal{E}_{\boldsymbol{X}}((\nabla_{\theta_m} \log p(\boldsymbol{X}|\boldsymbol{\theta}))^2)}} \end{bmatrix}.$$

で定義される．画像認識においては，対角共分散による正規混合分布を確率モデル $p(\boldsymbol{X}|\boldsymbol{\theta})$ とするのが典型である．K 混合の分布の確率密度は

$$p(\boldsymbol{X}\,|\,\boldsymbol{\theta}) = \prod_{t=1}^{T} \sum_{k=1}^{K} \pi_k \mathcal{N}(\boldsymbol{x}_t\,;\,\boldsymbol{\mu}_k, \mathrm{diag}(\boldsymbol{\sigma}_k)^2)$$

のように表される．π_k は $\sum_k \pi_k = 1$ なる混合率，$\boldsymbol{\mu}_k \in \mathbb{R}^d$ は平均パラメータ，$\boldsymbol{\sigma}_k \in \mathbb{R}^d$ は標準偏差パラメータを表わす．この場合，モデルパラメータは

$$\boldsymbol{\theta} = \left[\pi_1, \ldots, \pi_K, \boldsymbol{\mu}_1^\top, \ldots, \boldsymbol{\mu}_K^\top, \boldsymbol{\sigma}_1^\top, \ldots, \boldsymbol{\sigma}_K^\top\right]^\top$$

なる $(2d+1)K$ 次元ベクトルで表される．よって，得られる FV も $(2d+1)K$ 次元ベクトルになる．このうち最初の K 次元，すなわち $\pi_1, \ldots, \pi_K$ に対応する部分ベクトルはヒストグラム符号化をソフト量子化に置き換えたものに等しいことを示すことができる．しかし，最近の多くの研究では，この最初の K 次元を FV から除外した $2dK$ 次元ベクトルが用いられていることから [21], [22], [23]，本研究でも $2dK$ 次元ベクトルを採用する．

FV 法 [24] では，正規化に必要な項 $\mathcal{E}_{\boldsymbol{X}}((\nabla_{\theta_i} \log p(\boldsymbol{X}|\boldsymbol{\theta}))^2)$ の値を得るため，次の２つの仮定をおいている：(a) 各画像パッチに対して，混合成分の負担率 (responsibility) [25] がほぼ一つの混合成分に集中している，(b) 画像パッチの個数 T が等しい．すると，FV は次のように与えられる：

$$\begin{aligned} &\boldsymbol{f}_{\mathrm{euc}}(\boldsymbol{X}) \\ &:= \left[\boldsymbol{f}_{\mathrm{euc}}^{\boldsymbol{\mu}_1}(\boldsymbol{X})^\top, \ldots, \boldsymbol{f}_{\mathrm{euc}}^{\boldsymbol{\mu}_K}(\boldsymbol{X})^\top, \boldsymbol{f}_{\mathrm{euc}}^{\boldsymbol{\sigma}_1}(\boldsymbol{X})^\top, \ldots, \boldsymbol{f}_{\mathrm{euc}}^{\boldsymbol{\sigma}_K}(\boldsymbol{X})^\top\right]^\top \end{aligned}$$

ただし，$k = 1, \ldots, K$ に対して，

$$\boldsymbol{f}_{\mathrm{euc}}^{\boldsymbol{\mu}_k}(\boldsymbol{X}) \approx \frac{1}{\sqrt{T\pi_k}}\mathrm{diag}(\boldsymbol{\sigma}_k)^{-1}\boldsymbol{Y}_{k,\mathrm{euc}}\boldsymbol{\gamma}_{k,\mathrm{euc}},$$
$$\boldsymbol{f}_{\mathrm{euc}}^{\boldsymbol{\sigma}_k}(\boldsymbol{X}) \approx \frac{1}{\sqrt{2T\pi_k}}\left(\mathrm{diag}(\boldsymbol{\sigma}_k)^{-2}\boldsymbol{Y}_{k,\mathrm{euc}} \odot \boldsymbol{Y}_{k,\mathrm{euc}} - \mathbf{1}_d\mathbf{1}_T^\top\right)\boldsymbol{\gamma}_{k,\mathrm{euc}}$$

のように近似的に与えられる．このとき $\odot$ は要素ごとの積を表し，また $\boldsymbol{Y}_{k,\mathrm{euc}} := \boldsymbol{X} - \boldsymbol{\mu}_k\mathbf{1}_T^\top$ と定義した．ベクトル $\boldsymbol{\gamma}_{k,\mathrm{euc}} \in \mathbb{R}^T$ は k 番目の混合成分に対する負担率であり，$\boldsymbol{\gamma}_{k,\mathrm{euc}}$ の第 t 要素は

$$\frac{\pi_k\mathcal{N}(\boldsymbol{x}_t\,;\,\boldsymbol{\mu}_k,\mathrm{diag}(\boldsymbol{\sigma}_k)^2)}{\sum_{k'}\pi_{k'}\mathcal{N}(\boldsymbol{x}_t\,;\,\boldsymbol{\mu}_{k'},\mathrm{diag}(\boldsymbol{\sigma}_{k'})^2)}.$$

で与えられる．

BoVW 法と FV 法を対比してみると，ひとつの混合成分 $\mathcal{N}(\boldsymbol{x}_t\,;\,\boldsymbol{\mu}_k,\mathrm{diag}(\boldsymbol{\sigma}_k)^2)$ が一つの Visual Word に対応していると考えることができる．正規混合分布による FV を用いるとき，分散のパラメータを変えることによって計量を変化させているという見方もできなくはない．ただし，これを認めたとしても対角共分散による正規分布の性質から，計量がスケーリングされる方向は元の軸に制限されたままである．

3. 提案手法：マハラノビス符号化法

本研究では，BoVW と FV に**マハラノビス計量** [1], [26] の考え方を導入する．2 つのベクトル $\boldsymbol{x}, \boldsymbol{x}' \in \mathbb{R}^d$ 間のマハラノビス距離は

$$D(\boldsymbol{x}, \boldsymbol{x}'; \boldsymbol{A}) := \sqrt{(\boldsymbol{x} - \boldsymbol{x}')^\top \boldsymbol{A}(\boldsymbol{x} - \boldsymbol{x}')}$$

と表される．ただし，$\boldsymbol{A}$ は**マハラノビス行列**と呼ばれる正定値行列である．ユークリッド計量では，ある点から等距離の点の集合は円 (もしくは球，超球) になるが，マハラノビス計量の場合は楕円 (もしくは楕球，超楕球) になることが知られている．

3.1 マハラノビス BoVW 法

武井ら [18] は，BoWV 法の符号化ステップにおいてユークリッド距離の代わりにマハラノビス距離を採用し，マハラノビス距離を用いた符号化を**マハラノビス符号化**と名付けた．さらに武井らはコードブック構築の際もマハラノビス計量を用いた．K 個の Visual Word を含むコードブック $\boldsymbol{V} := [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K] \in \mathbb{R}^{d\times K}$ を得るために，目的関数

$$J_{\text{cb-ho}}(\boldsymbol{V}; \boldsymbol{A}) := \sum_{t=1}^{T} \min_{k_t\in\{1,\ldots,K\}} D(\boldsymbol{x}_t, \boldsymbol{v}_{k_t}; \boldsymbol{A})$$

を K 平均法のような Greedy な方法で最小化することによって求めた．このときの T はコードブック構築に用いる局所特徴の個数である．このように 1 つのマハラノビス行列を用いて符号化された BoVW を**一様マハラノビス** *Bag-of-Visual-Words*(HoMahaBoVW) と呼ぶこととする．これは d 次元の局所特徴空間全域で共通のマハラノビス計量を用いているとみなすことができる．これに対して，局所ごとに異なる計量を用いる方法も考えられる．本研究では，それぞれの Visual Word $\boldsymbol{v}_k$ ごとに異なるマハラノビス行列 $\boldsymbol{A}_k$ を許す方法を提案する．このように符号化された BoVW を**非一様マハラノビス** *Bag-of-Visual-Words*(HeMahaBoVW) と呼ぶこととする．HeMahaBoVW では局所特徴 $\boldsymbol{x}$ は

$$k_\star \in \operatorname*{argmin}_{k\in\{1,\ldots,K\}} D(\boldsymbol{x}, \boldsymbol{v}_k; \boldsymbol{A}_k).$$

によって 1 つの Visual Word $\boldsymbol{v}_{k_\star}$ に量子化される．

3.2 マハラノビス FV 法

武井ら [18] はさらに FV のフレームワークに対してもマハラノビス計量を組み込んでいる．マハラノビス行列の固有値分解 $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top$ から，**白色化行列** $\boldsymbol{W} := \boldsymbol{\Lambda}^{1/2}\boldsymbol{U}^\top$ を得る．各点 $\boldsymbol{x}$ に白色化行列をかけると，任意の 2 点間の計量はユークリッド計量として計算できる (i.e. $D(\boldsymbol{x}, \boldsymbol{x}'; \boldsymbol{A}) = \|\boldsymbol{W}\boldsymbol{x} - \boldsymbol{W}\boldsymbol{x}'\|$)．この性質に基づいて武井らは白色化した局所特徴を正規混合分布に導入した．すなわち，武井ら [18] は次の確率モデルを用いた：

$$p_{\mathrm{ho}}(\boldsymbol{X}|\boldsymbol{\theta}) := |\det(\boldsymbol{W})|^T \prod_{t=1}^{T}\sum_{k=1}^{K} \pi_k\mathcal{N}(\boldsymbol{W}\boldsymbol{x}_t\,;\,\boldsymbol{\mu}_k,\mathrm{diag}(\boldsymbol{\sigma}_k)^2).$$

この確率モデルから導出される FV $\boldsymbol{f}_{\mathrm{ho}}(\boldsymbol{X})$ を**一様マハラノビスフィッシャーベクトル** (HoMahaFV) と呼ぶこととする．HoMahaFV は

$$\boldsymbol{f}_{\mathrm{ho}}(\boldsymbol{X}) := \left[\boldsymbol{f}_{\mathrm{ho}}^{\boldsymbol{\mu}_1}(\boldsymbol{X})^\top, \ldots, \boldsymbol{f}_{\mathrm{ho}}^{\boldsymbol{\mu}_K}(\boldsymbol{X})^\top, \boldsymbol{f}_{\mathrm{ho}}^{\boldsymbol{\sigma}_1}(\boldsymbol{X})^\top, \ldots, \boldsymbol{f}_{\mathrm{ho}}^{\boldsymbol{\sigma}_K}(\boldsymbol{X})^\top\right]^\top$$

で表わすとする．正規化項 $\mathcal{E}_{\boldsymbol{X}}((\nabla_{\theta_i}\log p_{\mathrm{ho}}(\boldsymbol{X}|\boldsymbol{\theta}))^2)$ の算出に，オリジナルの FV でも用いられていたものと同じ仮定 (2.2 節参照) をおくと，各部分ベクトルは，

$$\boldsymbol{f}_{\mathrm{ho}}^{\boldsymbol{\mu}_k}(\boldsymbol{X}) \approx \frac{1}{\sqrt{T\pi_k}}\mathrm{diag}(\boldsymbol{\sigma}_k)^{-1}\boldsymbol{Y}_k\boldsymbol{\gamma}_k,$$
$$\boldsymbol{f}_{\mathrm{ho}}^{\boldsymbol{\sigma}_k}(\boldsymbol{X}) \approx \frac{1}{\sqrt{2T\pi_k}}\left(\mathrm{diag}(\boldsymbol{\sigma}_k)^{-2}\boldsymbol{Y}_k \odot \boldsymbol{Y}_k - \mathbf{1}_d\mathbf{1}_T^\top\right)\boldsymbol{\gamma}_k$$

と近似できる．ただし，$\boldsymbol{Y}_k := \boldsymbol{W}\boldsymbol{X} - \boldsymbol{\mu}_k\mathbf{1}_T^\top$ とおいた．ベクトル $\boldsymbol{\gamma}_k \in \mathbb{R}^T$ の第 t 要素は

$$\frac{\pi_k\mathcal{N}(\boldsymbol{W}\boldsymbol{x}_t\,;\,\boldsymbol{\mu}_k,\mathrm{diag}(\boldsymbol{\sigma}_k)^2)}{\sum_{k'}\pi_{k'}\mathcal{N}(\boldsymbol{W}\boldsymbol{x}_t\,;\,\boldsymbol{\mu}_{k'},\mathrm{diag}(\boldsymbol{\sigma}_{k'})^2)}.$$

で与えた．

このように武井らによる HoMahaFV は共通の白色化行列 $\boldsymbol{W}$ を用いている．本研究ではさらに，局所ごとに異なる計量を HoMahaFV に対しても導入する**非一様マハラノビスフィッシャーベクトル** (HeMahaFV) を提案する．この新しい FV 法では，混合成分ごとに異なる白色化行列 $\boldsymbol{W}_1, \ldots, \boldsymbol{W}_K$ を許しており，以下のような確率モデルを用いる：

$$p_{\mathrm{he}}(\boldsymbol{X}|\boldsymbol{\theta}) := \prod_{t=1}^{T}\sum_{k=1}^{K}\pi_k|\det(\boldsymbol{W}_k)|\mathcal{N}(\boldsymbol{W}_k\boldsymbol{x}_t\,;\,\boldsymbol{\mu}_k,\mathrm{diag}(\boldsymbol{\sigma}_k)^2).$$

$\boldsymbol{W}_k$ は k 番目の混合成分から得られたマハラノビス行列 $\boldsymbol{A}_k$ を固有値分解することによって計算される．$\boldsymbol{Y}'_k := \boldsymbol{W}_k\boldsymbol{X} - \boldsymbol{\mu}_k\boldsymbol{1}_T^\top$ としたとき，確率密度 $p_{\mathrm{he}}(\boldsymbol{X}|\boldsymbol{\theta})$ から $\boldsymbol{\mu}_k$，$\boldsymbol{\sigma}_k$ それぞれに対する FV の近似式は

$$\boldsymbol{f}_{\mathrm{he}}^{\boldsymbol{\mu}_k}(\boldsymbol{X}) \approx \frac{1}{\sqrt{T\pi_k}}\mathrm{diag}(\boldsymbol{\sigma}_k)^{-1}\boldsymbol{Y}'_k\boldsymbol{\gamma}'_k,$$
$$\boldsymbol{f}_{\mathrm{he}}^{\boldsymbol{\sigma}_k}(\boldsymbol{X}) \approx \frac{1}{\sqrt{2T\pi_k}}\left(\mathrm{diag}(\boldsymbol{\sigma}_k)^{-2}\boldsymbol{Y}'_k \odot \boldsymbol{Y}'_k - \boldsymbol{1}_d\boldsymbol{1}_T^\top\right)\boldsymbol{\gamma}'_k$$

と導出される．ただし，T 次元ベクトル $\boldsymbol{\gamma}'_k$ の第 t 要素は

$$\frac{\pi_k|\det(\boldsymbol{W}_k)|\mathcal{N}(\boldsymbol{W}_k\boldsymbol{x}_t\,;\,\boldsymbol{\mu}_k,\mathrm{diag}(\boldsymbol{\sigma}_k)^2)}{\sum_{k'}\pi_{k'}|\det(\boldsymbol{W}_{k'})|\mathcal{N}(\boldsymbol{W}_{k'}\boldsymbol{x}_t\,;\,\boldsymbol{\mu}_{k'},\mathrm{diag}(\boldsymbol{\sigma}_{k'})^2)}.$$

で与えた．

4. 実装の詳細

マハラノビス符号化を画像分類に適用するには，事前にマハラノビス行列を求めておく必要がある．加えて，HeMahaFV 法 については確率モデルのパラメータの値も決めておかなくてはならない．本節では，提案法におけるマハラノビス行列とパラメータの決め方について述べる．

4.1 HeMahaBoVW 法

提案する HeMahaBoVW 法では，K 個の Visual Word それぞれに異なるマハラノビス行列 $\boldsymbol{A}_1,\ldots,\boldsymbol{A}_K$ を必要とする．HeMahaBoVW 法は，広大な局所特徴空間に対して用いる共通な計量よりも，異なる局所領域ごとに適した計量のほうが存在する，という期待に基づいている．しかしながら，マハラノビス行列の柔軟性のため，訓練データを使って何の制約もなしにマハラノビス行列のフィッティングさせてしまうと，過整合をおこすかもしれない．このような考えから，本研究では 1 つのクラスに属する局所特徴の分布に対し同じ計量を共有することとする．クラス c で用いるマハラノビス行列 $\boldsymbol{A}_{(c)}$ を得るために，クラス c に属する画像パッチの特徴の集合から共分散行列を計算し，その逆行列で $\boldsymbol{A}_{(c)}$ を与えることとした．この際，共分散行列の非特異性を保証するために，共分散行列の対角成分に小さな正の定数を加算する．この小さな正の定数には共分散行列の最大固有値の 0.05 倍を用いた．Visual Word の集合をクラスごとに計算した後，それらを統合して，ひとつのコードブックを得る．最後に，訓練データ全体から計算された Visual Word を結合する．

4.2 HeMahaFV 法

提案する HeMahaFV 法では，クラスを考慮した方法で求められる $\frac{K}{2}$ 混合の正規混合成分と，訓練データ全体から求められる $\frac{K}{2}$ 混合の正規混合成分を最終的に組み合わせることで K 混合の正規混合分布とする．まず $\frac{K}{2}$ 個の混合成分をそれぞれのクラスで小さな正規混合分布を持つように，ほぼ均等に割り振る．このとき，HeMahaBoVW 法と同様にして，クラス内で共通のマハラノビス行列 $\boldsymbol{A}_{(c)}$ を考え，各クラスの混合成分に対応する白色化行列 $\boldsymbol{W}_{(c)}$ を与える．クラスごとにモデルパラメータの推定を行ったのち，それぞれのクラスに割り振られた正規混合成分を再び混合する．さらに，訓練データ全体に対しても $\frac{K}{2}$ 混合の正規混合分布を求め，先に求めた $\frac{K}{2}$ 混合の正規混合成分と組み合わせることで最終的な K 混合の正規混合分布 $p_{\mathrm{he}}(\boldsymbol{X}\,|\,\boldsymbol{\theta})$ とする．

クラスごとの正規混合分布のパラメータはそのクラスに属するデータのみを使って最尤推定する． 最尤推定には，尤度関数を Greedy に最大化する EM アルゴリズムを用いる．EM アルゴリズムは E-step と M-step を収束するまで繰り返す反復法である．クラス c の正規混合分布が K' 個の混合成分を持っているとして，その分布のパラメータ $\boldsymbol{\mu}_{1,c},\ldots,\boldsymbol{\mu}_{K',c} \in \mathbb{R}^d$, $\boldsymbol{\sigma}_{1,c},\ldots,\boldsymbol{\sigma}_{K',c} \in \mathbb{R}^d$, $\boldsymbol{\pi}_c \in \mathbb{R}^{K'}$ の最尤推定値を求めるために，T 個の局所特徴ベクトル $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_T$ を使うとする．その尤度関数は次のようにあらわされる：

$$\mathcal{L}(\boldsymbol{\theta}_{(c)}) := |\det(\boldsymbol{W}_{(c)})|^T\prod_{t=1}^{T}\sum_{k=1}^{K'}\pi_{k,c}\mathcal{N}(\boldsymbol{W}_{(c)}\boldsymbol{x}_t\,;\,\boldsymbol{\mu}_{k,c},\mathrm{diag}(\boldsymbol{\sigma}_{k,c})^2)$$

ただし，$\boldsymbol{\theta}_{(c)} := \{\boldsymbol{\mu}_{1,c},\ldots,\boldsymbol{\mu}_{K',c},\boldsymbol{\sigma}_{1,c},\ldots,\boldsymbol{\sigma}_{K',c},\boldsymbol{\pi}_c\}$ とおいた．この尤度関数を最大化するための EM アルゴリズムは次のようになる．

Algorithm 1 EM Algorithm for HeMahaFV Method

Input: Observation $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_T \in \mathbb{R}^d$ and initial values $\boldsymbol{\mu}^{(0)}_{1,c},\ldots,\boldsymbol{\mu}^{(0)}_{K',c} \in \mathbb{R}^d, \boldsymbol{\sigma}^{(0)}_{1,c},\ldots,\boldsymbol{\sigma}^{(0)}_{K',c} \in \mathbb{R}^d, \boldsymbol{\pi}_c \in \mathbb{R}^{K'}$

1: **for** $l = 1, 2, \ldots$ **do**

2: **E step** Compute

$$\gamma^{(l)}_{t,k} := \frac{\pi^{(l-1)}_{k,c}\mathcal{N}(\boldsymbol{W}_{(c)}\boldsymbol{x}_t\,;\,\boldsymbol{\mu}^{(l-1)}_{k,c},\mathrm{diag}(\boldsymbol{\sigma}^{(l-1)}_{k,c})^2)}{\sum_{k'}\pi^{(l-1)}_{k',c}\mathcal{N}(\boldsymbol{W}_{(c)}\boldsymbol{x}_t\,;\,\boldsymbol{\mu}^{(l-1)}_{k',c},\mathrm{diag}(\boldsymbol{\sigma}^{(l-1)}_{k',c})^2)}.$$

for $(t,\,k) \in \{1,\ldots,T\}\times\{1,\ldots,K'\}$;

3: **M step** Update the parameter values by

$$\boldsymbol{\mu}^{(l)}_{k,c} := \frac{\sum_t\gamma^{(l)}_{t,k}\boldsymbol{W}_{(c)}\boldsymbol{x}_t}{\sum_{t'}\gamma^{(l)}_{t',k}}.$$

$$(\boldsymbol{\sigma}^{(l)}_{k,c})^2 := \frac{\sum_t\gamma^{(l)}_{t,k}((\boldsymbol{W}_{(c)}\boldsymbol{x}_t - \boldsymbol{\mu}^{(l)}_{k,c}) \odot (\boldsymbol{W}_{(c)}\boldsymbol{x}_t - \boldsymbol{\mu}^{(l)}_{k,c}))}{\sum_{t'}\gamma^{(l)}_{t',k}}.$$

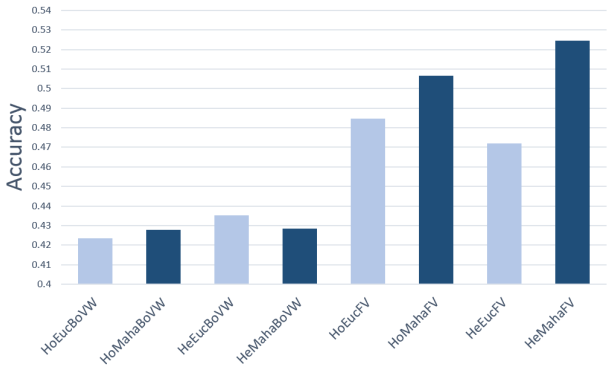
$$\pi^{(l)}_{k,c} := \frac{1}{T}\sum_t\gamma^{(l)}_{t,k}.$$

for $k = 1,\ldots,K'$;

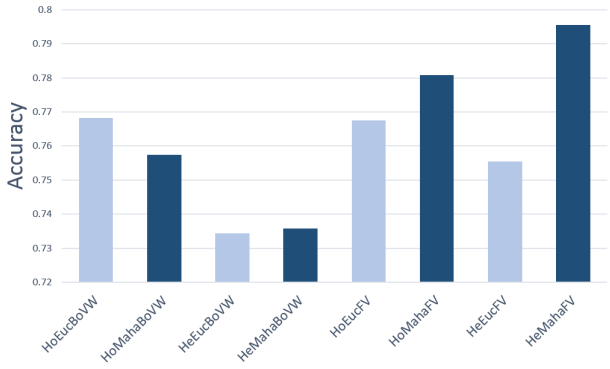
4: **end for**

5. 実験

提案法 HeMahaBoVW, HeMahaFV の性能を評価するため，次の 6 つの方法と比較する：



(a) FMD



(b) LSP15

図 1　Categorization Performance.

- 従来のユークリッド距離による BoVW，FV (HoEucBoVW, HoEucFV)
- HeMahaBoVW と同様な方法でコードブックを求めるが計量行列はすべて単位行列にしたもの (HeEucBoVW)
- HeMahaFV と同様な方法で $\boldsymbol{\theta}$ を求めるが計量行列はすべて単位行列にしたもの (HeEucFV)
- 武井ら [18] の方法 (HoMahaBoVW, HoMahaFV)

5.1　実験条件

画像から局所特徴を抽出するステップにおいて 3 画素間隔にキーポイントを設置した Dense SIFT を使用する．Visual Word の個数については，BoVW では $K = 1024$，FV では $K = 256$ とした．FV に対する正規化には Power 正規化，L2 正規化 [27] を用いた．ただし HeEucFV と HeMahaFV に対しては，次の L2 正規化を行った．その方法とは，一度クラスごとで正規化しその後全体を再度正規化するというものである．2 つのデータセット，Flickr Material Database(FMD)[28]，LSP15[5] 上で，One-vs-rest SVM による多クラス識別を行った．FMD は Flickr.com で集められた多種多様な画像を含んでいるデータセットであり，10 カテゴリかつ各クラスにつき 100 枚の画像から構成されている．クラス数は小規模だが，クラス内分散が非常に大きいデータセットであるため，クラス毎の共分散行列から計算されたマハラノビス行列を用いることで，符号化ステップでのより良いクラスタリングが期待できる．LSP15 は 15 クラスのシーン画像のデータセットで, 各クラスに 200 枚以上の画像がある．両データセット共に訓練用画像，評価用画像を各クラス 30 枚ずつとし，各画像を無作為に選出しながら 10 回実験を行いそれらの平均識別率を報告する．

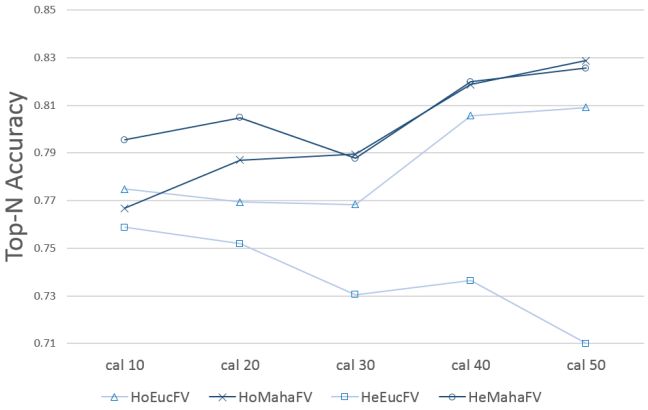


図 2　クラス数に伴う性能の変化

5.2　手法間の性能比較

図 1 に FMD および LSP15 上での実験結果を示す．実験の結果，BoVW については提案手法による効果は得られなかった．しかし，FV については，両データセット共に，提案手法である HeMahaFV によって顕著な性能向上が得られることを確認した．

5.3　クラス数に伴う性能の変化

上記の実験結果はいずれもクラス数が少ないデータセットを用いていた．ここでは，識別対象のクラス数を変動させたとき，各手法の性能がどのように変化するのかを示す．実験に伴い，我々は 101 クラスの物体画像データセットである CalTech 101(Cal 101)[29] を基にした独自のデータセット (Cal 10，Cal 20，... ，Cal 50) を作成した．その独自のデータセット (Cal 10，Cal 20，... ，Cal 50) の作成方法は，Cal 101 の先頭クラスから 10 クラス，20 クラスと 10 クラス間隔で最大 50 クラスまで取っていき，それぞれをデータセットとするというものである．この Cal 101 から派生した各データセットに対する実験から得られた Top-N Accuracy を図 2 に示す．このとき，N はクラス数の 10 分の 1 とした．図 2 から，提案手法である HeMahaFV の性能は識別対象のクラス数に関わらず，通常の FV(EucHoFV) と比べて良くなることがわかった．

6.　おわりに

本論文では，画像識別に用いられる BoVW，FV に対して画像のクラスを考慮したマハラノビス計量を組み込む手法 HeMahaBoVW，HeMahaFV を提案した．2 個の公開データセットを使った多クラス認識実験を行った結果，HeMa-

haBoVW については効果は得られなかったが，HeMahaFV については既存の方法に比べて識別率の向上を確認した．さらにクラス数を変動させて実験を行ったが，既存手法に対する HeMahaFV の有効性はクラス数に依存しないということがわかった．

本研究では，2 つの標準的な方法，BoVW 法およびフィッシャーベクトル法，に限定して，マハラノビス計量の効果を検証したが，コンピュータビジョンのコミュニティでは，さまざまな視点から数多くの改良が試みられてきたことは 1 節で述べた通りである．たとえば，背景雑音 (background clutter) の影響を軽減するために，Ramazan ら [22] は，物体切り出しの仮説を大量に作った上で，背景に当たる部分の重みを小さくし，物体である可能性が大きい部分の重みが大きくなるようにフィッシャーベクトルを作っている．Fraz ら [4] は，パッチごとにフィッシャーベクトルを計算する Mid-Level 表現を提案している．これらにマハラノビス計量を適用しても識別性能のさらなる向上が得られるかもしれない．BoVW 法およびフィッシャーベクトル法はパッチ特徴がほぼ 1 個の Visual Word で近似できるという仮定に基づいているが，スパース符号化法 [11], [12], [13] のような Visual Word を基底とみて，個々のパッチをその成分で表す方法も注目されている．その方法でも，パッチ特徴と再構築特徴とのユークリッド距離を最小にするように成分が求められるが，再構築誤差をユークリッド計量で測るのが最適かどうかは明らかではない．このようにマハラノビス符号化の導入は，物体認識や情景認識のためにこれまで開発されてきた数多くの方法をさらに改良できる可能性を残しており，マハラノビス符号化の考え方を新たな軸とした物体認識および情景認識のさらなる発展が期待される．

参考文献

[1] Kato, T., Takei, W. and Omachi, S.: A Discriminative Metric Learning Algorithm for Face Recognition, *IPSJ Transactions on Computer Vision and Applications, Presented at MIRU2013 as Oral Presentation*, Vol. 5, pp. 85–89 (2013).

[2] Csurka, G., Dance, C., Fan, L., Willamowski, J. and Bray, C.: Visual categorization with bags of keypoints, *Workshop on statistical learning in computer vision, ECCV*, Vol. 1, p. 22 (2004).

[3] Sánchez, J., Perronnin, F., Mensink, T. and Verbeek, J.: Image classification with the Fisher vector: Theory and practice, *International journal of computer vision*, Vol. 105, No. 3, pp. 222–245 (2013).

[4] Fraz, M., Edirisinghe, E. A. and Sarfraz, M. S.: Mid-level-Representation Based Lexicon for Vehicle Make and Model Recognition, *Pattern Recognition (ICPR), 2014 22nd International Conference on*, IEEE, pp. 393–398 (2014).

[5] Lazebnik, S., Schmid, C. and Ponce, J.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories, *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, Vol. 2, IEEE, pp. 2169–2178 (2006).

[6] Sivic, J. and Zisserman, A.: Efficient Visual Search of Videos Cast as Text Retrieval, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 31, No. 4, pp. 591–606 (2009).

[7] Farquhar, J., Szedmak, S., Meng, H. and Shawe-Taylor, J.: Improving “bag-of-keypoints” image categorisation: Generative models and pdf-kernels (2005).

[8] Winn, J., Criminisi, A. and Minka, T.: Object categorization by learned universal visual dictionary, *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, Vol. 2, IEEE, pp. 1800–1807 (2005).

[9] Jégou, H., Perronnin, F., Douze, M., Sánchez, J., Pérez, P. and Schmid, C.: Aggregating local image descriptors into compact codes, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 34, No. 9, pp. 1704–1716 (2012).

[10] Zhou, X., Yu, K., Zhang, T. and Huang, T. S.: Image classification using super-vector coding of local image descriptors, *Computer Vision–ECCV 2010*, Springer, pp. 141–154 (2010).

[11] Wang, J., Yang, J., Yu, K., Lv, F., Huang, T. and Gong, Y.: Locality-constrained linear coding for image classification, *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, IEEE, pp. 3360–3367 (2010).

[12] Boureau, Y.-L., Bach, F., LeCun, Y. and Ponce, J.: Learning mid-level features for recognition, *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, IEEE, pp. 2559–2566 (2010).

[13] Yang, J., Yu, K., Gong, Y. and Huang, T.: Linear spatial pyramid matching using sparse coding for image classification, *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, IEEE, pp. 1794–1801 (2009).

[14] Chatfield, K., Lempitsky, V., Vedaldi, A. and Zisserman, A.: The devil is in the details: an evaluation of recent feature encoding methods (2011).

[15] Boiman, O., Shechtman, E. and Irani, M.: In defense of nearest-neighbor based image classification, *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, IEEE, pp. 1–8 (2008).

[16] Cinbis, R. G., Verbeek, J. and Schmid, C.: Image categorization using Fisher kernels of non-iid image models, *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, IEEE, pp. 2184–2191 (2012).

[17] Tanaka, M., Torii, A. and Okutomi, M.: Fisher Vector based on Full-covariance Gaussian Mixture Model, IPSJ Transactions on Computer Vision and Applications (CVA)， Vol. 5, pp. 50–54 (2013).

[18] 武井　渉，細堀勝也，加藤　毅，大町真一郎：画像認識のためのマハラノビス符号化法の提案，電子情報通信学会技術報告書. PRMU，Vol. 113, No. 403, pp. 201–206 (2014).

[19] Lowe, D. G.: Distinctive image features from scale-invariant keypoints, *International journal of computer vision*, Vol. 60, No. 2, pp. 91–110 (2004).

[20] Jaakkola, T., Haussler, D. et al.: Exploiting generative models in discriminative classifiers, *Advances in neural information processing systems*, pp. 487–493 (1999).

[21] Ji, Z.: Decoupling Sparse Coding with Fusion of Fisher Vectors and Scalable SVMs for Large-Scale Visual Recognition, *Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on*, IEEE, pp. 450–457 (2013).

[22] Cinbis, R. G., Verbeek, J. and Schmid, C.: Segmentation driven object detection with Fisher vectors, *Computer Vi-*

sion (ICCV), 2013 IEEE International Conference on, pp. 2968–2975 (2013).

[23] Sydorov, V., Sakurada, M. and Lampert, C. H.: Deep Fisher Kernels–End to End Learning of the Fisher Kernel GMM Parameters.

[24] Perronnin, F. and Dance, C.: Fisher kernels on visual vocabularies for image categorization, *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, IEEE, pp. 1–8 (2007).

[25] Bishop, C. M.: *Pattern Recognition and Machine Learning*, Springer Science+Business Media, LLC, New York, USA (2006).

[26] Weinberger, K. Q. and Saul, L. K.: Distance Metric Learning for Large Margin Nearest Neighbor Classification, *J. Mach. Learn. Res.*, Vol. 10, pp. 207–244 (online), available from ⟨http://dl.acm.org/citation.cfm?id=1577069.1577078⟩ (2009).

[27] Perronnin, F., Sanchez, J. and Mensink, T.: Improving the fisher kernel for large-scale image classification, *Computer Vision–ECCV 2010*, Springer, pp. 143–156 (2010).

[28] Sharan, L., Rosenholtz, R. and Adelson, E.: Material perception: What can you see in a brief glance?, *Journal of Vision*, Vol. 9, No. 8, pp. 784–784 (2009).

[29] Fei-Fei, L., Fergus, R. and Perona, P.: Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories, *Computer Vision and Image Understanding*, Vol. 106, No. 1, pp. 59–70 (2007).