3R-06

画像認識のためのマハラノビスフィッシャー符号化法の提案

松澤　知己 †　　共著　武井　渉 ‡　　加藤　毅 ‡

† 群馬大学 工学部　　‡ 群馬大学 理工学部

1　はじめに

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 などの局所記述子を符号化し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．符号化の中でも**フィッシャーベクトル** [1] が最も成功している．そのフィッシャーベクトルに対し武井ら [2] 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，本研究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によって，非一様なマハラノビス計量により更なる識別性能の向上が得られることを示す．

2　フィッシャーベクトル

フィッシャーベクトル (以下，FV) は，データ $\boldsymbol{X}$ から確率モデル $p(\boldsymbol{X}|\boldsymbol{\theta})$ を介して，特徴抽出を行うフィッシャーカーネルを基盤とし，画像認識の分野では，Bag-of-Visual-Words 法に代わるベクトル量子化手法として広く使われている．画像認識においては，ひとつの画像から大量の局所記述子を得ることができるが，FV はこれらの分布の統計値を得るための一つの手段となっている．モデルパラメータ $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_m]^\top$ に対し，一画像から得られる局所記述子集合 $\boldsymbol{X}$ の FV は正規化勾配ベクトル

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

で定義される．画像認識においては，確率モデルとして，対角共分散による正規混合分布が専ら用いられている．$\boldsymbol{X}$ に T 個の局所記述子 $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T \in \mathbb{R}^d$ を含む時，その密度は

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

と表される．ただし，K 混合のモデルとし，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，$\boldsymbol{\sigma}_k \in \mathbb{R}^d$ は分散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．このうち最初の K 次元，すなわち $\pi_1, \ldots, \pi_K$ に対応する次元を FV から除外した $2dK$ 次元ベクトルがしばしば用いられている．

Bag-of-Visual-Words 法と FV 法を対比してみると，ひとつの混合成分 $\mathcal{N}(\boldsymbol{x}_t\,;\, \boldsymbol{\mu}_k, \mathrm{diag}(\boldsymbol{\sigma}_k))$ が一つの Visual Word に対応していると考えることができる．

3　提案法

本研究では，FV に**マハラノビス距離**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

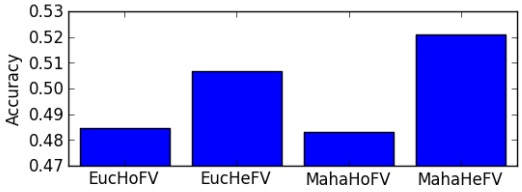
$$D(\boldsymbol{x}, \boldsymbol{x}'; \boldsymbol{A}) := \sqrt{(\boldsymbol{x} - \boldsymbol{x}')^\top \boldsymbol{A} (\boldsymbol{x} - \boldsymbol{x}')}$$

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．マハラノビス行列の固有値分解を $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top$ から，**白色化行列** $\boldsymbol{W} := \boldsymbol{\Lambda}^{1/2}\boldsymbol{U}^\top$ を考える．各点 $\boldsymbol{x}$ に白色化行列をかけると，任意の２点間の距離はユークリッド距離として計算できる (i.e. $D(\boldsymbol{x}, \boldsymbol{x}'; \boldsymbol{A}) = \|\boldsymbol{W}\boldsymbol{x} - \boldsymbol{W}\boldsymbol{x}'\|$)．武井ら [2] は，この性質に基づいて，白色化した局所特徴を混合正規分布に導入した．すなわち，武井らは次の確率モデルを用いた：

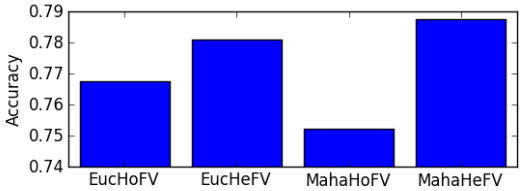
$$p_{\mathrm{homo}}(\boldsymbol{X}|\boldsymbol{\theta}) = |\det(\boldsymbol{W})| \prod_{t=1}^{T} \sum_{k=1}^{K} \pi_k \mathcal{N}(\boldsymbol{W}\boldsymbol{x}_t\,;\, \boldsymbol{\mu}_k, \mathrm{diag}(\boldsymbol{\sigma}_k)).$$

この確率モデルから導出される FV $\boldsymbol{f}_{\mathrm{homo}}(\boldsymbol{X})$ を**一様マハラノビスフィッシャーベクトル** (MahaHoFV) と呼ぶことにする．これは，全空間で同一のマハラノビス行列を用いているとみなすことができる．これに対して，局所ごとに異なるマハラノビス行列を用いる方法も考え

Copyright ©2015 Information Processing Society of Japan.
All Rights Reserved.



(a) FMD



(b) LSP15

図 1　**Categorization Performance.**

られる．本研究では混合成分ごとに異なる白色化行列 $\boldsymbol{W}_1, \ldots, \boldsymbol{W}_K$ を許す以下の確率モデルを用いる：

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

この確率モデルから導出される FV $\boldsymbol{f}_{\text{hetero}}(\boldsymbol{X})$ を**非一様マハラノビスフィッシャーベクトル** (MahaHeFV) と呼ぶ．

パラメータの決め方　MahaHeFV を計算するには，K 個の白色化行列の元となる K 個のマハラノビス行列 $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_K$ および，確率モデルのパラメータ $\boldsymbol{\theta}$ を求めておく必要がある．

本研究では，n_c クラスの多クラス分類への応用を考え，K 個の混合成分を各クラスにほぼ等分になるように割り振る．クラス c に割り振られた混合成分の個数を K_c と表すことにする．同じクラスに割り振られた混合成分における K_c 個のマハラノビス行列は共有することにし，クラス c のマハラノビス行列を $\boldsymbol{A}_{(c)}$ と表すことにする．$\boldsymbol{A}_{(c)}$ は，クラス c の訓練用画像における局所特徴の共分散行列の逆行列で定める．確率モデルのパラメータ $\boldsymbol{\theta}$ は，各クラスに割り振られた混合成分ごとに EM アルゴリズムで求める．これによって，合計 K 個の正規分布が得られる．さらに，武井ら [2] のようにクラス情報なしで K 混合正規分布を求め，これと混合することで，ひとつの確率モデル $p_{\text{hetero}}(\boldsymbol{X}|\boldsymbol{\theta})$ を得る．詳細は紙面の制約から割愛する．

4　実験

提案法 MahaHeFV の性能を評価するため，次の 3 つの方法と比較する：

- 従来のユークリッド距離による Fisher Vector (EucHoFV),
- MahaHeFV 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 (EucHeFV) ,
- 武井らの方法 (MahaHoFV) .

実験条件　画像から局所記述子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．$K = 256$ とした．正規化には Power 正規化, L2 正規化を用いた．ただし EucHeFV と MahaHeFV に対しては，独自の方法で 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), LSP15 上で, One-vs-rest SVM による多クラス識別を行った．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 枚ずつとした．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．

実験結果　図 1 に FMD および LSP15 上での実験結果を示す．実験の結果，一様 FV でも非一様 FV でもユークリッド計量よりもマハラノビス計量を用いたほうが識別率が高かった．さらに，MahaHeFV が MahaHoFV よりも若干ながら高い識別性能が得られた．

5　おわりに

本論文では，画像識別に用いられる 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 MahaHeFV 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通して，既存の方法に比べて識別率の向上を確認した．

参考文献

[1] Florent Perronnin and Christopher Dance. Fisher kernels on visual vocabularies for image categorization. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pp. 1–8. IEEE, 2007.

[2] 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. In *PRMU*, 2014.

Copyright ©2015 Information Processing Society of Japan.
All Rights Reserved.