確率的正準相関分析による単語ラベルを用いた局所特徴量の圧縮と詳細画像識別への利用

津田 智哉† 中山 英樹†
†東京大学 大学院情報理工学系研究科
E-mail: {tsuda, nakayama}@nlab.ci.i.u-tokyo.ac.jp

Abstract

画像認識の過程においてしばしば，ノイズを除去する目的や計算量を削減する目的で，主成分分析を用いて局所特徴量を圧縮する．しかし，主成分分析による局所特徴量の圧縮では，クラスの情報を活用しないため，ノイズと共に識別に重要となる情報が欠落してしまう．

そこで我々は，確率的正準相関分析を用いた局所特徴量の圧縮手法を提案する．確率的正準相関分析を用いることで，クラスの情報を活用しつつ，より識別に適した局所特徴量を生成することが可能となる．

また，詳細画像識別のデータセットに対して，提案手法を用いることで提案手法の有効性を示す．

1 はじめに

画像認識の過程において，ノイズを除去する目的や計算量を削減する目的で，前処理として局所特徴量の圧縮を行うことが少なくない．この圧縮の際によく使用される手法として，主成分分析 (PCA) が挙げられる．PCA は局所特徴量の元の分布を出来るだけ保存するように局所特徴量を潜在空間に射影することで圧縮する．しかし，画像に対応する単語情報を無視しているため，識別に重要となる情報が欠落してしまいがちであるという問題がある．

この問題を改善するために，我々は既に正準相関分析 (CCA) を用いた局所特徴量の圧縮手法 [1] を提案している．CCA を用いる場合，各画像に対応する単語情報を単語ラベル特徴量としてまず数量化する．そして，局所特徴量と単語ラベル特徴量の双方を活用して，圧縮後の潜在空間への射影行列を学習する．しかし，この手法では圧縮された局所特徴間の距離尺度が考慮されないため，相関の高い上位の固有ベクトルでもそうでない下位の固有ベクトルでも，距離の計算に同じ寄与をもってしまう．これは，局所特徴量を bag-of-visual-words (BoVW) [2] などへコーディングする際に問題となる．また，単語ラベル特徴量が寄与するのは射影行列の学習時のみで，実際にサンプルを潜在空間へ射影する際には無視されるという問題もある．

図 1 Caltech-UCSD Birds-200-2011

そこで本研究では CCA を用いた局所特徴量の圧縮手法をさらに発展させ，確率的正準相関分析 (PCCA) [3] の枠組みを適用することで，CCA を用いた圧縮手法における問題を克服し，識別に重要となる情報をより残しつつ局所特徴量を圧縮する手法を提案する．

この提案手法は一般画像認識において汎用的に利用可能であるが，特に効果を発揮すると期待できる分野として詳細画像識別 (fine-grained visual categorization, FGVC) [4] が挙げられる．FGVC は，ドメインこそ限定されているものの，非常に類似した画像群を多クラスに分類する困難なタスクである．代表的な例としては，動植物の品種の識別 [5][6] 等が挙げられる．FGVC では，クラス間の差異が極めて微小なため，この微小な差異を識別できるだけの強力な局所特徴量が求められる．

そこで，FGVC の代表的なデータセットである Caltech-UCSD Birds-200-2011[7] (図 1) に対して提案手法を用いた実験を行い，PCA や CCA による圧縮によって得られる局所特徴量に比べ，提案手法によって得られる局所特徴量がより識別に重要となる情報を有した強力な特徴量であることを示す．

2 提案手法

画像に対応する単語が既知である訓練データに対しては，PCCA を適用することで局所特徴量と各画像に対応する単語情報を数量化した単語ラベル特徴量の双方を用いて潜在空間へと射影する．一方，テストデータは当然のことながら画像しか与えられないため，局所特徴量のみを用いて潜在空間へと射影する．ここで，

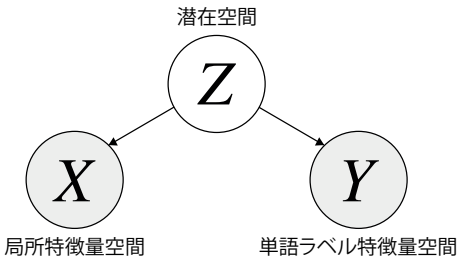

図 2　PCCA の確率モデル

訓練データを射影した潜在空間とテストデータを射影した潜在空間は，共通の潜在空間であることが理論的に保証されている (図 2).

2.1 前処理

訓練データとして画像とそれに対応する単語が与えられ，テストデータとして画像のみが与えられるとする．まず，訓練データとテストデータの全画像から SIFT 等の p 次元の局所特徴量を dense sampling する．その局所特徴量を $\boldsymbol{x} = (x_1, x_2, \cdots, x_p)^T$ とする．

また，総クラス数を q とすると，訓練データの各画像に対して単語ラベル特徴量を $\boldsymbol{y} = (y_1, y_2, \cdots, y_q)^T$ と定義する．なお，クラス k $(k = 1, 2, \cdots, q)$ に属する画像の単語ラベル特徴量は，$(y_1, y_2, \cdots, y_k, \cdots, y_q)^T = (0, 0, \cdots, 1, \cdots, 0)^T$ である．

そして，訓練データに関しては，画像の各局所特徴量 $\boldsymbol{x}$ にその画像の単語ラベル特徴量 $\boldsymbol{y}$ が対応するとする．これは，画像中の対象物体および背景領域に共通して単語ラベル特徴量 $\boldsymbol{y}$ を対応させるという，近似的なアプローチとなるが，FGVC においてはユーザーが対象物を中心に添えるシーンが多いと考えられるため，比較的妥当であると言える．

2.2 確率的正準相関分析

テストデータのように，局所特徴量 $\boldsymbol{x}$ のみが与えられ，画像の単語情報がない場合，PCCA による射影後の変数 z は正規分布をなし，その中心 $E(\boldsymbol{z}|\boldsymbol{x})$ と分散 Φ_x はそれぞれ次の式で表される．

$$E(\boldsymbol{z}|\boldsymbol{x}) = M_x^T A^T (\boldsymbol{x} - \overline{\boldsymbol{x}}), \tag{1}$$

$$\Phi_x = var(\boldsymbol{z}|\boldsymbol{x}) = I - M_x M_x^T. \tag{2}$$

一方，訓練データのように，局所特徴量 $\boldsymbol{x}$ と単語ラベル特徴量 $\boldsymbol{y}$ の両方が与えられた場合，PCCA による射影後の変数 z は同じく正規分布をなし，その中心 $E(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{y})$ と分散 Φ_{xy} はそれぞれ次の式で表される．

$$E(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{y}) = \begin{pmatrix} M_x \\ M_y \end{pmatrix}^T \cdot \begin{pmatrix} (I-\Lambda^2)^{-1} & -(I-\Lambda^2)^{-1}\Lambda \\ -(I-\Lambda^2)^{-1}\Lambda & (I-\Lambda^2)^{-1} \end{pmatrix} \begin{pmatrix} A^T(\boldsymbol{x} - \overline{\boldsymbol{x}}) \\ B^T(\boldsymbol{y} - \overline{\boldsymbol{y}}) \end{pmatrix}, \tag{3}$$

$$\Phi_{xy} = var(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{y}) = I - \begin{pmatrix} M_x \\ M_y \end{pmatrix}^T \begin{pmatrix} (I-\Lambda^2)^{-1} & -(I-\Lambda^2)^{-1}\Lambda \\ -(I-\Lambda^2)^{-1}\Lambda & (I-\Lambda^2)^{-1} \end{pmatrix} \begin{pmatrix} M_x \\ M_y \end{pmatrix}. \tag{4}$$

ここで，学習データの分散共分散行列を $C = \begin{pmatrix} C_{xx} & C_{xy} \\ C_{yx} & C_{yy} \end{pmatrix}$ と書くと，上の式の行列 A, B は次の一般化固有値問題の解である．

$$C_{xy} C_{yy}^{-1} C_{yx} A = C_{xx} A \Lambda^2 \ \ (A^T C_{xx} A = I_d), \tag{5}$$

$$C_{yx} C_{xx}^{-1} C_{xy} A = C_{yy} B \Lambda^2 \ \ (B^T C_{yy} B = I_d). \tag{6}$$

ただし，Λ は大きい順に d 個の正準相関係数を並べた対角行列であり，d は潜在空間の次元数である．

また，M_x, M_y は $M_x M_y^T = \Lambda$ かつ spectral norm がそれぞれ 1 未満という条件を満たす任意の行列である．ここでは単純に次の対角行列で与えることとする．

$$M_x = \Lambda^\beta, \ M_y = \Lambda^{1-\beta} \ (0 < \beta < 1). \tag{7}$$

β は潜在空間の学習において，局所特徴量と単語ラベル特徴量の寄与を調整するパラメータとなる．

なお，固有値問題の解を安定させ過学習を防ぐために，局所特徴量の共分散行列に正則化項を加える．すなわち，$C_{xx} \to C_{xx} + \alpha I$ とする．α は汎化を決めるパラメータである．

PCCA によって射影された潜在空間における正規分布の中心を圧縮後の局所特徴量とみなす．つまり，局所特徴量 $\boldsymbol{x}$ のみが与えられる場合と局所特徴量 $\boldsymbol{x}$ および単語ラベル特徴量 $\boldsymbol{y}$ が与えられる場合において，それぞれ最終的な局所特徴量を次の式で表す．

$$\boldsymbol{v}_x = E(\boldsymbol{z}|\boldsymbol{x}), \tag{8}$$

$$\boldsymbol{v}_{xy} = E(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{y}). \tag{9}$$

圧縮に際して，汎化性を決めるパラメータ α，および局所特徴量と単語ラベル特徴量の寄与を調整するパラメータ β を決める必要がある．また，これによって得られた新たな局所特徴量を BoVW や Fisher Vector [8] に適用して最終的な画像特徴ベクトルを得る．

3 実験

FGVC において標準的に用いられるベンチマークである，Caltech-UCSD Birds-200-2011 を用いて評価を行った．Caltech-UCSD Birds-200-2011 は 200 クラスの鳥の画像からなるデータセットであり，5994 枚の訓練データと 5794 枚のテストデータがある．画像から RGB-SIFT 特徴量 [9] を dense sampling し，PCA, CCA, および提案手法 (PCCA) を用いて 16, 32, および 64 次元に圧縮したものを局所特徴量として比較した．

最終的には，これらの局所特徴量を ガウシアンの数が 32 の Fisher Vector へ加工し，ロジスティック回帰により識別器を構築し，各クラスごとの識別正解率の平均をスコアとして評価した．

表 1 に実験結果を示す．PCCA 適用の際のパラメータ β の値は 0.01 (16 次元), 0.005 (32 次元), 0.01 (64 次元) である．

表 1　Caltech-UCSD Birds-200-2011 (%)

Descriptor	16	32	64
PCA(Baseline)	19.09	27.53	31.65
CCA	28.93	33.26	33.29
PCCA	**29.34**	**33.29**	**33.71**

また，PCA, CCA, および PCCA によって圧縮した局所特徴量を用いて識別を行った際の，ロジスティック回帰の出力をクラスごとに複数掛け合わせることで得たスコアを表 2 に示す．ここで，PCCA 適用の際のパラメータ β は，表 1 中の各圧縮次元中の最高スコアに対応する β の値である．

表 2　Late-fusion (%)

Descriptor	16	32	64
PCA+CCA	30.05	34.58	35.99
PCA+PCCA	**30.09**	**34.61**	**36.24**
CCA+PCCA	29.27	33.37	33.98
PCA+CCA+PCCA	**31.77**	**35.01**	**36.51**

4　結論・考察

表 1 より，提案手法である PCCA による局所特徴量の圧縮が詳細画像識別に対して有効であることが確認できる．PCA のスコアに比べて CCA および PCCA のスコアが大きく優っていることから，PCA による圧縮では欠落した識別に重要な情報を，単語ラベルの情報を用いる PCCA および CCA による圧縮では抽出できていると考えられる．また，より低次元に圧縮するほど単語ラベル特徴量を用いる CCA および PCCA の優位性が大きくなることが分かる．これより，提案手法がより識別に重要な情報に重みを置きつつ局所特徴量を圧縮していることが言える．

表 2 において，PCA+CCA および PCA+PCCA によって得られるスコアは単体のスコアを大きく上回るのに対し，CCA+PCCA ではスコアの上昇幅が小さいことが分かる．これは互いの潜在空間が類似しているためだと考えられる．

そして，表 1 と表 2 から分かるように，PCCA 単体および PCA+PCCA のスコアが CCA 単体および PCA+CCA のスコアをいずれも上回っていることから，提案手法は 1 章で挙げた CCA の問題点を克服し，提案手法が CCA に対して優位性をもつと言える．

また，PCA+CCA に比べて PCA+CCA+PCCA のスコアが高いことから，提案手法では PCA および CCA による圧縮では抽出できていない，識別に重要となる情報を抽出できていると考えられる．

5　将来展望・応用先

今回は画像から抽出した全ての局所特徴量に対して，近似的にその画像の単語ラベル特徴量を対応させたが，画像中の物体領域から得られた局所特徴量と景領域から得られた局所特徴量にそれぞれ異なる単語ラベル特徴量を対応させることで自動的に前景抽出を行い，より識別精度を向上させるアプローチを検討したい．

提案手法は画像間の微小な差異の識別に有効であるため，医用画像における微小な差異を識別し病気を特定する等の応用例が考えられる．

謝辞

本研究は公益財団法人栢森情報科学振興財団の支援を受けて実施された．

参考文献

[1] H. Nakayama, “Augmenting descriptors for fine-grained visual categorization using polynomial embedding”, *In Proc. ICME*, 2013.

[2] G. Csurka, *et al*, “Visual Categorization with Bags of Keypoints”, *In Proc. ECCV*, 2004.

[3] F. R. Bach, *et al*, “A probabilistic interpretation of canonical correlation analysis”, Technical Report 688, Department of Statistics, University of California, Berkeley, 2005.

[4] I. Biederman, *et al*, “Subordinate-level object classification reexamined”, Psychological Research, vol. 62, pp. 131-153, 1999.

[5] M-E. Nilsback, *et al*, “Automated flower classification over a large number of classes”, Proceedings of the Indian Conference on Computer Vision, Graphics and Image Processing, 2008.

[6] S. Yang, *et al*, “Unsupervised template learning for fine-grained object recogmition”, *In Proc. NIPS*, 2012.

[7] C. Wah, *et al*, “The Caltech-UCSD Birds-200-2011 Dataset”, Computation & Neural Systems Technical Reports, CNS-TR-2011-001, 2011.

[8] Perronnin, *et al*, “Improving the Fisher kernel for large-scale image classification”, *In Proc. ECCV*, 2010.

[9] Koen E. A. van de Sande, *et al*, “Evaluating Color Descriptors for Object and Scene Recognition”, IEEE Transactions on Pattern Analysis and Machine Intelligence, volume 32 (9), pages 1582-1596, 2010.