

令和元年6月13日現在

機関番号：82626

研究種目：基盤研究(C)（一般）

研究期間：2015～2018

課題番号：15K00261

研究課題名（和文）パターン認識のための特徴量変換に関する研究

研究課題名（英文）A study on feature transformation in pattern recognition

研究代表者

小林 匠（Kobayashi, Takumi）

国立研究開発法人産業技術総合研究所・情報・人間工学領域・主任研究員

研究者番号：30443188

交付決定額（研究期間全体）：（直接経費） 3,700,000円

研究成果の概要（和文）：パターン認識において認識性能を左右する重要なデータ特徴表現に関して、抽出された特徴量の弁別力を向上させる様々な特徴変換手法を提案した。入力データの内容を効率的に表現する特徴量においては、特徴抽出器やデータ分布の特性に由来する構造的情報が内在している。提案法では、そのような特徴量の本質的構造に基づき、弁別的情報を抽出し強調する。特に、構造としてヒストグラムのボケや事前確率モデル、さらには物理次元や不変性といった観点に着目し、計算量も低コストな手法を構築した。提案した特徴変換法により、特徴表現の高精度化、ひいては認識システムの高性能化を達成した。

研究成果の学術的意義や社会的意義

近年、計測データの大規模化・多様化が進み、それらデータをサービス等へ利活用するためのデータ自動認識技術、いわゆるAIの需要が急速に拡大している。本研究は、自動認識の中核を成すパターン認識の性能改善に資するものである。特に、特徴抽出の後処理という位置づけで、様々な既存認識システムに容易かつ計算量的にも低コストで導入できるため波及効果も期待できる。さらにそのような実用面のみでなく、既存特徴量の変換処理に着目し、そこに数理的視点を導入した点でも学術的な意義が大きい。

研究成果の概要（英文）：In this study, we have proposed various feature transformation methods to enhance the discriminative power of features. In general, the feature to represent the content of input data contains structural information which is derived from the characteristics of feature extractors and input data distribution. The proposed methods leverage the essential structures to improve the discriminativity of the features. Those methods are formulated especially by focusing on the deblurring of histogram, prior probabilistic models, physical structures and invariance to input data perturbation. We can apply the methods in a computationally efficient manner, while contributing to the improvement of feature representation as well as performance of the whole recognition systems.

研究分野：パターン認識

キーワード：特徴抽出 特徴変換 画像認識 動画認識

様式 C - 19、F - 19 - 1、Z - 19、CK - 19 (共通)

1. 研究開始当初の背景

(1) 計測センサや記憶装置の発展に伴い、大規模な計測データを蓄積することが可能となり、それらデータをサービス等に利活用するためにはデータの自動認識処理、いわゆる AI (人工知能) 技術が必要不可欠となる。

(2) 自動認識の中核を成すパターン認識の枠組みでは、右図に示されるように、入力データを特徴抽出技術により特徴量 (一般にはベクトル形式) として表現し、認識課題に沿って適切な識別・弁別処理を施すことで、所望の認識結果を得ることができる。

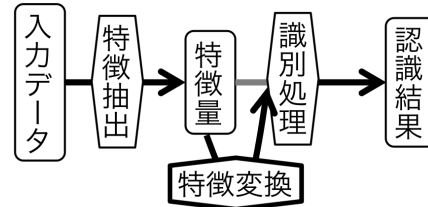


図: パターン認識処理の流れ

(3) そのため、入力データの内容をより良く反映した特徴表現が高性能な認識システムを実現するためには必要であり、これまで様々な特徴抽出技術が提案されている。その一方で、抽出した特徴量を後処理として変換する特徴変換については、これまであまり注目されていなかった。それ故、広く適用されている変換手法としては、ベクトルノルム正規化程度に留まっていた。

2. 研究の目的

(1) そこで、本研究では特徴表現の高精度化、ひいては認識システムの高性能化を目指し、抽出された初期特徴量をより識別に有利な弁別力の高い形式へと変換する新たな特徴量変換技術を開発する。

(2) 初期特徴量はランダムな値ではなく、特徴抽出器の性質や入力データの分布特性によって、特徴量の中にはある構造が内在している。そのような、特徴量の持つ本質的な情報を抽出し、強調することによって特徴量変換手法を定式化する。

(3) また、特定のタスクや特徴抽出器に特化することなく、様々な特徴量へ適用可能な汎用的な技術を開発することも目的となる。同時に、計算量的観点からも効率が良く、従来の認識システムへ容易に導入可能となる手法の構築を目指す。

3. 研究の方法

本研究では、その汎用性や適用範囲の広さにより、画像や動画から抽出される特徴量を対象として特徴変換の研究を進めた。特徴量に内在する構造として、本研究ではヒストグラム形式や物理座標系、さらには入力変動への不変性に着目して下記に示す様々な手法を構築した。

(1) ヒストグラム特徴のボケ除去

画像の検索や認識を目的として、局所特徴量に基づく visual word 特徴表現が広く適用されている。これは局所特徴量の分布を量子化し離散化することにより、ヒストグラム形式で画像特徴を表現している。この特徴抽出過程では、本来は連続的な分布を離散量子化しているため、量子化誤差等が避けられない。本研究では、これをヒストグラムのボケ過程とみなした。つまり、近傍の量子化点へ情報 (特徴量) が漏れ、本来のヒストグラム特徴からボケた特徴量が実際には算出されている。このボケ過程により特徴表現の弁別力が損なわれるため、ボケを除去する特徴変換を定式化する。ボケ過程は次式のランダムウォークとして数理的に定式化できる。

$$z = Pz' \quad (z: \text{得られた } C \text{ 次元特徴}, z': \text{本来の } C \text{ 次元特徴}, P: C \times C \text{ 遷移行列}) \dots \text{式}(3.1.1)$$

ここで遷移行列 P は C 個の量子化点間での類似度行列により導出可能であり、スパース行列として表現される。上式より、本質的なヒストグラム特徴 z' は、その非負性に注意して、

$$z' = \max[P^{-1}z, 0] \quad \dots \text{式}(3.1.2)$$

と解析的に陽に求まる。しかしながら、遷移行列 P 自体はスパースであってもその逆行列 P^{-1} は密な行列となる。さらに、一般に量子化数 C が大きい (~数千) ことも加味すると、式(3.1.2)の変換には計算コストがかかる。そこで、 P のスパース性を活用した次の近似方式を提案した。

$$P = I + (1 - \alpha)Q, \quad P^{-1} \approx I + \alpha(I - Q), \quad z' = \max\{\{I + \alpha(I - Q)\}z, 0\} \quad \dots \text{式}(3.1.3)$$

ここで、 Q は P の非対角行列でスパース性を維持しており、 α は から算出できるパラメータである。式(3.1.3)により、特徴変換はスパース行列とベクトルとの単純な積から成る低計算コストの手法となり、容易にボケ除去を実現できる。

(2) ヒストグラム特徴の事前確率モデル

ヒストグラム特徴は上述のように、分布の離散表現であるため、それ自体が離散確率密度関数とみなせる。そのような離散確率に対する事前確率モデルとしてはディリクレ分布を自然に

仮定することができる。

$$p(z|\theta) \propto \prod_i^C z_i^{\theta_i - 1} \dots \text{式(3.2.1)}$$

ここで θ はディリクレ分布のパラメータ (C 次元) を示す。このディリクレ分布の特性をより良く表現する手法としてフィッシャーカーネル[1]を採用し、これを特徴変換として定式化する。フィッシャーカーネルは、 \log 尤度のパラメータ微分とフィッシャー情報行列により構成され、

$$H^{-\frac{1}{2}} \nabla_{\theta} \log[p(z;\theta)] = H^{-\frac{1}{2}} [\log(z) - \mu_{\theta}] \dots \text{式(3.2.2)}$$

と表現できる。ここでは μ は \log ヒストグラム特徴の平均で算出できるため、残る問題はフィッシャー情報行列の効率的な表現である。ここでは、トリガンマ関数 ψ の特性などを考慮して、十分大きなヒストグラム特徴次元 C では次の近似表現が得られることを導いた。

$$H \approx \text{diag}(\psi'(\theta) - \psi'(\theta_0)\mathbf{1}) = \text{diag}(\sigma_{\theta}^2) \dots \text{式(3.2.3)}$$

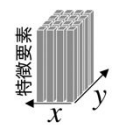
ここで σ_{θ}^2 は \log ヒストグラム特徴量の分散として算出することができる。以上をまとめると、事前確率モデルに基づく特徴変換として

$$\text{diag}(\sigma_{\theta}^{-1})(\log(z) - \mu_{\theta}) \dots \text{式(3.2.4)}$$

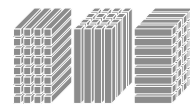
という簡便かつ計算コストも低い表現を実現した。

(3) 特徴の物理座標系

画像内に写る対象は空間的構造を有しており、その構造特性を抽出する目的で特徴抽出においても、局所的な物理座標系に沿った形で特徴量が抽出されることが多い。例えば2次元座標系沿って対象領域を小領域に分割し、各々の領域から特徴量を抽出するなどの方法がある(右図)。従来はそのように抽出された特徴量も1次元へ伸長してベクトルとして扱ってしまっていたが、それでは対象のもつ物理座標での空間特性を十分に反映した表現とは言い難い。そこで、このような物理座標系を考慮した特徴変換を考えることで、物理座標上の特性を反映した特徴表現を得ることを試みた。座標系を加味した特徴量は高階テンソル(上図では3階テンソル)となるが、その各軸での特徴に分割して、特徴変換手法を構築する。上図の特徴量は、右図のような各軸での特徴チューブに分割される。この特徴チューブに対して、類似度をまず与える。特徴表現の形式はその類似度と密接な関係にあり、例えば内積類似度を考える場合には一般のベクトル形式での特徴が得られる。ここでは類似度として SSIM (Structural Similarity Index Measure) [2] を考えた。さらに類似度の近似表現を導出することで、最終的な特徴変換が定式化した。



図：物理座標を加味した特徴



(4) 特徴量の反転不変

画像や動画では、対象を撮影する向きが変わることで、頻繁に水平方向での左右反転が起こる。しかしながら、一般の特徴量はそのような左右反転の入力変動に対して不変ではなく、反転することで異なる特徴量となってしまふ。入力の画像や動画の向きを揃える操作なども有効と考えられるが、そのような前処理は計算コストもかかり、その他の外乱により向き検出の精度自体も高いとは言えない。そのため、ここでは特徴変換により、そのような左右反転変動に不変な(頑健な)特徴表現を獲得した。上記(3)で述べたように特徴量は物理座標系に従って抽出され、そのような場合には左右反転変動は特徴要素間のスワップ操作として表現できる。そのため、一般にそのような変動は、特徴量 z に対して変動による変換行列を T と表すと、

$$\hat{z} = T^{\top} z \dots \text{式(3.4.1)}$$

と定式化される。また、ここで特徴変換として線形モデル(変換ベクトル w) を考えると、変動に対する不変性は

$$w^{\top} \hat{z} = w^{\top} T^{\top} z = w^{\top} z \Rightarrow T w = w \dots \text{式(3.4.2)}$$

を満たす w により与えられることがわかる。これは変換行列 T の固有値 1 に対する固有ベクトルが w となることを意味している。変換行列 T が固有値 1 以外をとる一般の場合へ拡張すると、特徴変換は変換行列 T の固有ベクトル w と固有値 λ を用いて

$$f_i(z) = \begin{cases} w_i^{\top} z & \lambda_i \geq 0 \\ |w_i^{\top} z| & \lambda_i < 0 \end{cases}, i \in \{1, \dots, C\} \dots \text{式(3.4.3)}$$

と定式化される。ここでの変換行列 T は陽に与えることも、学習で獲得することも可能である。

4. 研究成果

様々な画像認識実験において提案法の有効性を定量的に確認した。

(1) ヒストグラム特徴のボケ除去

局所特徴量として SIFT 記述子[3]を用いた visual word 特徴に対してボケ除去特徴変換を適用した。量子化数の増加に伴ってボケ除去による性能改善が明確になることを確認し、これは、量子化数の増大により近傍量子化点への特徴量の漏れ（ボケ）が起こりやすくなることに起因していると考えられる。ここでの word 数（量子化数）は 16384 とした。一般画像認識課題である VOC2007[4]や Caltech-256[5]、風景認識の MIT-67[6]へ適用した結果、下表のように良好な性能改善を得た。

表：Visual Word に基づく画像認識性能（正解率[%]）

	VOC2007 [4]	Caltech-256 [5]	MIT-67 [6]
特徴変換前	60.6%	54.1%	58.4%
特徴変換後	61.5%	55.5%	60.2%

(2) ヒストグラム特徴の事前確率モデル

ヒストグラムの事前確率モデルであるディリクレ分布モデルに基づく特徴変換を画像認識課題に適用した。ここでは、広く用いられている混合ガウス分布モデルに基づくフィッシャーカーネル表現[7]を比較対象とした。下表に示すように、ヒストグラム特徴の事前分布を考慮することでより良い特徴表現へと変換され、その結果性能向上に寄与していることがわかる。

表：フィッシャーカーネルに基づく画像認識性能（正解率[%]）

	VOC2007 [4]	Caltech-256 [5]	MIT-67 [6]
特徴変換前	60.0%	55.4%	60.3%
特徴変換後	63.8%	57.4%	63.4%

(3) 特徴の物理座標系

上述の局所記述子 SIFT も局所領域の分割によって構成される特徴量である。SIFT に対して物理座標を考慮した特徴変換を施し、画像検索タスクで性能を評価した結果が左下表となる。さらに、ここでは深層畳み込みニューラルネットワーク(CNN)[8]による特徴に対しても同様に適用した。ニューラルネットワークの中間層では、畳み込み特徴が2次元座標に並んだ特徴マップが得られているため、特徴マップが変換対象となる。風景認識課題に対して適用した結果が右下表である。いずれの場合も良好な性能改善が実現できていることがわかる。

表：SIFT による画像検索性能（mAP）

	Oxford building [9]
特徴変換前	67.6
特徴変換後	73.2

表：CNN 特徴による画像認識性能（正解率[%]）

	Scene-15 [10]	SUN-397 [11]
特徴変換前	90.3	52.1
特徴変換後	91.2	54.3

(4) 特徴量の反転不変

左右反転が頻繁に起こる行動認識に対して反転不変特徴変換を適用した。動画像を用いた行動認識では、ビデオカメラの向きが変わることで同一行動が反転した形で動画像において表現されることになる。この動画像認識に対して局所特徴量として MBH (Motion Boundary Histogram) [12]などの Hand-craft 特徴や畳み込みニューラルネットワーク[13]による学習的特徴量を抽出し、変換手法を施した。前者の Hand-craft 特徴では反転による変換行列が陽に求まるのに対して、後者の学習的特徴では変換行列が未知であるため、学習により獲得したものをを用いた。下表が認識性能の結果である。不変性を特徴表現の段階で導入することにより、識別器の学習効率も上がり、性能が向上していることがわかる。

表：動画像を用いた行動認識性能（正解率[%]）

	Hollywood2 [14]	HMDB-51 [15]	UCF-101 [16]
特徴変換前	70.5%	64.4%	89.7%
特徴変換後	72.1%	65.1%	90.3%

< 引用文献 >

- [1] T.S.Jaakkola and D.Haussler. "Exploiting generative models in discriminative classifiers," In NIPS, 1999.
- [2] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. "Image quality assessment:

From error visibility to structural similarity, " IEEE Transactions on Image Processing, 13(4):600-612, April 2004.

[3] D.Lowe. "Distinctive image features from scale invariant features," International Journal of Computer Vision, 60:91-110, 2004.

[4] "The PASCAL Visual Object Classes Challenge 2007 (VOC2007)," <http://www.pascal-network.org/challenges/VOC/voc2007/index.html>.

[5] G.Griffin, A.Holub,and P.Perona, "Caltech-256object category dataset," Tech. Rep. 7694, Caltech, 2007.

[6] A. Quattoni and A. Torralba, "Recognizing indoor scenes," in CVPR, 2009, pp. 413-420

[7] J. Sanchez, F. Perronnin, T. Mensink, and J. Verbeek. "Image classification with the fisher vector: Theory and practice," International Journal of Computer Vision, 105(3):222-245, 2013.

[8] K. Simonyan and A. Zisserman. "Very deep convolutional networks for large-scale image recognition," CoRR, abs/1409.1556, 2014.

[9] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman. "Object retrieval with large vocabularies and fast spatial matching," In ICCV, 2007.

[10] S. Lazebnik, C. Schmid, and J. Ponce. "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," In CVPR, pages 2169-2178, 2006.

[11] J. Xiao, J. Hays, K. A. Ehinger, A. Oliva, and A. Torralba. "Sun database: Large-scale scene recognition from abbey to zoo," In CVPR, 2010

[12] N. Dalal, B. Triggs, and C. Schmid. "Human detection using oriented histograms of flow and appearance," In ECCV, pp. 428-441, 2006.

[13] L. Wang and Y. Qiao. "Action recognition with trajectory-pooled deep-convolutional descriptors," In CVPR, pp. 4305-4314, 2015

[14] M. Marszalek, I. Laptev, and C. Schmid. "Actions in context," In CVPR, pages 2929-2936, 2009.

[15] H. Kuehne, H. Jhuang, E. Garrote, T. Poggio, and T. Serre. "Hmdb: A large video database for human motion recognition," In ICCV, pages 2556-2563, 2011.

[16] K. Soomro, A. R. Zamir, and M. Shah. "Ucf101: A dataset of 101 human action classes from videos in the wild," In CRCV-TR-12-01, 2012.

5 . 主な発表論文等

〔雑誌論文〕(計 0 件)

〔学会発表〕(計 4 件)

T. Kobayashi, Flip-Invariant Motion Representation, 査読有, Proc. International Conference on Computer Vision (ICCV), 2017, pp. 5629-5638

小林匠, 機械学習を用いた画像特徴量変換, 招待講演, 第 20 回情報論的学習理論ワークショップ, 2017

T. Kobayashi, Structured Feature Similarity with Explicit Feature Map, 査読有, Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 1211-1219

T. Kobayashi, Histogram Feature Deblurring, 査読有, Proc. International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2016, pp. 1871-1875

〔図書〕(計 0 件)

〔産業財産権〕

出願状況 (計 0 件)

取得状況 (計 0 件)

〔その他〕

ホームページ等

6 . 研究組織

(1)研究分担者：なし

(2)研究協力者：なし

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。