

# 距離情報を手がかりとした画像認識における転移学習手法の改良

岡本昌也† 中山英樹†

† 東京大学大学院 情報理工学系研究科

E-mail: {okamoto,nakayama}@nlab.ci.i.u-tokyo.ac.jp

## Abstract

本研究では、画像認識における転移学習手法を改良し、これまで画像情報のみを用いて行われてきた転移学習の問題に対して、付加的な情報を用いる、新たな転移学習の枠組みを提案する。具体的には、画像と同時に取得された距離情報を用いた転移学習手法を提案する。本手法では、より精度良く知識の転移を行うため、画像と共に取得された距離情報を弱教師信号と解釈し、転移先の特徴空間において PLS(Partial Least Squares)を行い、意味的に近いサンプルが近くに配置される様な部分空間を選択することで、転移学習の精度を向上させる手法を提案する。実験では、従来手法と比較して 3.49%精度が向上することを示し、提案手法の有効性を示した。

## 1 はじめに

### 1.1 背景

画像認識の分野では大量の教師データが必要とされる中で、データに対する手動でのラベル付けは非常に人的コストが大きい。そこで、既にラベルが付いているドメイン（照明や解像度などの環境）が異なる画像群で学習した知識を用いて、対象とするドメインを認識するという、画像認識における転移学習手法が Saenko らの手法 [1] をはじめとして多数提案され、注目を集めている。

転移学習とは、異なるドメイン（照明や背景など環境的条件が同じデータの集まり）から学習した知識を活用する方法のことである。異なるドメインの画像を図 1 に示す。



図 1 異なるドメイン間の違いの例

Saenko らは、ソースドメインの画像特徴をうまく変換することで、実際にターゲットドメインに学習データを用意する場合の精度に近づけることが出来る事を示し、転移学習が画像認識の分野にも応用できる可能性を示した。

しかし、従来の研究では転移先の空間において画像情報のみ、もしくは、画像情報と少数のラベルのみを使用しており、それら以外の情報を使用した転移学習手法は提案されてこなかった。

一方で、ウェアラブルコンピューティングの発展や、センシングデバイスの小型化などに伴い、画像の取得と同時に、位置情報や方位情報など付加的な情報を容易に取得出来る様になってきている。特に距離情報に関しては、Kinect などの機器が普及し、容易に画像に対する付加情報として取得する事が出来るようになってきた。そこで、本論文では、距離情報をはじめとする付加情報を用いた転移学習の新たな枠組みを提案する。

### 1.2 目的

本論文では、画像と同時に取得された付加的な情報を手がかりとして、転移学習を行う手法を提案する。具体的には、深度画像を付加情報として使用する。

本論文の目的は、距離情報を活用した新たな転移学習手法の提案と同時に、今までの様な画像情報のみを使用した転移学習手法ではなく、付加的な情報を追加で使用する新たな問題の枠組みを提案することである。

実験では、収集が容易でかつ、大量に存在する Web 上に存在する一般的な画像群を既知ドメインとして、実際の環境下で収集された少数の画像に対して、物体認識することとした。つまり、Web 上に存在する画像をソースドメインの画像とし、実際の環境下で撮影された画像をターゲットドメインにするという事である。

## 2 関連研究

物体認識におけるドメイン適用の問題は、Saenko らの論文 [1] が提起した。この研究は、information-theoretic metric learning(ITML) の手法 [2] を応用し、異なるドメイン間のサンプルが、同じクラス間では近く、異なるクラスのサンプル間では離れる様に、マハラノビス距離尺度を求める手法を提案している。ターゲット側

にも少数の教師データを前提とする転移学習の教師学習の手法である．

目的を考えた場合，ターゲット側に教師データを一切前提としない教師なしの手法が望ましいが，非常に難しい問題である．教師なしの手法では，ソースとターゲット間の補完的な部分空間を用いるアプローチが有望であると考えられており，その最初の研究が Gopalan らのもの [3] である．この手法は，ソース側の特徴空間からターゲット側へ明示的に部分空間を作り中間的な部分空間ではソースとターゲットの分布がある程度近づいている．それらの中間的な部分空間へ射影した点を全てつなげて特徴ベクトルとする手法である．

Gopalan らの手法を改良した，解析的に解く方法として Boqing らの Geodesic Flow Kernel(GFK) の手法 [4] がある．この手法は，明示的に部分空間を作り出すのではなく，問題を固有値問題として解析的に解く方法である．ソースドメインについては画像各々について正確なラベルが付いており，PLS[5] を適用してクラスを表すラベルが同じサンプルが近くに配置される様な部分空間を選ぶことで，PCA を適用する場合に比べて最終的な分類精度が向上することが知られている．

PLS とは，出力変数との共分散を最大化するような射影を見つける手法である．

いずれの手法もターゲット側のデータとして，画像情報のみ，もしくは，画像情報と少数の教師信号を用いる手法である．本論文では，GFK の手法を改良して，ターゲット側の画像に対応する付加的な情報，距離情報を用いる，新たな転移学習手法を提案する．

### 3 提案手法

本論文で提案する手法の詳細を説明する．

#### 3.1 処理の流れ

処理の大まかな流れは図 2 のようになる．

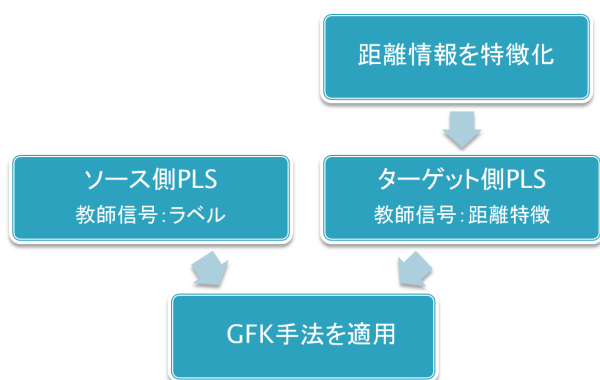


図 2 処理の流れ

1. ソース側の特徴空間においてラベルを教師信号として PLS を行う

2. ターゲットドメインの距離情報から，各画像についての距離特徴を求める
3. ターゲット側の特徴空間において距離特徴を教師信号として PLS を行う．この部分が本手法の最もコアとなる部分である．
4. GFK の手法を適用する

以下でそれぞれについて，説明を行う．

#### 3.2 ソース側の PLS

この部分は，従来手法でも行われている試みであり，提案手法のコントリビューションではない．4 節の実験でも，この処理を行わず，PCA を適用することで代用する手法（提案手法 1）と，この処理を行う手法（提案手法 2）の 2 つの比較を行う．

転移学習ではソース側の画像には，正確なラベルが付いているということが前提になっているので，その正確な教師信号を手がかりに PLS を適用し，分布を改善する．

#### 3.3 距離情報から距離特徴を生成

本論文で使用する距離情報からの特徴は，Bo らの提案した Kernel descriptors という手法 [6] を使用して，深度画像から生成した．Bo らの論文では，画像情報と深度画像から生成される特徴量と，深度画像からのみから生成されるものの 2 つが提案されているが，今回の実験では後者を利用した．実験では，Bo らの配布しているコード<sup>1</sup>を利用した．

#### 3.4 ターゲット側の PLS

ソース側で PLS を適用して分布を改善することにより転移の精度が向上する事が知られている事を受けて，ターゲット側でも，何らかの情報を教師信号と解釈して PLS を適用すれば同様の効果が期待できると考えられる．

3.3 節で生成した，距離情報から与えられる特徴をクラス内の特徴にある程度の相関を期待する，画像特徴を説明する説明変数と解釈して，ターゲット側の特徴空間において PLS を適用する．つまり，同じクラスのサンプル同士が近くなり，異なるクラスのサンプル間が離れるというある種の分布の改善を予め行うことで，精度の向上を目指す．

#### 3.5 GFK 手法の適用

転移学習のコアとなる，固有値問題を解き，転移学習を行う部分は GFK と同じ手法をとる．配布されているコードをベースとして，それを改良することで実装を行った．

この後，分類手法を用いて最終的な物体認識の結果を得る．転移学習の精度を比較する為，実験では分類手法として一番単純な最近傍法を使用した．

<sup>1</sup><http://www.cs.washington.edu/robotics/projects/kdes/>

## 4 実験

### 4.1 使用するデータセット

提案手法を行う為、ソース側のデータとして ImageNet[7]、ターゲット側のデータとして Janoch らの RGB-D 物体データセット (B3DO)[8] を使用した。B3DO は、各画像に対応する距離データを表す深度画像が、画像に加えて提供されている。B3DO に含まれる画像数が多い6クラスを対象とし、対象とする物体の部分のみを、画像及び深度画像を切り出して使用した。

結果として、収集したデータセットは次の表1のようになった。

表1 データセットの画像数

| クラス名     | B3DO  | ImageNet |
|----------|-------|----------|
| bottle   | 238   | 920      |
| bowl     | 142   | 919      |
| cup      | 256   | 919      |
| keyboard | 129   | 1512     |
| monitor  | 243   | 1134     |
| sofa     | 109   | 982      |
| 合計       | 1119  | 6386     |
| 平均       | 186.5 | 1064.3   |

また、実際の画像の例を図3に示す。ターゲットドメインの B3DO には、画像に対応する深度画像が付随する事に注意する。図中の深度画像は見やすさのため、カラーレンジを調節しているが、実際の実験には B3DO データセット内データをそのまま使用した。



図3 データセットの画像例

### 4.2 精度比較実験

まず最初に、本論文で提案した手法と従来手法の精度比較を行う。

#### 4.2.1 実験方法

4.1 節のデータセットに対して、6クラス分類の精度比較実験を行う。分類手法は最近傍法を用いた。比較する手法は次の4つである。

1. 提案手法1 (ソース: PCA ターゲット: PLS)
2. 従来手法1 (ソース: PCA ターゲット: PCA)
3. 提案手法2 (ソース: PLS ターゲット: PLS)
4. 従来手法2 (ソース: PLS ターゲット: PCA)

2,4 は GFK 法を適用する。ソース側に PCA を適用する場合である提案手法1と従来手法1を比較し、PLS を適用する場合である提案手法2と従来手法2を比較する。部分空間の次元数は、それぞれの手法について最も精度が高くなるように、10,20,30,40,50の中から選択した。各種法は異なる部分空間を生成するので、次元数を固定する事で特定の手法に有利にならない様、手動で選択している。

画像特徴として、SIFT 特徴 [9] の Bag of Feature(BoF) 表現 [10] を使用する。コードブックはソース側の画像のみから作成し、サイズは1000とした。

ターゲット側の画像は全てテスト画像とし、ソース側の画像は一定数を使用する。ソース側の画像数を1クラス100枚の計600枚とした。

#### 4.2.2 実験結果

精度比較実験を行った結果は、次の表2のようになった。

表2 各手法の分類精度。表中、次元数とは部分空間の次元数であり、精度が最大になる次元数を表す。提案手法1は部分空間の次元数が20と50の時、精度が最大であったことを示している。

| 手法名               | 分類精度 (%)     |
|-------------------|--------------|
| 提案手法1 (次元数=20,50) | <b>29.04</b> |
| 従来手法1 (次元数=30)    | 28.60        |
| 提案手法2 (次元数=10)    | <b>32.53</b> |
| 従来手法2 (次元数=50)    | 29.04        |

結果として、ソース側、ターゲット側の両方で PLS を行った提案手法2が最も良いという結果になった。

#### 4.3 ソース側データ数変更実験

次に、ソース側のデータ数を変化させ、各々で本論文で提案した手法と従来手法の精度比較を行う。ソース側により大規模な知識が提供される場合の、手法ごとの振る舞いの違いを検証する。

#### 4.3.1 実験方法

ソース側の画像を 1 クラス 20,100,500 枚ずつ、すなわち総データ数を 120,600,3000 枚と変化させ、それぞれについて、各手法の精度比較を行う。

部分空間の次元数は 4.2 節と同様に、それぞれの手法について最も精度が高くなるように、10,20,30,40,50 の中から選択した。

#### 4.3.2 実験結果

ソース側データ数変更実験を行った結果は、次の表 3 の様になった。

表 3 ソース側データ数変更実験

| 総データ数 | 提案 1         | 従来 1         | 提案 2         | 従来 2  |
|-------|--------------|--------------|--------------|-------|
| 120   | 28.33        | <b>28.95</b> | <b>32.35</b> | 31.64 |
| 600   | <b>29.04</b> | 28.60        | <b>32.53</b> | 28.87 |
| 3000  | <b>33.42</b> | 31.72        | <b>34.94</b> | 33.92 |

結果として、いずれのデータ数についても提案手法 2 が最も良いという結果になった。ソース側のデータ数が少ないときの方が、相対的な精度上昇率は高いことが確認出来る。

### 5 考察

#### 5.1 精度比較実験

精度比較実験の結果をみると、提案手法は、従来手法の最も精度の良い、ソース側で PLS を行う場合に比べて、3.49% 上昇しており、従来手法に対して、本手法の方が精度が高くなる事を実証出来た。

また、ソース側に PLS を適用せず、PCA を適用する場合（提案手法 1 及び従来手法 1）についても、従来手法に比べて 0.44% 精度が上昇することが確認出来た。これにより、単純にターゲット側に PCA を適用する場合でも、一定程度の効果がある事を実証出来たと考える。

#### 5.2 ソース側データ数変更実験

ソース側データ数変更実験の結果をみると、いずれの場合も精度比較の実験同様、提案手法 2 が最も精度が高いという結果になった。このことから、ソース側のデータ数が上昇した場合でも、提案手法の有効性が示せた。一方で、ソース側のデータ数が増えるにつれて、従来手法との差が小さくなっている。これは、全体としてソース側で行う PLS の与える影響が大きくなっている為だと考えられる。ターゲット側のデータが増える場合についても検討して行く必要があるだろう。

ソース側に PLS を適用しない手法 2 つ（提案手法 1 及び従来手法 1）については、データ数が少ない場合はあまり差がなく、120 の場合には従来手法 1 の方が良いという結果だが、データ数が多くなるにつれ提案手法 1 の方が良く、また精度差も大きくなっている。このことから、ソース側に PLS を適用しない場合は、提案手法によって転移学習の大規模データに対する可用性が向上している事を示している。

### 6 おわりに

#### 6.1 まとめ

本研究では、画像認識における画像とともに取得された距離情報を使用する新たな転移学習手法を提案した。また同時に、付加情報を用いた転移学習の枠組みの提案も行った。結果として、単純な従来手法よりも提案手法の方が、物体認識精度において、優れていることを示した。実験では、従来一般的な手法に対して精度が 3.49% 向上することを示した。また、データ数が増えた場合の精度比較も行い、より実用的な設定での提案手法の有効性を示した。特に、物体認識における付加情報を用いた転移学習手法の標準的な手法を確率できた。

#### 6.2 今後の課題

実用的なシステムを考えると、ターゲット側にも少数の教師データが存在する事は考えられる。本手法では、ターゲット側の教師信号は一切考慮していないが、正確な教師信号と、付加情報からの弱教師信号の両方を使用して、転移学習の精度を向上させる半教師付学習の手法を検討していく必要があるだろう。

また、現状ではひとつの付加情報をもとに分布の改善を行っているが、実際に使用する場合、音声や位置情報など、他の多くの付加情報が同時に収集できると考えられる。そういった情報を統合的に扱い、転移学習に応用していく手法を検討して行く予定である。

#### 6.3 今後の発展

近年のウェアラブルデバイスの発展に伴って、ウェアラブルデバイスでは、画像のみならず様々な情報がとれるはずであり、画像以外の付加的な情報を用いるアプローチはより一般的になってくると考えている。この研究はそういった環境で、対象とする画像に対して学習データを一切用意する必要のないアプリケーションを構築する事を目指す基盤的な研究である。実際の環境で動く画像認識を用いたアプリケーションが今まであまり提案されてこなかったのは、実環境での学習データを収集する事が難しいという問題に起因しており、この研究はその問題に対する新たなアプローチを提案している。

実用的なアプリケーションとして、現在のライフロ

グを発展させ，ライフログ映像を解析して，映像に登場する物体を認識する事で，ブログの記事の作成をアシストするシステムなどへの応用が期待出来ると考えている．

## 謝辞

本研究は，JST CREST の研究課題「複雑データからのディープナレッジ発見と価値化」の支援を得て実施された．

## 参考文献

- [1] K. Saenko, B. Kulis, M. Fritz, and T. Darrell. Adapting visual category models to new domains. In *Proc. of ECCV*, 2010.
- [2] Jason V. Davis, Brian Kulis, Prateek Jain, Suvrit Sra, and Inderjit S. Dhillon. Information-theoretic metric learning. In *ICML*, 2007.
- [3] R. Gopalan, R. Li, and R. Chellappa. Domain adaptation for object recognition: An unsupervised approach. In *Proc. of ICCV*, 2011.
- [4] B. Gong, Y. Shi, and F. Sha. Geodesic flow kernel for unsupervised domain adaptation. In *Proc. of IEEE conference on Computer Vision and Pattern Recognition*, 2012.
- [5] H. Wold, S. Kotz, and N. L. Johnson. Partial least squares. In *Encyclopedia of Statistical Sciences*, 1985.
- [6] L. Bo, X. Ren, and D. Fox. Depth Kernel Descriptors for Object Recognition. In *Proc. of IROS*, 2011.
- [7] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*, 2009.
- [8] A. Janoch, S. Karayev, Y. Jia, J.T. Barron, M. Fritz, K. Saenko, and Darrell T. A category-level 3-d object dataset: Putting the kinect to work. In *Proc. of ICCV Workshop on Consumer Depth Cameras in Computer Vision*, 2011.
- [9] D.G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, Vol. 60, No. 2, pp. 91–110, 2004.
- [10] G. Csurka, C. Bray, C. Dance, and L. Fan. Visual categorization with bags of keypoints. *Workshop on Statistical Learning in Computer Vision, ECCV*, pp. 1–22, 2004.