

壁に反射した光を使った隠れた物体の分類

白樫奈々[†]
東京電機大学[†]

鶏内朋也[‡]
慶應義塾大学[‡]

小篠裕子[†]
東京電機大学[†]

1 はじめに

物体認識の研究では、カメラから見える物体を認識することが前提となっており、完全に遮蔽された物体を認識することは十分に想定されてこなかった。しかし、カメラから隠れた物体を認識できれば、防犯、セキュリティ、監視など様々な応用が期待される。

我々は、完全に遮蔽された物体を認識するため、物体から反射した光を他の物体に反射させ、その光から隠れた物体を推測するタスクを考える。本研究では、その第一歩として、図1のように、光源から物体に入射した光を壁に反射させ、その壁を撮影することで、直接カメラから観測できない物体を分類する。本稿では、このように撮影した画像を間接光画像と呼ぶ。

間接光画像は、認識対象となる物体の形状情報をほとんど失っており、RGB画像での分類は困難である。そこで我々は壁の撮影にハイパースペクトル(以降HS)カメラを採用する。HSカメラは、波長分解能が高く、1画素に100バンド以上のスペクトル情報を保有するため、物体の形状に依存しない情報を取得できる。よって、HSカメラで撮影した間接光画像(以降、HS間接光画像)であれば物体を分類可能と考える。

HS間接光画像の分類には、HS画像の分類に有効と評価されているニューラルネットワークを用いる。一般に、分類精度の高いモデルを訓練するには、大量のラベル付きデータが必要であるが、間接光画像のラベル付けは、物体から反射した光が当たる壁の領域を目視で判断することが難しいためコストが高い。一方、近年、物体を直接撮影したHS画像の公開が増えてきたため、収集は比較的容易である。

そこで我々は、物体を直接撮影したHS画像で学習した知識を、HS間接光画像の分類に転移する。一般に、大量の教師データが準備されたデータ(ソースドメイン)で学習されたモデルを、目標のデータ(ターゲット

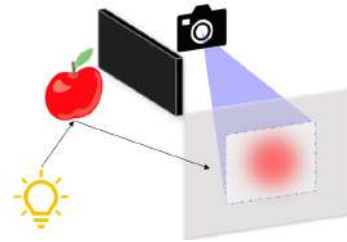


図1: タスク設定。

データ)にそのまま用いると、精度が低下することが多い。ドメインの相違がある状況で、ターゲットデータに対して高精度なモデルを学習しようと試みる手法をドメイン適応と呼ぶ。本タスクでは、物体を直接撮影したHS画像をソースデータ、HS間接光画像をターゲットデータとし、ドメイン適応を用いてHS間接光画像を分類する。

我々は、HS画像を用いた隠れた物体の分類に、ドメイン適応が有効か否かを検証する。本研究は、このタスクに挑む最初の試みであるため、検証用データセットを新たに構築する。

2 提案手法

間接光画像の空間情報から、隠れた物体を推測することは非常に困難である。我々は、空間情報に依存せず、波長情報のみから物体を分類するため、画像中の1ピクセル $x \in \mathbf{R}^d$ を入力とする画素毎分類を行う。ここで d は各画素が保持するデータの次元数である。RGB画像では $d=3$ であるが、HS画像は $d > 100$ と、高次元な波長情報を保有しているため、波長情報のみでも分類に有効な特徴を抽出することが期待できる。

物体を直接撮影したHS画像の物体領域のピクセル x_i^s と対応するラベル y_i^s のペアを $D_s = (x_i^s, y_i^s)_{i=1}^N$ 、間接光画像のピクセル x_i^t と対応するラベル y_i^t のペアを $D_t = (x_i^t, y_i^t)_{i=1}^M$ とし D_s をソースデータ、 D_t をターゲットデータと呼ぶ。 x_i^s と x_i^t は、入力空間においてそれぞれ別の分布からサンプルされたデータと見做せる。また、ターゲットデータのラベルも利用可能であ

Fully-occluded object classification from light spectra reflected off a wall

[†]Nana Shirakashi, Tokyo Denki University

[‡]Tomoya Kaichi, Keio University

[†]Yuko Ozasa, Tokyo Denki University

表 1: 分類精度.

Train data	S 100	S -	S 100	S -
	T 25	T 25	T 25	T 25
DA	✓		✓	
Instances	4	4	1	1
Accuracy(%)	81.61	89.48	47.77	72.15

る. D_t を少量しか利用できない設定でも, D_s で学習した知識を利用することで, ターゲットデータに対する認識精度を向上させる. 本研究では, 広範なデータに適用可能でかつ, end-to-end に学習できる教師ありドメイン適応手法である Motiiian らの手法 [1] を画素毎分類に拡張する.

Motiiian らが提案したモデルは, x_i^s と x_i^t をペアで入力し, 重みを共有した特徴抽出器 g を用いて, 特徴ベクトル z_i^s, z_i^t を得る. 続いて z_i^s を分類器 h に通すことで, 出力 \hat{y}_i^s を得る. この時, \hat{y}_i^s と y_i^s の分類誤差に加えて, $y_i^s = y_i^t$ の時には $d(z_i^s, z_i^t)$ が小さく, $y_i^s \neq y_i^t$ の時には $d(z_i^s, z_i^t)$ が大きくなるように誤差項を加えることで, ソースドメインのデータで得られる知識を転移する. ここで $d(\cdot)$ は特徴空間での距離を表す. 我々の提案手法では, 特徴抽出器 g と分類器 h を, HS 画像の画素毎分類で state of the art の Hu らの手法 [2] を参考に構築した. 具体的には, g は畳み込み層, プーリング層, 全結合層の 3 層で構成する. 中間層の活性化関数には \tanh 関数を用いる. h は全結合層のみで構成する.

3 実験

3.1 実験条件

実験のために構築したデータセットは, 青果物の HS 物体画像と HS 間接光画像で構成されている. 撮影には HS カメラ (COSMOS Eye) を用いた. 間接光画像の撮影状況を図 3 に示す. COSMOS Eye は, 350nm から 1050nm まで, 5nm のスペクトルの解像度で撮影することができ, 1画素当たり 141 バンドの情報を持つ. データセットは 5 種類の青果物のカテゴリから構成される. 種類毎に 5 つのインスタンスを用意し, 裏表両面から 1 回ずつハロゲン光のもとで撮影し, 物体画像 50 枚と間接光画像 50 枚の画像を用意した. モデルの訓練・評価には, 図 2 に示すように目視でマスクを生成することで物体領域と間接光領域を抽出し, その中の画素をランダムに選択して使用した. ソースドメインの物体画像には, 5 つのインスタンスのうち 4 つを学習に用いた. ターゲットドメインの間接光画像は 1 つ

のインスタンスを訓練データに使うものと 4 つを訓練データに使うものの 2 種類の実験を行った. データは全て最大値で正規化した. 訓練データは, マスク領域からランダムに 25pixel 抽出したものをを用いた. モデルは Adadelta を用いて最適化し, バッチサイズは 512 とした. また, ドメイン適応の効果を検証するための比較対象として, ターゲットドメインのデータのみを学習に用いる場合も実装した. 実験では, 過学習を防ぐために 100epoch 更新後, 検証データの損失の最低値が更新されなければ学習を停止した.

3.2 結果

5 クラス分類の実験結果を表 1 に示す. ドメイン適応を用いた手法はターゲットデータのみで学習した場合よりも精度が低かった. この理由として, ターゲットデータとソースデータの違いから起こるドメインシフトに対応しきれなかったことが考えられる. また, ドメイン適応を用いない方が分類精度が高いことから, ドメイン適応を活用する学習手法については改善の余地がある.

4 まとめ

本稿では, HS 間接光画像から物体を分類するタスクにおけるドメイン適応の有効性を検証した. 実験では十分な有効性は見られなかった. ファインチューニングや特徴変換など他のドメイン適応手法についても比較検討することを今後の課題とする.



図 2: 取得画像.

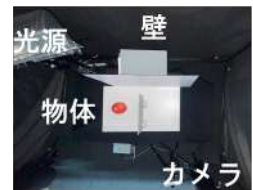


図 3: 撮影環境

参考文献

- [1] Saeid Motiiian, Marco Piccirilli, D. A. A. and Dorett, G.: Unified Deep Supervised Domain Adaptation and Generalization, *Proc. of ICCV*, (2017).
- [2] Hu, W., Huang, Y., Wei, L., Zhang, F. and Li, H.: “Deep Convolutional Neural Networks for Hyperspectral Image Classification,” *Journal of Sensors*, Vol. 2015 (2015).