

科学研究費助成事業 研究成果報告書

平成 27 年 6 月 6 日現在

機関番号：12601

研究種目：若手研究(A)

研究期間：2012～2014

課題番号：24680017

研究課題名(和文)膨大な画像データを用いた超多クラス一般物体認識システムの構築

研究課題名(英文)Super multi-class object recognition system using a large amount of image data

研究代表者

原田 達也 (Harada, Tatsuya)

東京大学・情報理工学(系)研究科・教授

研究者番号：60345113

交付決定額(研究期間全体)：(直接経費) 20,400,000円

研究成果の概要(和文)：本研究の目的は、膨大な画像データとテキストデータから、それらの関係性を統計的に学習し、超多クラスの一般的物体認識可能なシステムを構築することにある。この実現には、1)膨大なデータであっても破たんすることなく学習し続けられる手法が重要となる。また、2)複数の物体が存在しても、どこに何があるのか検出する手法も重要となる。3)一般的に教師データ作成の人的コストは高く、この学習データ構築コストを減らすことが、正確かつ多様な物体認識に重要となる。さらに、4)分からないものが分かる技術は、自律的に学習していく認識システムに必要不可欠である。本研究課題では、上記1)～4)に関して取り組み成果を得た。

研究成果の概要(英文)：The goal of our research is the construction of the super multi-class generic object recognition system by learning the relationship between a large amount of image and text data statistically. A method to continuously learn the classifiers from a huge amount of data without breakdown is crucial to realize this system. If there are many objects in one image, it is important to recognize where and what they are. A cost to construct high quality training dataset is so expensive that reducing the construction cost is also crucial. Moreover, a technique to find novel classes is a bottleneck for the continuously growing recognition system. In this research, we have tackled the above mentioned topics and produced some results.

研究分野：知能機械情報学

キーワード：画像認識 コンピュータビジョン 機械学習 人工知能

1. 研究開始当初の背景

画像から一般的な物体やシーンを認識することはコンピュータビジョンの究極の目標の一つである。一般的な物体・シーンを認識する分野は「一般物体認識」と呼ばれコンピュータビジョンの中でも競争が熾烈であり、その進化のスピードは目を見張るものがある。一般物体認識の発展は扱うクラス数の多さによって理解できる。2006年から2009年までは100クラスのオーダーであり、2010年から1000クラスを扱う研究がいくつか出現している。1000クラスのオーダーを高精度に認識可能なシステムは残念ながら多くない。なぜなら、多くの研究が100クラス程度のトイワールドに特化した手法となっており、扱うべきクラス数やデータ量が増えたときに計算が破綻してしまうためである。このようにコンピュータビジョンでは1000クラスの識別問題に取り組みはじめているが、たとえ1000クラスであったとしても実世界に存在する物体・シーンの多様性に対応できているとは言い難い。さらに実世界の多様性を表現するために対象クラス数を増やした際に、現在開発されている1000クラスレベルの認識手法が同様に有効に働く保証はない。

2. 研究の目的

本研究の目的は、Web等に存在する膨大な画像データとテキストから、それらの関係性を統計的に学習し、超多クラスの一般的な物体・シーンを認識可能な知的システムを構築することにある。現在実用化されている画像認識システムはデジカメに搭載されている顔認識システムに代表されるように一つもしくは少数のクラスのみに対応可能である。また一般物体認識の研究であっても平均的に100クラス、最大でも1000クラスを扱うに止まり、実世界に存在する物体・シーンの多様性に対応できているとは言い難い。本研究課題では、言葉で記述できる対象であれば認識可能とする超多クラス一般物体認識をめざし、その方法論の確立を目指す。

3. 研究の方法

膨大な多クラスの物体認識を行うには、まず、膨大なデータであったとしても破たんすることなく学習し続けられる識別器の学習手法が重要となる。また、画像中に複数の物体が存在したとしても、それぞれどこに何があるのか検出する手法の確立も重要となる。一般的に、教師データを与える人的コストは高く、この学習データの構築する労力を減らすことが、正確かつ多種多様な物体認識に重要となる。さらに、分からないものが分かる技術は、自律的に学習していく認識システムに必要不可欠であり、この困難な課題にも取

り組む必要がある。そこで本研究課題では上記の点を念頭に置き、1) 画像認識におけるオンライン学習、2) 多クラス物体検出、3) 物体セグメンテーションの弱教師付き学習、4) 異常検出、の4点を中心に取り組むこととした。

4. 研究成果

(1) 画像認識におけるオンライン学習

画像データが大量になると、全てのデータが主メモリに載らなくなるために、データを一括して学習するのではなく、少数のデータを取り込んで逐次的に学習するオンライン学習が重要となる。多種多様なオンライン学習手法が機械学習や自然言語処理の分野で提案されている。しかしながら、それらの手法を画像認識で利用される特徴に適用して比較評価されることはなかった。そこで、本課題では、線形識別器における様々なオンライン学習手法を、近年よく利用される画像特徴に適用することで実験的に比較し、画像認識におけるオンライン線形識別器のガイドラインを導いた。また、画像認識の観点から、これらのオンライン線形識別器手法の統一的な解釈を与えた。

得られた大規模画像認識のためのオンライン学習のガイドラインは以下の通りである。

- A) パーセプトロンでも最新のオンライン学習手法と同等の性能を得ることができる。
 - ただし、次のB)にあげるガイドラインをパーセプトロンに適用した場合に有効である。
- B) 平均化のテクニックがすべてのアルゴリズムに有効である。
 - 平均化のテクニックを用いない1次のアルゴリズムは2次のアルゴリズムと比較して性能が劣る。
 - 平均化のテクニックを用いると1次のアルゴリズムも2次のアルゴリズムも同等の性能が得る。
 - 平均化のテクニックは1次のアルゴリズムも2次のアルゴリズムも収束を速くする。
- C) マルチクラスの設定をはじめに考慮すべきである。
 - 1クラス対その他クラスの設定とマルチクラス設定とでは同等の識別性能である。
 - ただし、1クラス対その他クラスの設定はマルチクラス設定よりもより長いCPU時間が必要である。

また、図1に画像認識における線形識別器のオンライン学習の比較を示す。各手法の棒グラフの左側が平均化テクニックなしの場合であり、右側が平均化テクニックを適用した場合である。

Averaging Does Boost All

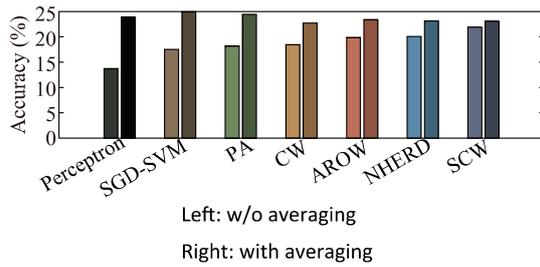


図 1：画像認識における線形識別器のオンライン学習の比較

(2) 多クラス物体検出

本課題では、大規模データを用いて、マルチクラスの物体検出器を同時に最適化する効率的な手法を提案し実現した。従来の物体検出器の学習は、各検出対象物体を正例サンプル、それ以外の物体を負例サンプルとして識別境界を決定する「1 クラス対その他クラス」のアプローチをとるものが主流であった。しかしながら、この方法では各クラスを独立に学習するため、異なるクラス間のスコアのバランスを調整できない。提案手法は、マルチクラスの識別手法を応用してマルチクラスの物体検出器を同時に学習することで、クラス間のバランスを最適化する。このとき、学習対象物体クラス間の差だけでなく、その他の大量の背景画像と各クラスとの差を考慮することで、未知物体の誤検出を抑える。実験では、大規模一般物体認識コンペティション ILSVRC 2011 で用いられた大量データセットのサブセットによる評価を行い、提案手法の有効性を示した。また、この手法は ILSVRC2014 でも利用されて、30 以上のチームが参加する中、4 位の成績を収めた。図 2 に本研究で実現した物体検出システムの出力例を示す。



図 2：物体検出の結果

(3) 物体セグメンテーションの弱教師付き学習

実環境を撮影した色・距離画像のペアと、各画像ペアにつけられた物体ラベル情報を用いて、画像中の各物体を自動的にセグメンテーションし、物体検出器を学習する手法を

提案し実現した。昨今、バウンディングボックスの教師情報を使わず、画像単位でつけられた物体ラベルのみを用いて物体検出器を学習する手法が盛んに研究されているが、色・距離画像を用いた研究例はない。色画像のみならず色・距離画像のペアを用いることで、空間的な連続性を考慮したセグメンテーションが可能になり、ひとまとまりの物体の候補を絞りやすくなる。また、二次元画像の特微量に比べて視点変化に頑健である三次元特微量を抽出することができるため、学習対象物体の正例画像群に共通して存在する物体の発見が容易になる。

そこで本研究課題では、色・距離画像の法線ベクトル推定に基づいたセグメンテーションとボクセルベースの三次元特微量を用いて Multiple Instance Learning を行うことで、バウンディングボックスを必要とせずに物体を学習する新規な手法を提案した。提案手法は、二次元画像のみを用いた弱教師付き学習による物体検出の最先端手法では完全に失敗するような物体の見えの変化の激しいデータセットにおいて、良い性能を示した。また三次元特微量として色テクスチャ、グレースケールテクスチャ、表面形状曲率の局所特徴を比較し、さらに、これらを組み合わせることで性能が向上することを確認した。

図 3 に弱教師付き学習により得られた物体検出結果を示す。中央の列の写真が提案手法の結果であり、左右の列の写真が比較した従来手法の結果である。青い四角が誤って検出された領域であり、赤い四角が正しく検出された領域を示している。



図 3：弱教師付き学習により得られた物体検出結果

(4) 移動体からの異常検出

実環境中に存在する新規な物体などを発見する機能はクラス数を限定しない物体認識において必要不可欠な機能である。新規な物体を実世界から発見する基盤技術として異常検出があげられる。特に自動車やロボットなどの移動体からの異常検出はニーズが高い。従来の移動体からの視覚的な異常検出には、大量のデータから通常モデルを作成し、そこからの逸脱を異常とする統計的異常検出の立場と、ある地点における観測画像をその地点における過去の観測と比較する画像比較型異常検出の立場があった。しかし、移動体は様々な環境を移動するため個々の環境において統計的に十分な画像を集めることは難しい。また、画像比較型の異常検出では同一地点の観測のみ参照するため照明条件などの環境の変化に柔軟に対応することができない。

そこで本課題では、同一地点における過去からの変化を異常検出の対象とすることにより、環境の変化に柔軟な異常検出の手法を提案する。その際に、通常データの組み合わせにより対象データの再構成が行えるかを異常の基準とする異常検出手法を導入し、移動体におけるサンプル数不足の問題の解消を試みる。まず、固定視点からの映像を用いた実験により使用可能な画像特徴量の比較を行った。そして、移動ロボットを用いて収集したデータを用いた実験により提案手法が少サンプル環境において従来手法では検出が不可能な異常の存在を検出することができることを示した。

5. 主な発表論文等

[雑誌論文] (計9件)

- ① Yoshitaka Ushiku, Masatoshi Hidaka, Tatsuya Harada. Three Guidelines of Online Learning for Large-Scale Visual Recognition. In The Twenty-Seventh IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2014), pp. 3574-3581, 2014. 査読有.
10.1109/CVPR.2014.457
- ② Asako Kanezaki, Sho Inaba, Yoshitaka Ushiku, Yuya Yamashita, Hiroshi Muraoka, Yasuo Kuniyoshi, Tatsuya Harada. Hard Negative Classes for Multiple Object Detection. 2014 IEEE International Conference on Robotics and Automation (ICRA 2014), pp. 3066-3073, 2014. 査読有.
10.1109/ICRA.2014.6907300
- ③ Emanuele Rodolà, Andrea Torsello, Tatsuya Harada, Yasuo Kuniyoshi, Daniel Cremers. Elastic Net

Constraints for Shape Matching. In the 14th International Conference on Computer Vision (ICCV 2013), pp. 1169-1176, 2013. 査読有.
10.1109/ICCV.2013.149

- ④ Emanuele Rodolà, Tatsuya Harada, Yasuo Kuniyoshi, and Daniel Cremers. Efficient Shape Matching using Vector Extrapolation. In The British Machine Vision Conference (BMVC 2013), pp. 91.1-91.11, 2013. 査読有.
<http://dx.doi.org/10.5244/C.27.91>
- ⑤ Asako Kanezaki, Yasuo Kuniyoshi, and Tatsuya Harada. Weakly-supervised Multi-class Object Detection Using Multi-type 3D Features. Proceedings of the 21th Annual ACM International Conference on Multimedia (ACMMM 2013), pp. 605-608, 2013. 査読有.
10.1145/2502081.2502159
- ⑥ Yuya Yamashita, Tatsuya Harada, and Yasuo Kuniyoshi. Causal Flow. IEEE Transactions on Multimedia, Vol. 3, No. 3, pp. 619-629, 2012. 査読有.
10.1109/TMM.2012.2191396
- ⑦ Yoshitaka Ushiku, Tatsuya Harada, and Yasuo Kuniyoshi. Efficient Image Annotation for Automatic Sentence Generation. The 20th Annual ACM International Conference on Multimedia (ACMMM 2012), pp. 549-558, 2012. 査読有.
10.1145/2393347.2393424
- ⑧ Hiroharu Kato, Tatsuya Harada, and Yasuo Kuniyoshi. Visual Anomaly Detection from Small Samples for Mobile Robots. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2012), pp. 3171-3178, 2012. 査読有.
10.1109/IROS.2012.6386031
- ⑨ Fumihiro Bessho, Tatsuya Harada and Yasuo Kuniyoshi. Dialog System Using Real-Time Crowdsourcing and Twitter Large-Scale Corpus. Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL2012), pp. 227-231, 2012. 査読有.
<http://dl.acm.org/citation.cfm?id=2392800>. 2392841

[学会発表] (計10件)

- ① Tatsuya Harada. Bridging the Gap between Visual Contents and Natural Language. The First Kyoto University-Inamori Foundation Joint Kyoto Prize Symposium, 2014. 7/13/2014. Kyoto University (Kyoto,

- Kyoto-shi)招待講演.
- ② Atsushi Kanehira, Masatoshi Hidaka, Yusuke Mukuta, Yuichiro Tsuchiya, Tetsuaki Mano, and Tatsuya Harada. MIL at ImageCLEF 2014: Scalable System for Image Annotation. CLEF 2014 Evaluation Labs and Workshop, pp. 372-379, 2014. 2014/9/16, (Sheffield, England)
 - ③ 椋田悠介, 原田達也. 確率的偏正準相関分析. 信学技報, vol. 113, no. 286, IBISML2013-58, pp. 169-176, 2013. 2013/11/13, 東京工業大学 (東京都・目黒区)
 - ④ 鎌田智恵, 原田達也. マルチモーダル情報の相互情報量に基づく異常検出に関する研究. 信学技報, vol. 113, no. 196, PRMU2013-48, pp. 167-172, 2013. 2013/9/3, 鳥取大学 (鳥取県・鳥取市)
 - ⑤ 日高雅俊, 郡司直之, 原田達也. ラベル間の階層構造を考慮したWeb画像アンテーション手法に関する研究. 信学技報, vol. 113, no. 196, PRMU2013-52, pp. 201-206, 2013. 2013/9/3, 鳥取大学 (鳥取県・鳥取市)
 - ⑥ 牛久祥孝, 原田達也, 國吉康夫. キーフレーズ推定と文法モデルによる画像説明文生成. 画像の認識・理解シンポジウム (MIRU2012), OS3-01, 2012. 2012/8/6, 福岡国際会議場 (福岡県・福岡市)
 - ⑦ 稲葉翔, 村岡宏是, 山下裕也, 牛久祥孝, 金崎朝子, 原田達也, 國吉康夫. 学習時間に着目した効率的な大規模画像分類. 画像の認識・理解シンポジウム (MIRU2012), OS6-03, 2012. 2012/8/8, 福岡国際会議場 (福岡県・福岡市)
 - ⑧ 金崎朝子, 原田達也, 國吉康夫. 多種類の三次元特徴量を用いた物体セグメンテーションの弱教師付き学習. 画像の認識・理解シンポジウム (MIRU2012), OS4-13, 2012. 2012/8/8, 福岡国際会議場 (福岡県・福岡市)
 - ⑨ 水落大, 原田達也, 國吉康夫. 大規模コーパスと対話の相互作用を活用した感情遷移推定. 第30回日本ロボット学会学術講演会予稿集, , 9月19日, 3D1-7, 2012. 札幌コンベンションセンター (北海道・札幌市)
 - ⑩ 金崎朝子, 稲葉 翔, 牛久祥孝, 山下裕也, 村岡宏是, 原田達也, 國吉康夫. 大規模画像データセットを用いたマルチクラス物体検出器の同時学習 ～ 物体毎に特化した負例クラスの導入 ～. 信学技報, vol. 112, no. 197, PRMU2012-42, pp. 105-112, 2012. 2012/9/3, 東京農工大 (東京都・小金井市)

[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
<http://www.mi.t.u-tokyo.ac.jp/>

6. 研究組織

(1) 研究代表者

原田 達也 (TATSUYA, Harada)
東京大学・大学院情報理工学系研究科・教授
研究者番号：60345113