様式Ｆ－１９

科学研究費助成事業（学術研究助成基金助成金）研究成果報告書

平成 25 年 6 月 3 日現在

機関番号：６２６１５
研究種目：挑戦的萌芽研究
研究期間：2011～2012
課題番号：23650093
研究課題名（和文）　物体領域の推定と物体認識・検出モデルの学習の同時最適化
研究課題名（英文）　Joint Optimization of Object Region Localization and Recognition/Detection Model Learning

研究代表者
佐藤 真一（SATOH Shin’ichi）
国立情報学研究所・コンテンツ科学研究系・教授
研究者番号：90249938

研究成果の概要（和文）：本研究では、写っているもののラベルが付いた大量の画像群から、対象物体の領域を自動抽出すると同時に､物体認識･検出器を学習する手法について検討する。学習用画像は、対象物体の外接矩形や物体領域を与える必要はなく、画像に対するラベルが与えられればよく、人手でも容易に付与可能であり、Flickr や文字字幕つき映像などからも大量に取得可能である。物体領域を自動抽出することにより、画像合成などにも利用できる上、より精密な認識・検出器の実現が可能となる。

研究成果の概要（英文）：We addressed a method to locate target objects and to learn object recognition and detection models simultaneously from given large collection of images with labels of target objects. The method does not require bounding rectangles or object regions of training images but just object labels, which are easily prepared by manual annotation, or easily obtained from Flickr or closed captioned videos. Object region localization functionality enables image synthesis and more precise object recognition and detection.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
交付決定額	2,800,000	840,000	3,640,000

研究分野：総合領域
科研費の分科・細目：情報学 知覚情報処理・知能ロボティクス
キーワード：一般画像認識, 物体領域検出

１．研究開始当初の背景

画像中の任意の物体の認識は古くから試みられているがいまだに解決を見ず、もし実現できればマルチメディア検索等きわめて多くの応用が考えられる。近年一般物体認識技術として研究が盛んであり、大量のラベル付き画像から物体認識システムを学習する方法が最も有望と考えられる。

これらのアプローチは図１のように分類でき、主として 1) ラベルは画像に与えられ、物体領域を特定しない特徴量を利用する方法(通常の Bag of Feature 法)、2) ラベル、特徴量とも物体を囲む矩形領域を利用する方法(PASCAL VOC の認識・検出タスク)、3) 物体領域に対するラベルを利用する方法(PASCAL VOC のセグメンテーションタスク、MIT LabelMe プロジェクト) がある。一般に画像に対するラベルは準備しやすいが認識精度に問題があり、物体領域に対するラベルは構築に大変な手間がかかるが認識精度は高くなる。これまで、画像に対するラベルのみで、物体領域に対するラベルつき画像と同等の認識精度を実現した方法はない。これに対し、本研究では、ラベルつき画像群のみから物体領域の自動抽出を行うと同時に高精度の物体認識を行う手法を検討する

(図中の 4) に相当)。

[Nguyen et al. 2009] では、Multiple Instance Learning(MIL) を用いて、ラベル付き画像群から物体の外接矩形を推定しており、本研究の着想に近いが、莫大な探索領域のため一般の物体領域抽出には対応できず、また矩形領域のため認識精度に問題がある。[Russell et al. 2006, Hoiem et al. 2005] は、パラメータを変えて複数のセグメンテーション結果を生成させておき、それらの中から認識精度等により適当なセグメントを選ぶという戦略を取っているが、物体領域は生成されたセグメントのいずれかとなり、高精度の領域抽出は実現できない。[Todorovic and Ahuja 2006] は、複数の詳細さのセグメンテーションを行い、セグメント間の包含関係に基づきセグメンテーション木を構築し、同一ラベルを持つ画像間で共通している部分木を物体領域とする方法だが、本質的に生成したセグメントのグルーピングにより物体領域を生成する方法であり、やはり精度に問題がある。

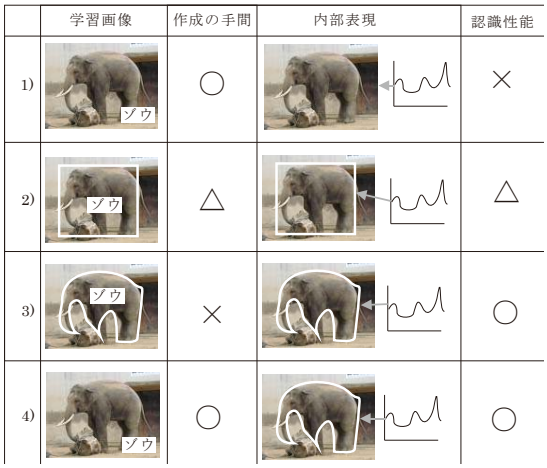

図 1. 学習用画像とシステム内部表現

２．研究の目的

大量のラベル付き画像から物体領域が自動抽出できること、こうして得られた物体領域情報により高精度の物体認識が可能なこと、これらを同時に最適化する定式化が可能であること、これにより物体抽出並びに物体認識システムの学習が同時に最適化でき、最も効果的な物体認識システムが実現可能であることを明確にする。

物体領域の自動抽出ができると、画像合成用の物体モデルが容易に構築可能となる。また、実現される物体認識手法は、大量マルチメディア情報の検索、マルチメディアマイニング、大量マルチメディア情報に基づく情報分析に有効であり、マルチメディア情報の意味解析のためのブレークスルーとなりうる。

３．研究の方法

われわれは、[Nguyen et al. 2009]と同様、この問題に対して Multiple Instance Learning(MIL)を利用することとした。すなわち、学習において、ラベルの付与されている画像を bag とみなし、その画像におけるすべての可能な物体領域を instance とする。MIL では、正のラベルの付いた bag (すなわち、ある特定の物体が画像中に存在する)には、少なくとも一つの instance が正のラベルに対応している(すなわち、少なくとも一つの領域がその物体に対応している)ことになり、負のラベルの付いた bag (すなわち、その画像にはその物体は存在しない)には、いずれの instance も正のラベルとならない(すなわちどの領域も物体に対応しない)ことになる。適切な MIL アルゴリズムにより、学習用画像には画像に対するラベルが与えられた状態から、物体領域に対する適切なラベルが推定され、高精度の物体レベルの認識・検出モデルが学習可能となることが期待される。

一般に MIL アルゴリズムには、Diverse Density(DD)アルゴリズムをはじめとする生成型手法と、MILES, mi-SVM, MI-SVM 等の識別型手法とが存在するが、最終的な識別性能の点では、識別型手法が有利であることが知られている。そこで、われわれも識別型手法に基づいた方法をとる。

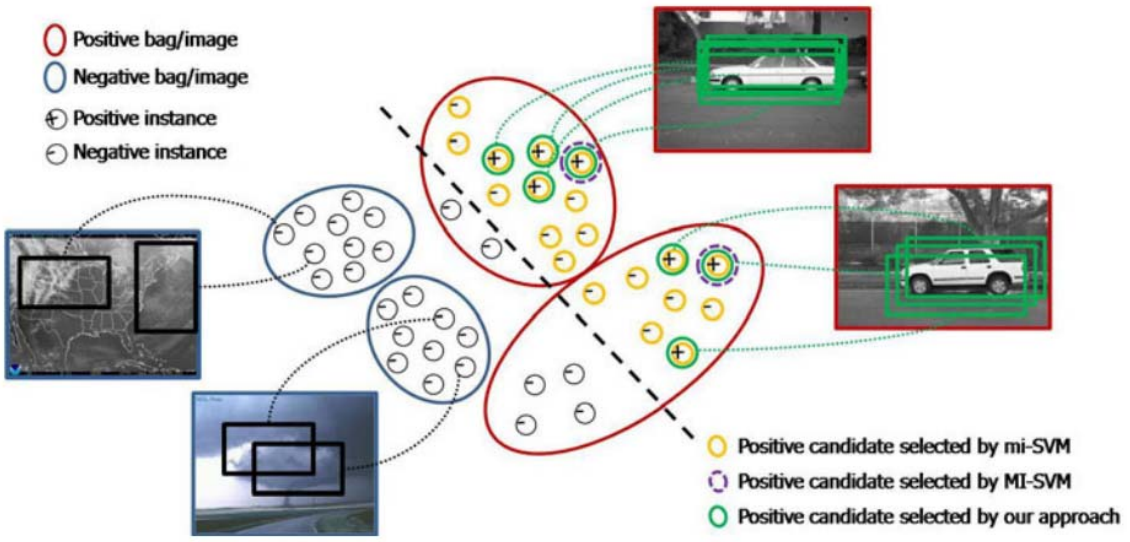

図 2. 物体領域に対応した MIL アルゴリズム

一方、こうした MIL アルゴリズムでは、各 instance のラベルは独立に扱われる。しかしながら、われわれの問題設定では、instance は物体領域に相当しており、例えば相互に重なり合うような場合には、こうした物体領域に対応する instance は同じラベルとなる可能性が非常に高いと考えられる。

そこで、われわれは、MIL アルゴリズムを拡張し、新たに、同じくらいの大きさで重なりの大きい instance どうしのラベルは同じになりやすいという拘束条件を追加した。概要を図 2 に示している。SVM の拘束式を拡張し、MIL に対応した上、instance 間の物体領域としての拘束条件を考慮した拘束式を以下に示す。

$$\min_{\{y_i\}} \min_{\{w,b,\xi\}} \quad \frac{1}{2}\|w\|^2 + C\sum_I \xi_I$$

$$\text{subject to} \quad \forall I:\ Y_I = -1 \wedge -\langle w, x_i\rangle - b \geq 1 - \xi_I\,,\ \forall i \in I\,,$$

$$\text{or} \quad Y_I = 1 \wedge \langle w, x^*\rangle + b \geq 1 - \xi_I\,,$$

$$\forall x^* \in \mathcal{SR}(x_{mm(I)}, T)\,, 0 < T \leq 1, \text{ and } \xi_I \geq 0$$

x_i は i 番目の instance に対応する特徴量、y_i はそのラベルであり、ξ_i は対応するスラック変数、他 C, b は SVM のパラメータである。SR の部分が物体領域の重なりを評価する項であり、T によりどの程度の重なりがあった場合に同じラベルとみなすかを制御することができる。これは、SVM と同様に二次計画問題として解くことが可能である。そのアルゴリズムを図 3 に示す。
詳細については、学会発表[1, 2]を参照いただきたい。

Pseudo code for heuristic algorithm

```
Initialize: for every positive bag B_I
  Compute x_I = Σ_{i∈I} x_i/|I|.
  SR_I = x_I.
REPEAT
 - Compute QP solution w,b for dataset with positive
     samples {SR_I : Y_I = 1} and negative samples {x_i : Y_I = −1}.
 - Compute outputs f_i = ⟨w, x_i⟩ + b for all x_i in positive bags.
 - FOR (every positive bag B_I)
     Set x_I = x_{mm(I)}, mm(I) = arg max_{i∈I} f_i
     SR_I = FindSurround(x_I, T)
 - END
WHILE ({mm(I)} have changed)
OUTPUT (w, b)
```

図 3. アルゴリズム

物体認識のためのベンチマーク画像セットである Caltech-101 を利用して性能評価を行っている。評価の結果、提案手法は[Nguyen et al. 2009]を含む従来手法の識別性能を上回る性能を達成したことが確認されている。識別性能の抜粋を図 4 に示す。MA は手で物体を含む矩形領域が与えられて学習した場合を示し(図 1 の 2)に相当)、GH は画像全体で学習したものを表し(図 1 の 1)に相当)、mi-SVM 並びに MI-SVM は従来手法に相当し、Ours が提案手法を表している。興味深いことに、MA よりも GH が高い性能を表し、特にこのデータセットでは、物体領域外の背景領域の情報が物体識別に有効であることが示されているが(こうした事実は最近他の研究者らによっても指摘されている[Liu and Wang 2012])、提案手法が全体的にいずれの手法よりも高い識別性能を達成していることが示されている。

	MA	GH	mi-SVM	MI-SVM	Ours
Butterfly	76.7	76.7	53.3	86.7	**93.3**
Camera	70.0	80.0	53.3	73.3	**86.7**
Ceiling_fan	70.0	**80.0**	53.3	66.7	**80.0**
Cellphone	80.0	**90.0**	63.3	83.3	**90.0**
Laptop	80.0	76.7	66.7	76.7	**86.7**
Motorbikes	73.3	**93.3**	63.3	80.0	90.0
Platypus	83.3	90.0	53.3	86.7	**100.0**
Pyramid	**90.0**	**90.0**	63.3	76.7	**90.0**
Tick	76.7	83.3	56.7	80.0	**90.0**
Watch	**80.0**	**80.0**	53.3	73.3	**80.0**

図 4. 提案手法の性能評価

４．研究成果

われわれは、写っているもののラベルが付いた大量の画像群から、対象物体の領域を自動抽出すると同時に、物体認識・検出器を学習する手法について検討した。この問題に対して Multiple Instance Learning(MIL)を適用し、かつ MIL に対して物体領域識別に特化した拡張を施した手法を提案した。本手法により、他の既存手法よりも高い認識性能を達成することができ、この点では当初予定よりも高い成果を上げることができた。

一方、本手法では物体を含む矩形領域を抽出するようになっており、当初予定である物体領域の正確な推定とはなっていない。物体領域の推定のため、階層型セグメンテーション手法[Sande et al. 2011]を用い、分枝限定法により効率よく目標の物体領域を探索する手法を開発したが(本成果は未発表)、認識精度の向上にはつながらなかった。これは、正確な物体領域の推定は認識精度の向上には必ずしも直結しないという最近の他の研究者らの知見[Liu and Wang 2012]とも呼応し、われわれもこうした事実にいち早く気が付いていたことになる。

このように当初の研究予定については一通り検討を行っており、高い認識性能を達成すると同時に、次の研究にもつながる知見も得ており、十分な研究成果を上げたと考えている。

〔参考文献〕

[Nguyen et al. 2009] M. H. Nguyen, L. Torresani, F. D. la Torre, and C. Rother, “Weakly supervised discriminative localization and classification: A joint learning process,” in Proc. of ICCV, pp. 1925–1932, 2009.

[Russell et al. 2006] B. C. Russell, W. T. Freeman, A. A. Efros, J. Sivic, and A. Zisserman, “Using multiple segmentations to discover objects and their extent in image collections,” in Proc. of CVPR, vol. 2, pp. 1605–1614, 2006.

[Hoiem et al. 2005] D. Hoiem, A. A. Efros, and M. Hebert, “Geometric context from a single image,” in Proc. of ICCV, 2005.

[Todorovic and Ahuja 2006] S. Todorovic and N. Ahuja, “Extracting subimages of an unknown category from a set of images,” in Proc. of CVPR, pp. 927–934, 2006.

[Sande et al. 2011] K. E. A. van de Sande, J. R. R. Uijlings, T. Gevers, and A. W. M. Smeulders. Segmentation as selective search for object recognition. ICCV, 2011.

[Liu and Wang 2012] L. Liu and L. Wang, What has my classifier learned? Visualizing the classification rules of bag-of-feature model by support region detection, CVPR, 2012.

５．主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 3 件）

[1] 佐藤 真一, ``画像・映像意味解析技術の過去、現在と今後の可能性,'' 電子情報通信学会誌, Vol.96, No.1, pp.58-59, 2013.

[2] 佐藤真一, ``大規模データに基づく画像・映像意味解析,'' 映像情報メディア学会誌, Vol.66, No.11, pp.887-890, 2012.

[3] 佐藤 真一, ``マルチメディア内容解析技術による放送映像アーカイブの構造化,'' 電子情報通信学会誌, Vol.95, No.1, pp.68-73, 2012.

〔学会発表〕（計 2 件）

[1] Thanh Duc Ngo, Duy-Dinh Le, and Shin'ichi Satoh, ``Boosting global scene classification accuracy by discriminative region localization,'' International Conference on Image Processing (ICIP2011), Sep. 11-14 (presented on Sep. 12), Brussels, Belgium, 2011.

[2] Thanh Duc Ngo, Duy-Dinh Le, and Shin'ichi Satoh, ``Improving Image Categorization by Using Multiple

Instance Learning with Spatial Relation," International Conference on Image Analysis and Processing (ICIAP 2011), Sep. 14-16 (presented on Sep. 14), Ravenna, Italy, 2011.

〔図書〕（計１件）

[1] Xiaomeng Wu, Sebastien Poullot, and Shin'ichi Satoh, ``Multimedia Duplicate Mining toward Knowledge Discovery," in Frank Y. Shih ed., ``Multimedia Security: Watermarking, Steganography, and Forensics," pp. 3-30, Taylor & Francis Group, CRC Press, 2012, 423 pages.

〔その他〕

[1] 佐藤真一, ``マルチメディア解析・検索研究のための大規模コーパスの動向," 画像ラボ, Vol.22, No.6, pp.19-26, 2011.

６．研究組織

(1)研究代表者

佐藤 真一 (SATOH Shin'ichi)
国立情報学研究所・コンテンツ科学研究系・教授
研究者番号：90249938

(2)研究分担者

なし

(3)連携研究者

なし