

ウェブ画像を用いたカテゴリ別 Visual Words による未知物体判別

田中 雄翔[†] 滝口 哲也^{††} 有木 康雄^{††}[†] 神戸大学大学院システム情報学研究科 〒657-8501 兵庫県神戸市灘区六甲台町 1-1^{††} 神戸大学自然科学系先端融合研究環 〒657-8501 兵庫県神戸市灘区六甲台町 1-1E-mail: [†]tytanaka@me.cs.scitec.kobe-u.ac.jp, ^{††}{takigu, ariki}@kobe-u.ac.jp

あらまし 本論文では、既知の物体と同様に、未知の物体も判別できるカテゴリ別 Visual Words を提案する。最も広く用いられている物体認識の手法は、Bag of Features(BoF) 手法である。これは、SIFT(Scale-Invariant Feature Transform) などの局所特徴を量子化することによって、Visual Words と呼ばれるコードブックを作成し、その出現頻度ヒストグラムとして画像を表現する手法である。しかし、この手法には既知の物体にしか適用できないという問題点がある。従って、BoF 手法は未知の物体を含む物体認識/照合問題に適している手法とはいえない。この観点から、本論文は未知のカテゴリの物体も表現することができるカテゴリ別 Visual Words と、それによる物体認識/照合手法を提案する。10 クラスの物体認識において、提案手法は従来の BoF 手法より 8.0%の精度が得られた。

キーワード 一般物体認識, SIFT, Bag of Features

1. はじめに

一般物体認識は、実世界において一般的な名称でコンピュータが物体を認識することである。これはコンピュータビジョンにおいて最も困難な課題の一つである。さらに、近年高解像度デジタルカメラや、大容量の HDD の登場により、膨大な量の動画や画像を分類したり、検索したりすることが困難になりつつある。それゆえに、それらの膨大な動画や画像を自動的に分類し、検索することができるようになることがコンピュータに求められている。この観点からも、一般物体認識は今後さらに重要になっていく課題である。

一般物体認識で広く用いられているアプローチは、Bag of Features(BoF) [1] である。このアプローチは画像中の物体のカテゴリを認識するものである。BoF は、物体の画像から SIFT(Scale-Invariant Feature Transform) [2] [3] や SURF(Speed Up Robust Features) のような局所特徴を抽出し、それらを k-means 手法によって W 個のクラスタに分類する、物体の見えを元にした手法である。それぞれのクラスタのセントロイドとなるベクトルのことを Visual Word と呼び、その数は経験的に決定される。この方法では、物体の画像は Visual Words の出現頻度ヒストグラムによって表現される。

BoF による物体の画像表現は、物体のオクルージョンに頑健である。なぜなら BoF は、局所特徴の集合体としての表現であり、また k-means によるベクトル量子化をすることによって、見え方の変化に強いためである。しかし、画像中にカテゴリとは関係のない特徴が多く含まれていると信頼性が失われたり、学習に用いる画像セットを用意するのが困難といった問題点もある。

図 1 は、BoF 手法の概略図である。図において、学習

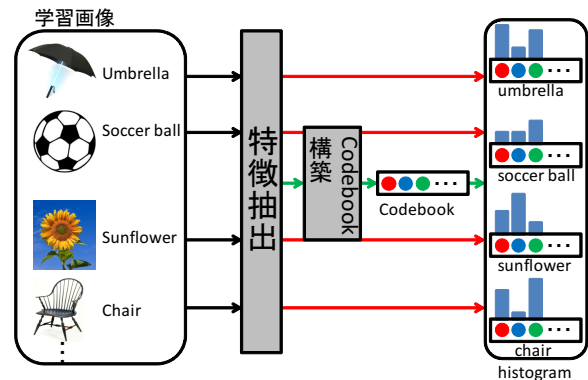


図 1 従来の BoF 手法のフロー図

画像は、認識対象となるさまざまなカテゴリの物体画像であり、SIFT などの局所特徴量が抽出される。次に、得られた特徴から k-means 法などによって特徴量のクラスタリングを行い、得られたクラスタを Visual Words とすることで、コードブックが作成される。最後に、物体画像の特徴はそれぞれ、コードブックのどの Visual Words に属するか判定され(基本的にはユークリッド距離が最も近いものが選ばれる)、ベクトル量子化される。従って、画像 1 枚に対して 1 つのヒストグラムが得られるので、それらを学習データとすることで識別器が構築でき、認識が可能となる。しかし、この BoF 手法には、学習画像に含まれなかったカテゴリの物体(未知物体)は認識できないという問題点がある。

図 2 は、ある物体 A が既知か未知かを判定するシステムのフロー図を示したものである。図において、ユーザが「A はどこですか?」と問いかける状況を考える。もしシステムが A を知っている(既知)場合には、システムはその物体を持ってきて、「これが A です。」と答える

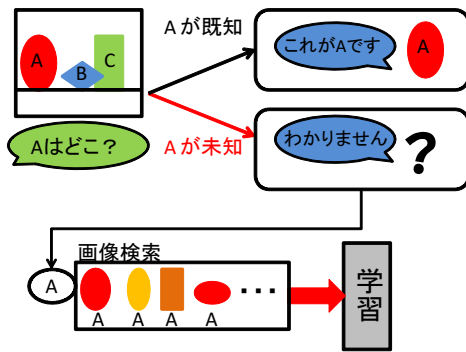


図 2 既知/未知物体に対するフロー図

ことができる。しかし、システムが A について何も知らなければ (未知), 答えることはできない。この問題を解決するためには、システムが A について学習をしなければならない。しかし、未知物体 A に関する情報は、何も持ち得ていないというのが普通である。そこで、A についての情報を得るために、ウェブから情報を得ることを考える。

ウェブには大量の画像が存在し、画像にはそれぞれタグが付けられている。A という名称をもとに検索を行い、検索したタグをラベルとして学習させることで、擬似的に教師あり学習を実行できる。これにより、システムが A という未知物体を学習して取ってくるというタスクが達成できる。

物体の知識表現としては、先ほど述べた BoF 手法を用いる。この手法では、既知の物体を表現するために Visual Words からなるコードブックを構築している。従って、本研究のタスク設定のように、未知の物体を認識する場合、BoF 手法では、既知の物体の画像特徴量に未知の物体の画像特徴量を加えて、クラスタリングを行い、新たに Visual Words を作り直さなければならない。これは手間がかかることであり、認識対象が増加していくと、認識率は低下していくという問題もある。この観点から、BoF 手法は未知のカテゴリの物体を認識するのには適していないといえる。この問題を解決するために、本論文では、既知の物体と同様に未知の物体も認識できるような、カテゴリ別 Visual Words による画像認識/照合の手法を提案する。

本論文は次のように構成されている。2 章で、提案手法のカテゴリ別 Visual Words による画像認識/照合の手法を述べる。3 章で、10 クラスの画像データセットを用いて認識/照合実験を行い提案手法を評価する。4 章で、論文のまとめと今後の課題について述べる。

2. 提案手法

提案手法では、認識対象となる物体カテゴリ毎に、Visual Words からなるコードブックを構築する。従来の BoF 手法との違いを図 3 に示す。

図 3 は、左側が従来の BoF 手法の概略図、右側が提

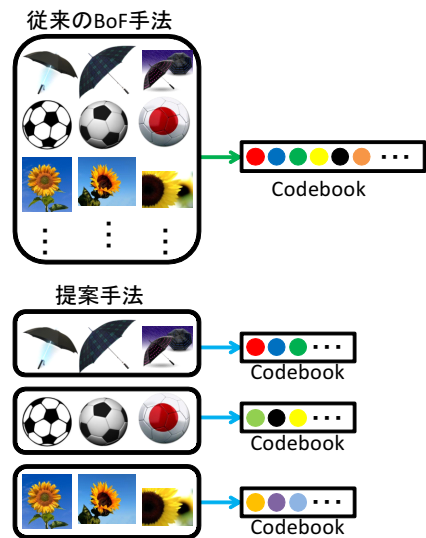


図 3 従来手法と提案手法の違い

案手法の概略図を表している。従来の BoF 手法では、認識対象となるすべてのカテゴリの学習画像を用いて、1 つのコードブックを作成する。こうすることで、作成されるヒストグラムには、カテゴリ毎に大きな違いができるため、認識には効果的である。しかし、この手法では新しいカテゴリの物体を認識しようとする際に、コードブックを再構築する必要がある。提案手法では、個々のカテゴリの学習画像を使って、カテゴリ毎にコードブックを構築するので、新たに認識対象が増えても、コードブックを再構築する必要がない。このため、計算時間が削減され、認識対象が増えると認識率が低下するといった問題も回避できる。

2.1 カテゴリ別 Visual Words

カテゴリ別に Visual Words を作成する利点は主に 2 つある。一つ目は、認識対象の数が増加しても認識率の低下が少ないことである。一般的に、認識対象の数が増加すると、認識率は低下する。実際に、Caltech-101 や Caltech-256 を用いた研究でも、後者の方が認識率は大幅に減少している。しかし、提案手法では、カテゴリ別に個々にコードブックを作成するため、認識対象となるカテゴリ数が増加しても、認識率の低下が少ない。二つ目は、一度あるカテゴリのコードブックを作成すると、それはもはや未知の物体ではないため、再び同じタスクが与えられても、学習する必要がない点である。つまり同じカテゴリに関しては、コードブックを作り直す必要がなくなる。

2.2 Reject 領域

2.1 節では、カテゴリ別にコードブックを作る利点を 2 つ述べた。しかし、それぞれの物体カテゴリ別に作られたコードブックは、従来の BoF 手法では考慮する必要のなかった図 4 に示すようなベクトル量子化の問題点を含



図 4 ベクトル量子化の問題

んでいる。

図 4 に示すように、一般的に、画像には物体の特徴とともに、背景やオクルージョンなどの、物体には関係の無い特徴も多く含まれている。画像認識のための Caltech-101 などのデータセットでも、そういったノイズが含まれているので、本研究で用いるウェブ画像には数多くのノイズが含まれることになる。Visual Words を作る段階では、学習のために収集した画像にノイズが存在していても、k-means クラスタリングによって、物体に関係する Visual Words と、ノイズに関係する Visual Words に分類される。しかし、BoF を作成する段階では、画像単位でヒストグラムを作成するので、入力画像の質によって得られるヒストグラムは大きく変わってしまう。つまり、画像にノイズが多ければ、認識したいカテゴリに関係のない特徴点が多く検出されてしまう。システムには、どの特徴点が物体に関係しているかそうでないかは判断できないので、そのまま BoF を求めると、物体の真のヒストグラムとは大きく異なったヒストグラムが得られてしまう。

この問題を解決するために、本研究では図 5 に示すような "Reject 領域" を提案する。

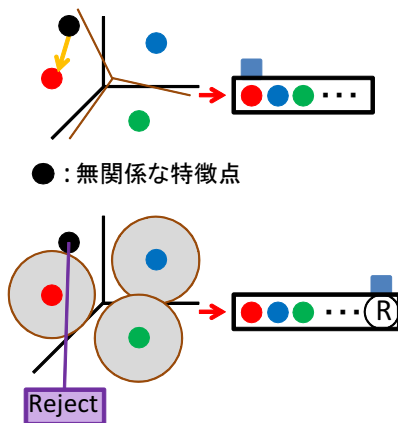


図 5 Reject 領域

図 5 の左に示すような従来の BoF 手法では、物体には関係の無い黒い点で表された特徴 (背景やオクルージョン) は、赤い点で表された Visual Words に分類される。Visual Words は、物体に関係のある特徴で作られているという前提なので、これらを積み重ねていくと、結果として得られるヒストグラムは真のヒストグラムとは大きく異なってしまう。しかし、それぞれの Visual Words に対して、半径によって領域を指定し、その円に入ら

ない領域を "Reject 領域" とすると、関係の無い特徴は、Reject できるので、その点はヒストグラムに数えられなくなる。このように、関係のない特徴を、"R" と書かれた Reject 領域を表すところに貯めていくようにすることで、図 6 で示すような真のヒストグラムにより近いものが得られる。図は、Reject 領域なしと Reject 領域ありの場合のコードブックに関して、傘の画像をベクトル量子化したときの図である。Reject 領域がない場合には、得られるヒストグラムは真の傘のヒストグラムとは異なってしまう。しかし、Reject 領域がある場合には、不必要な特徴が Reject されるので、真の傘のヒストグラムに近いものが得られる。

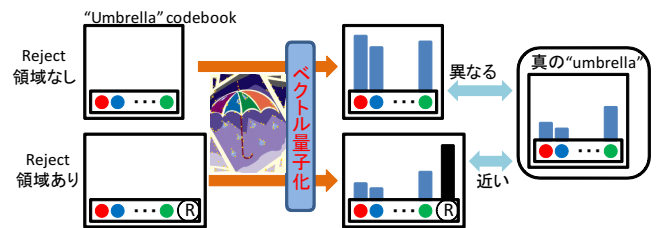


図 6 Reject 領域を考慮したヒストグラム

2.3 物体照合

物体を照合する際には、k Nearest Neighbor(kNN) 法や Support Vector Machine(SVM) などがよく用いられる。これらは、画像をベクトル量子化して得られるヒストグラムを学習データとする識別器である。しかし、コードブックをカテゴリ別に作成することで、図 7 に示すように、対象画像のヒストグラムを学習することなく、物体照合を実行できる。

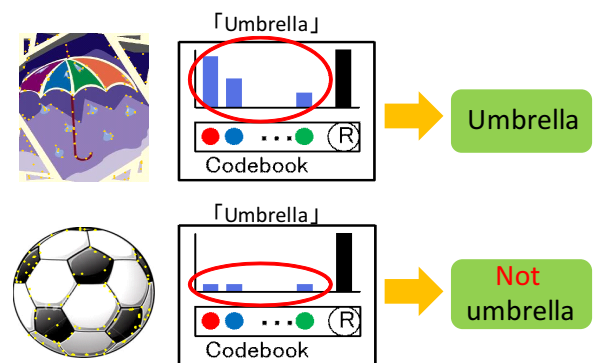


図 7 物体照合

図 7 には 2 つの例を示している。まず、一つ目は、"傘" のコードブックを用いて "傘" の画像をベクトル量子化した場合である。二つ目は、"傘" のコードブックを用いて "サッカーボール" の画像をベクトル量子化した場合である。一つ目の傘の画像は、傘を表す特徴を多く含んでいるので、ベクトル量子化を行うと、傘に関係のある特徴は Visual Words に入り、赤線で囲まれた領域のヒ

ストグラムは多くなるため、傘として認識されやすいヒストグラムとなる。一方、サッカーボールの画像は、傘を表す特徴はほぼ含んでいないので、多くの傘に関係のない特徴は Reject 領域に落とされる。このため、Visual Words に入る特徴の数は少なくなり、傘としては認識されないヒストグラムとなる。つまり、出現頻度ヒストグラムで、Reject 領域ではない Visual Words に注目し、そこに入った特徴のトータルの量を閾値処理することによって、照合することができる。もし、トータルの量が閾値より大きければ、その物体はそのコードブックが表すカテゴリとして照合される。対照的に、もしトータル量が閾値より小さければ、その物体はそのコードブックが示すカテゴリではないものとなる。

kNN [5] や SVM [6] などの従来の識別器は、学習画像のヒストグラムすべてを学習データとする必要がある。しかし、提案手法は学習画像のヒストグラムを用いる必要がなく、入力画像のヒストグラムだけで物体を照合できる。従って、計算時間を大幅に削減できる。

3. 実 験

3.1 データセット

提案手法の有効性を評価するために、Google 画像検索から上位 30 枚の画像を学習画像として収集し、Caltech-101 データセットから 20 枚をそれぞれのカテゴリのテスト画像として用いた。学習画像を 30 枚にした理由は、画像検索の精度がカテゴリによってばらつき、精度が低いカテゴリの検索結果では、ランキングが 30 位以下になる場合があり、このときの画像は、対象とはまったく異なっているためである。この点から、本研究では、30 枚を学習の枚数とした。また、テスト画像には、そのカテゴリの物体の全景が写っているものが望ましいため、Caltech-101 のデータセットを用いた。

3.2 実 験

本稿では、2 つの実験を行った。それぞれ物体認識実験と物体照合実験である。一つ目の物体認識実験では、10 クラス (ダルメシアン、ドル紙幣、ハリネズミ、ピザ、サッカーボール、標識、ひまわり、傘、イス、陰陽) の物体認識を行った。二つ目の物体照合実験では、同じ 10 クラスを用いて物体照合を行った。

3.3 結 果

図 8 に物体認識実験の結果を示す。

図 8 において、縦軸は認識率を表し、横軸はコードブックのサイズを表している。コードブックのサイズは、様々な研究で扱われているが、ベストなサイズは明確に定まっておらず、カテゴリ毎に異なるのが普通である。本研究では、カテゴリ別にコードブックを作成するので、従来の BoF に比べて必要なコードブックのサイズは小さくてよいと考えている。そこでコードブックのサイズ

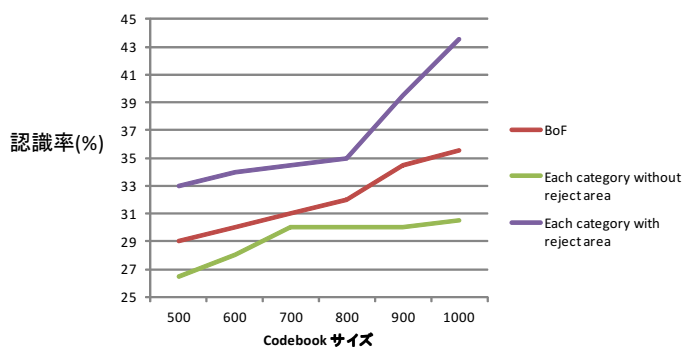


図 8 物体認識の結果

は 500 から 1000 の範囲で設定している。

図中の紫の線は、提案手法でカテゴリ別に Visual Words を作り Reject 領域を導入した結果である。緑の線は、カテゴリ別に Visual Words を作ってはいないが、Reject 領域を導入しなかった結果であり、赤い線は、従来の BoF を表している。この図から、Reject 領域を設けなかった場合のカテゴリ別手法の認識率は、コードブックのサイズを変化させても、常に従来手法の BoF よりも劣っていることがわかる。これは、図 6 で述べたように、Reject 領域を導入しなければ、真に近いヒストグラムが得られないためだと考えられる。また、紫の線のように、カテゴリ別手法で Reject 領域を正しく設定すれば、従来の BoF 手法よりも認識率が上回っている。これらより、Reject 領域は、物体認識において、重要な役割を果たしていると言える。

図 9 に物体照合実験の結果を示す。照合とは、画像 1 枚に対して「これは A ですか?」とタスクが与えられ、Yes/No で返答する問題と定義する。その結果は、Precision Recall 曲線 [4] で評価する。この実験では、図 7 で示した提案手法と kNN 法を比較する。実験で用いる画像セットは 1 つ目の物体認識実験と同じものを使用した。つまり、ある 1 つカテゴリのコードブックを作成し、それに対してテスト画像 (10 カテゴリ*20 枚) 全てに対して、「これは A ですか?」という質問を与え、その返答に対する Precision Recall の平均を求めた。

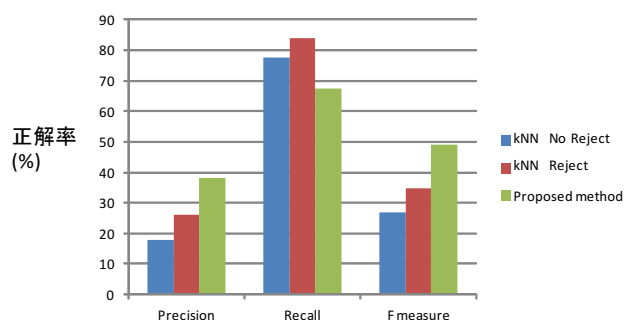


図 9 物体照合の結果

図 9 において、青色は、kNN 法で Reject 領域を導入

しなかった結果、赤色は、kNN 法で Reject 領域を導入した結果、緑色は、図 7 で提案した照合法による結果を表している。また、図 7 で提案した照合法では、Reject 領域含んでいる。

実験結果から、kNN 法において Reject 領域を用いなければ、F 値で結果が最も低くなる。Reject 領域ありの kNN 法と比較すると、提案手法は、Precision と F 値において、よい結果を示している。更に、計算時間も大幅に削減することができた。

3.4 考 察

物体認識と物体照合の 2 つの実験を行い、提案手法の有効性を示した。問題となるのは、認識と照合の精度が、他の研究成果と比べると全体的に低いことである。この原因は図 10 に示すように、ウェブから集めた画像には、異なる物体が多数含まれていることから、Visual Words を正しく構築することが難しく、ベクトル量子化によって作成されるヒストグラムも粗になっているためと考えられる。この問題を解決するために、画像検索そのものの精度を改善する研究 [7] や、検索した画像から重要な特徴を見つけ、画像をランキングする手法 [8] [9] [10] なども研究されている。これらの手法を用いることで、学習の精度が向上し、全体的に認識と照合の精度が改善されることが考えられる。

カテゴリ別に Visual Words を作成した当初の動機は、システムが未知の物体に遭遇する場面は多く、その度に、従来の BoF では、知識を再構築しなければならないという問題があったためである。

また、コードブックのサイズや Reject 領域のサイズは、対象となるカテゴリごとに異なると考えるのが普通であるので、今後は、それらの自動最適化などにも取り組んでいく予定である。



図 10 ウェブから収集された間違った画像

4. 結 論

この論文では、従来の Bag of Features 手法の問題点である、未知の物体への適応の問題を解決するために、カテゴリ別に Visual Words を作成することを提案した。具体的には、ベクトル量子化の問題を解決するために、Reject 領域を導入することによって、未知の物体の認識に用いることができる新しい手法を提案した。さらに、学習画像のヒストグラムを用いることなく、物体照合ができる手法も提案した。結果として、認識率は 8.0% 改善され、照合率は従来の手法よりも 14.1% 改善され、計算時間も大幅に削減することができた。

今後は、Reject 領域やコードブックのサイズを自動的に決定する手法について研究していく予定である。

文 献

- [1] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," Proc. ECCV Workshop on Statistical Learning in Computer Vision, pp.1-22, 2004.
- [2] D. G. Lowe, "Object recognition from local scale-invariant features," Proc. IEEE International Conference on Computer Vision, pp.1150-1157, 1999.
- [3] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," Journal of Computer Vision, Vol.60, No.2, pp.91-110, 2004.
- [4] J. Davis and M. Goadrich, "The relationship between precision-recall and roc curves." Technical report #1551, University of Wisconsin Madison, January 2006.
- [5] B. Dasarathy, "Nearest Neighbor Pattern Classification Techniques." IEEE Computer Society Press, Los Alamitos, CA, 1991.
- [6] J.A.K. Suykens and J. Vandewalle, "Least Squares Support Vector Machine Classifiers." Neural Processing Letters 9: 293-300, 1999.
- [7] H. Jegou, M. Douze, and C. Schmid, "Improving bag-of-features for large scale image search." IJCV, vol. 87, no. 3, pp. 316-336, 2010.
- [8] Yushi Jing, and Shumeet Baluja, "Visual Rank: Applying PageRank to Large-Scale Image Search." IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, vol.30, no.11, 2008.
- [9] Nobuyuki Morioka, and Jingdong Wang, "Robust Visual Reranking via Sparsity and Ranking Constraints."
- [10] Yushi Jing, and Shumeet Baluja, "PageRank for Product Image Search", WWW 2008 / Refereed Track: Rich Media, 2008.