Vol.2013-ICS-170 No.2
2013/3/10

物体類似度知識ベースを用いた物体認識

八木亮[†1]　吉村枝里子[†2]　土屋誠司[†2]　渡部広一[†2]

近年，画像情報はインターネット上のコンテンツなどを通してその利用が急激に増加している．その画像情報の検索技術についても web 上でのテキストと関連付けた手法だけではなく，類似画像検索などの画像情報を頼りにした検索手法も多くなり，画像認識の研究が活躍している．画像認識でも一般の物体をその対象のクラス名で認識することを一般物体認識と言い，数十年の間研究がなされている．一般物体認識の入力画像には実世界のシーンを用いるため，特徴量を算出する際に物体以外の余分な情報である背景などの特徴量を取得してしまうことがある．そこで，物体の類似性を考慮して構築した知識ベースを用いて，背景などの余分な情報を削除し，物体の特徴量だけを抽出する手法を提案した．

Object Recognition with a Knowledge Base for Object Similarity

RYO YAGI[†1]　ERIKO YOSHIMURA[†2]
SEIJI TSUTHIYA[†2]　HIROKAZU WATABE[†2]

Recently, image information is used by lots of people through contents in the internet. For searching image information, there are not only the ways that connect with texts in the web, but also the ways that use similarity image.
Recognizing the general objects as class names of target, called the general object recognition, has studied for a few decades. It can't be help to get extra information such as back ground when the system computes feature amount, because it is used the real scenes as input image of generic object recognition. So, the author proposed the way to delete extra information and extract feature amounts of objects using the knowledge base with information from similarity points.

1. はじめに

近年，画像情報はインターネット上のコンテンツなどを通してその利用が急激に増加している．その画像情報の検索技術についても web 上でのテキストと関連付けた手法だけではなく，類似画像検索などの画像情報を頼りにした検索手法も多くなり，画像認識の研究がこれまでより重要となっている．このように画像認識技術の需要が高まっている中で，画像認識でも物体をその一般的な名称(クラス名)で認識することを一般物体認識と言い，数十年の間研究[1]がなされている．

現在の一般物体認識で広く用いられている手法に，Bag of Features(BoF) [2]がある．この手法は画像中の物体のクラスを認識するものである．BoF は物体の画像から局所特徴量を抽出し，それらを k-means 手法によって k 個のクラスタに分類する，物体の見え方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(重心)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．この手法では，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 による物体表現は，物体のオクルージョン(物体が他の物体によって部分的に隠れてしまうこと)に強いといわれている．なぜなら，BoF は局所特徴量の集合であり，また k-means 法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，画像中にその物体のクラスに関係のない特徴や背景が多く含まれていると信頼性が失われるといった問題がある．

そこで，本稿では複数のクラスに共通して現れる特徴を，そのクラス特有の特徴ではないと考え，雑音として認識する．そこで，そのクラスに固有の特徴量のみを考慮して物体認識を行うために，物体類似度知識ベースを用いて物体認識を行う手法を提案する．物体類似度知識ベースには複数のクラスに共通して現れる特徴量を格納する．入力画像の特徴量に対してこの物体類似度知識ベースを用い補正をかけて物体認識を行う．

2. 研究概要

2.1 前提条件

本稿のシステムは次の 3 つの条件を満たすものを前提条件とする．学習画像のクラス分けは手動で行うこと，入力画像には分類対象のクラスの物体が 1 つのみ，入力画像のクラスは学習画像のクラスに含まれていることである．

2.2 システム概要

本システムは入力画像のクラスを認識し，出力する．システムの流れを図 1 に示す．物体類似度知識ベースとヒストグラム DB の構築の流れを図 2 に示す．

†1 同志社大学大学院工学研究科
Graduate School of Engineering, Doshisha University
†2 同志社大学理工学部
Faculty of Science and Engineering, Doshisha University

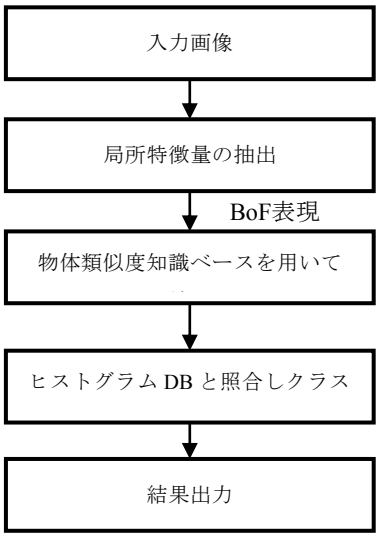

図 1 システムの流れ

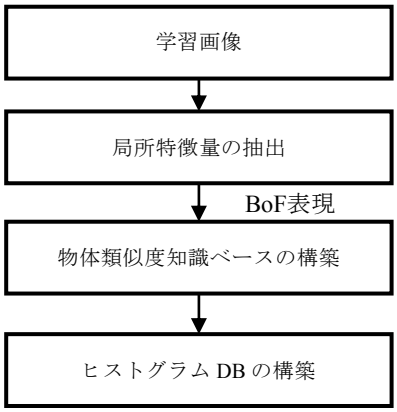

図 2 物体類似度 KB とヒストグラム DB の構築の流れ

2.3 前提条件

本稿のシステムは次の 3 つの条件を満たすものを前提条件とする．学習画像のクラス分けは手動で行う，入力画像には分類対象のクラスの物体が 1 つ，入力画像のクラスは学習画像のクラスに含まれている．

3. 物体類似度 KB とヒストグラム DB の構築

この章では図 2 に示した各ステップの詳細について述べる．

3.1 学習画像

本稿では知識ベース作成のための学習用画像に Caltech-101[3]を用いた．Caltech-101 はカルフォルニア工科大学の Fei-Fei によって収集され作られた画像データベースである． 101 の物体カテゴリと追加背景からなる 102 のクラスで構成され，31 から 800 の画像をカテゴリ毎に含んでおり，全体で 9145 枚の画像で構成されている．本稿ではこの背景を除いた 101 クラスの中からランダムに 25 クラス，各クラスからそれぞれ 30 枚の画像を学習画像とした．

3.2 局所特徴量の抽出

入力画像と学習画像から局所特徴量を抽出する．局所特徴とは画像中の濃淡変化が大きい特徴点を検出し，その特徴点の周りの領域を微分値により特徴ベクトルにしたものである．この特徴は対象が同一のものであれば，視点変化や回転変化，スケールの異なる画像であっても，同じ特徴点が検出されやすい．本稿で用いる局所特徴量として，SIFT[4][5]を用いた．SIFT は前述の通り，画像のスケール・回転変化に対してロバストに働くため，図3のように同一人物の対応点をSIFTにより求めることが可能である．そのため，特定物体の同定には非常に有効な特徴量と言える．しかし，図4のように異なる人物においては対応点を求めることができないため，一般物体認識問題などに関するクラス分類に対して，SIFT 特徴量をそのまま利用する事は困難である．そこでBag-of-Features のアプローチを用いる．

図 3 同一人物による対応点

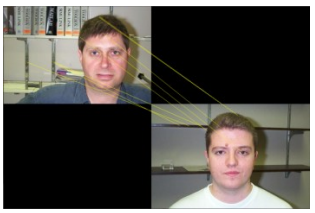

図 4 異なる人物での対応点

3.3 BoF(Bag-of-Features)表現

BoF は文書分類手法である Bag-of-words[6]を画像に適用した手法である．Bag-of-words では文章を単語の集合と見なし，単語の語順を無視してその頻度で文章の分類を行う．これと同様に，BoF では画像を局所特徴量の集合と見なし，その位置情報を無視して画像のクラス認識を行う．BoF の作成手順は以下の流れとなる．

1. 全ての学習画像から局所特徴量を抽出する．
2. クラスタリング手法を用い，局所特徴で構成される特徴ベクトルをベクトル量子化させ，局所特徴を word として扱えるようにする．
3. k 個の特徴ベクトルのうち最も近い特徴ベクトルに投票を行うことで，出現回数のヒストグラムで画像を表

現する．

4.

局所特徴量には SIFT を用いる．クラスタリングとは，学習パターンの空間内での分布状態を見て，クラスタに分割する処理のことである．クラスタリングのまでの流れを図 5 に示す．

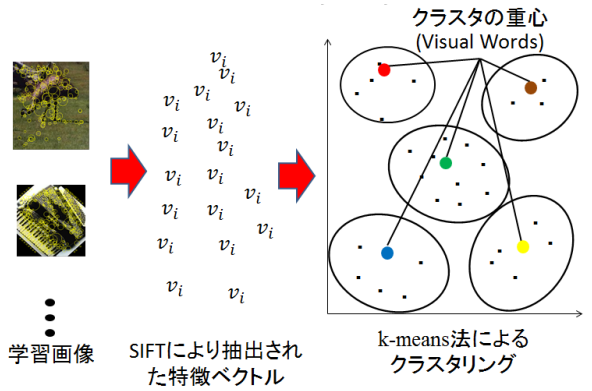

図 5 クラスタリングの流れ

本稿ではクラスタリングの手法として k-means 法を用いた．クラスタリングを行うことで得られた各クラスタの重心ベクトルが Visual Word になる．学習画像から SIFT 特徴量を抽出し，その特徴量を最も距離が近い Visual Word に分類する．Visual Word の出現回数をベクトル量子化し，ヒストグラムを得る．最後にその画像の局所特徴量の総数で各 bin を割ることで正規化されたヒストグラムが作成される．本稿ではクラスタ数を 200 個としてクラスタリングを行った．このように画像 1 枚を Visual Word の出現頻度ヒストグラムで表現することを BoF 表現と呼ぶ．全ての学習画像を BoF 表現で表す．BoF 表現の例を図 6 に示す．

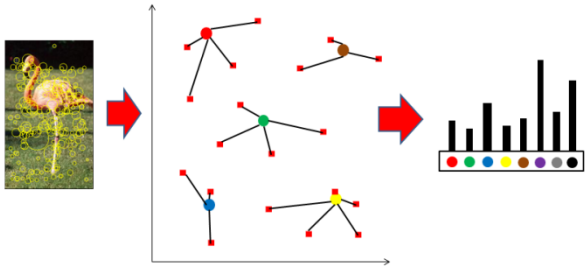

図 6 BoF 表現

3.4 物体類似度 KB の構築

一般的に，画像には物体の特徴のほかに，背景などの物体には関係のない特徴(ノイズ)も多く含まれており，本稿で使用した学習画像の中でもノイズが大きく影響してしまうクラス が存在する．BoF の特徴として，背景が似ているクラスは背景の特徴の出現頻度から物体自身のクラスと関係なく同じクラスとして認識されることがある．

この誤認識に対応するため，本稿では物体類似度知識ベースを作成した．物体類似度知識ベースには，学習画像クラス名と複数のクラス間に共通して現れる特徴の Visual Word ID を格納する．ここで Visual Word ID とはクラスタ分類した際の重心ベクトルのことである．クラス分類の際にその ID の Visual Word 出現頻度を無視することで，複数の学習画像のクラスに共通した特徴を無視して分類できる．複数の学習画像クラスと共通して現れる特徴はそのクラス固有の特徴ではないため，雑音として処理することができると考えられる．

学習画像からそれぞれのクラスの学習画像の Visual Word 出現頻度の平均ヒストグラムを算出する．そして，そのヒストグラム同士でそれぞれの特徴量の出現頻度の上位 n 位までに同じ Visual Word があった場合，その Visual Word の ID を物体類似度知識ベースに登録する．物体類似度知識ベースに Visual Word ID を登録する例を図 7 に示す．図 7 では例として学習画像の特徴量を 8 個にクラスタリングした際の，Visual Word ID を色付きの円で示している．emu クラスと flamingo クラスの学習画像の Visual Word 出現頻度の平均ヒストグラムをそれぞれ算出する．そして，それぞれのクラスの Visual Word ID の出現頻度の上位 n 位を比較する．この例では n=2 としている．emu クラスは赤と紫，flamingo クラスでは紫と黒の Visual Word ID を比較する．紫が重複しているため，物体類似度知識ベースの emu クラス，flamingo クラスの項目に紫の Visual Word ID を登録する．

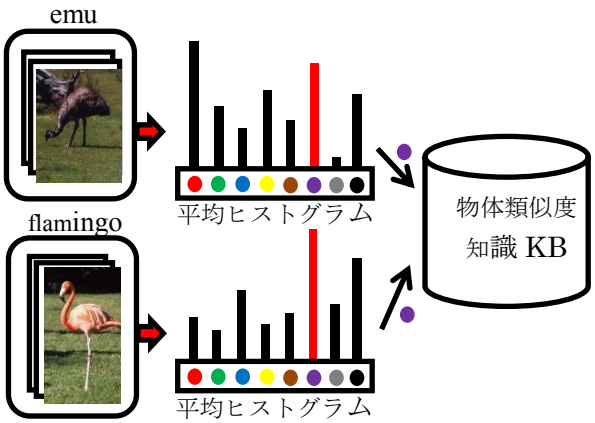

図 7 物体類似度 KB 構築の流れ

この処理を全てのクラスに対して行う．以下に物体類似度知識ベース作成の流れを示す．

(a) 学習用画像から BoF を用いてクラスそれぞれの平均ヒストグラムを算出する．
(b) BoF を使って求めた出現頻度のヒストグラムを $ck(i)1$（$k \in 1, \ldots K$）まで比較し，最も出現頻度の数が近い Visual Word から順位づけを行う．
(c) 順位の高い上位 n 位までの Visual Word ID を持っ

た知識ベースを構築.

物体類似度知識ベースのイメージを表 1 に示す.

表 1 物体類似度 KB 例

accordion	●
emu	●
flamingo	●
・・・	・・・

3.5 ヒストグラム DB の構築

3.4 節で得られた物体類似度知識ベースを用いて，学習画像ヒストグラムに補正をかける．物体類似度知識ベースに格納されている Visual Word ID はそれぞれのクラスにとって雑音であるので，各クラスの学習画像ヒストグラムに対して，物体類似度知識ベースに登録されている Visual Word ID の出現頻度を 0 にする．そして，Visual Word 出現頻度の合計が 1 になるように正規化する．この処理を全てのクラスの学習画像ヒストグラムに対して行い，ヒストグラムを作成しなおす．得られた学習画像のヒストグラムを保持したヒストグラム DB を構築する．ヒストグラム DB の構築例を図 8 に示す.

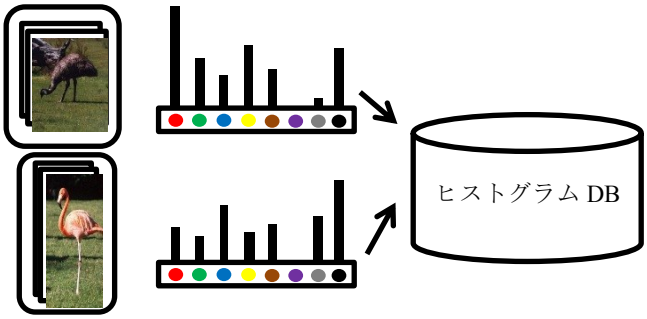

図 8 ヒストグラム DB 構築例

4. 物体類似度 KB を用いた物体認識システム

この章では図 1 に示した各ステップの詳細について述べる.

入力画像は 2 章で示した前提条件を満たすものを用いる．局所特徴量の抽出は 3.2 節の処理を入力画像に対して行う．得られた入力画像の特徴量を BoF 表現で表す．物体類似度知識ベースを用いた補正では，3.4 節で学習画像ヒストグラムに対して行った処理と同様の処理を入力画像のヒストグラムに対して行い補正をかける．このとき，入力画像のクラスは不明であるので，クラス毎の補正をかけるために入力画像を学習画像のクラス分だけ複製する．複製された入力画像ヒストグラムに対して，クラス毎の補正をかける．この処理のイメージを図 9 に示す.

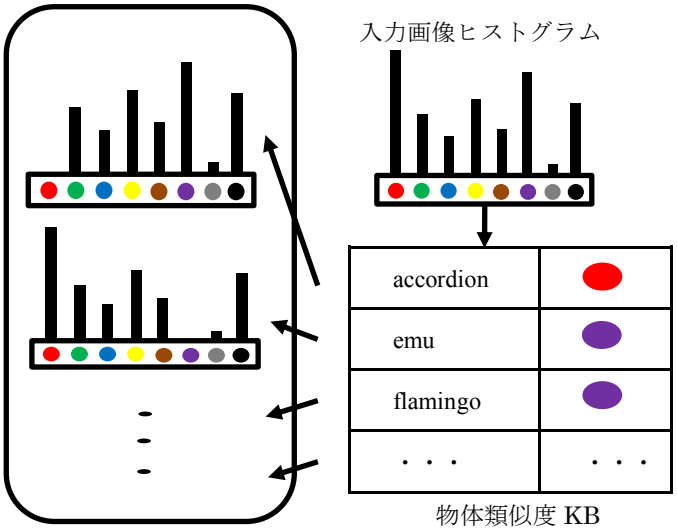

図 9 物体類似度 KB を用いた補正の例

次に，ヒストグラム DB と照合してクラス認識では，3.4 節で得られたヒストグラム DB 内の全てのヒストグラムと物体類似度知識ベースを用いて補正をかけた入力画像ヒストグラム間の類似度を計算する．ヒストグラムの類似度の計算方式として Histogram Intersection[7]を用いる．以下に類似度計算の式を示す.

$$\sum_i \min(H_1[i], H_2[i])$$

H_1 と H_2 はそれぞれ入力画像ヒストグラムと学習画像ヒストグラムを表す．それぞれのヒストグラムの i 番目のうち小さい方の値を返す．つまり 2 つのヒストグラムの対応する各 bin の値の小さい方を足し合わせていくことで，2 つのヒストグラムの類似度を得る．クラス認識の流れは以下のようになる.

1. 入力画像を BoF 表現で表し，ヒストグラムを得る.
2. ヒストグラムを学習クラス数と同じ 25 個に複製し，物体類似度知識ベースを参照し，25 クラスに応じて補正する.
3. 入力画像ヒストグラム群に対し，ヒストグラム DB の全画像ヒストグラムと類似度計算を行う.

ヒストグラムの類似度を計算し，上位 m 位までに存在するクラス毎に類似度を足し合わせる．全体で最も類似度の最も高くなったクラスを出力する.

5. 評価

入力画像として学習画像のクラスと同じクラスの画像を各 30 枚ずつ用意し, BoF 表現のみで物体分類を行ったものと，物体類似度知識ベースを用いて補正をかけ物体分類を行ったものを実験した．入力画像のクラスの分類精度を評価した．分類精度は以下の式によって表される．図 10 に分類精度を示す.

$$\text{分類精度}(\%) = \frac{\text{正しくクラス認識された入力画像数}}{\text{入力画像総数}} * 100$$

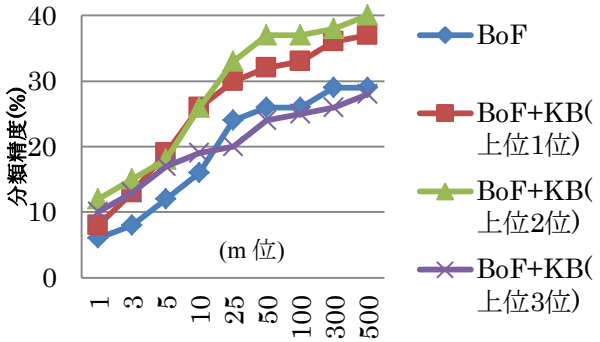

図 10 **分類精度**

6. 考察

BoF 手法のみで分類した場合よりも物体類似度知識ベースを用いて補正をかけたときのほうが最大で 10.3%精度が良くなり，分類精度として 39.6%の精度を得た．物体類似度知識ベースに格納する Visual Word ID は上位 2 位よりも範囲を大きくすると精度が下がっていく傾向が見られた．これは，物体類似度知識ベースに含める Visual Word ID を多くすればするほど，ノイズだけでなく，物体自身の特徴も無視してしまうために起こると考えられる．

7. おわりに

本稿では，物体類似度知識ベースを用いて物体を分類する手法を提案した．入力画像から特徴量を抽出し，その特徴量と BoF 表現で構成されるデータベースを作成し，物体類似度知識ベースを用いて補正をかけることによって，分類精度を 10.3%向上させることができた．

今後の課題として，背景を考慮するだけでなく，同一クラス内でも特徴が大きく違うクラスはそれぞれもっと詳細なクラス設定(例えば「モンシロチョウ」や「アゲハ蝶」など)を行わなければ, 正しく分類することは難しいと考えられる．

謝辞 本稿の一部は，科学研究費補助（若手研究(B)24700215) の補助を受けて行った．

参考文献

1) 柳井啓司,一般物体認識の現状と今後，情報処理学会論文誌：コンピュータビジョン・イメージメディア, 48,SIG16 (CVIM19), pp. 1-24 (2007).
2) G. Csurka, C. R. Dance, L. Fan, J. Willamowski,and C. Bray, Visual categorization with bags of key-points, Proc. ECCV Workshop on Statistical Learn-ing in Computer Vision, pp.1-22, 2004.
3) Caltech 101 image dataset. http://www.vision.caltech.edu/Image Datasets/Caltech101/
4) D. G. Lowe, Object recognition from local scale-invariant features, Proc. IEEE International Con-ference on Computer Vision, pp.1150-1157, 1999.
5) D. G. Lowe, Distinctive image features from scale-invariant keypoints, Journal of Computer Vision,Vol.60, No.2, pp.91-110, 2004.
6) C.D. Manning, H. SchFutze. Foundation of Statistical Natural Language Processing,The MIT Press, 1999.
7) Swain,M. J. and Ballard, D. H.: Color Indexing,International Journal of Computer Vision, 7(1), pp.11-32, 1991.