Vol.2014-ICS-174 No.5
2014/3/4

# SURF 特徴量を用いた BoF 法による 複数物体認識

古川拓也[†1]　吉村枝里子[†2]　土屋誠司[†2]　渡部広一[†2]

近年，デジタルカメラや写真共有サービス等の普及によって，膨大な画像が存在するようになり，従来の画像周辺のテキストとの関連を用いた検索手法だけでは，目的の画像を見つけ出すことが難しくなってきた．そこで近年は，類似画像検索など，画像情報を頼りにした検索手法も多く提案されるようになり，画像中に何が写っているのか認識する物体認識が重要になってきている．本稿では，BoF 法を用いた物体認識を複数物体が写った画像にも適応できるようにし，実用的な物体認識の実現を目指した．

# Multiple Object Recognition By BoF using SURF

TAKUYA FURUKAWA[†1]　HIROKAZU WATABE[†2]
SEIJI TSUCHIYA[†2]　ERIKO YOSHIMURA[†2]

Recently, there are numerous images because of spreading service of sharing pictures or digital cameras. It is difficult to find　it is becoming too difficult to find the objective image from web site using character information nearby the image. So recently it is suggested to find the image using similarity image. So I propose the way to recognition for images which it is taken multiple objects by BoF using SURF.

## 1. はじめに

画像処理の分野で重要な課題の 1 つである物体認識は，近年，さらに重要になりつつある．物体認識とは，画像が与えられたとき，写っている物体が何かをコンピュータに理解させることである．デジタルカメラや，写真共有サービスの普及などに伴い，誰もが手軽に画像を扱えるようになった．それに伴い，ウェブ上においても実世界シーンの画像が爆発的に増加してきている．誰もが手軽に画像をウェブ上にあげられるため，タグ付けも行われていないような画像が増え，画像につけられたファイル名や画像と同じサイト上にある情報による検索も難しくなってきている．今後は，画像中に写っている物体そのものを認識し，検索する手法が求められる．また，今後身近になっていく人間の手伝いをするパートナーロボットや車の自動運転システムなどにも物体認識は有用である．従来のようにセンサで何らかの物体が存在することを認識するだけでなく，物体が何かを画像から得ることで，パートナーロボットであれば，指示されたものを取ることができ，自動運転システムであれば，避けなければならないものかどうかを判断することが出来る．このように，今後さまざまな方面で重要となる物体認識に，本研究では近年の物体認識で主に用いられ，背景の影響や個体差の影響による誤差も許容できる Bag-of-Feature[1] (以降 BoF，詳細は後述)を用いて取り組む．

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

## 2. 研究目的

複数物体が写った入力画像に対し物体認識を行い，物体名を出力することを目的とする．具体例を図 1 に示す．

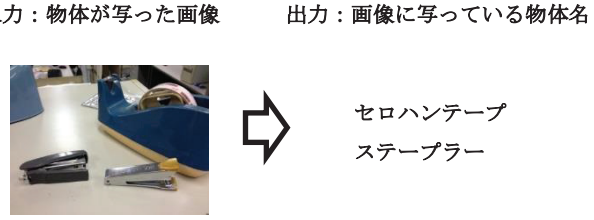



図 1　本研究の入出力の具体例

入力画像は複数物体が写った自然に撮影された画像とする．複数物体とは 0 個以上の物体を指し，画像中に物体が写っていない場合でも，1 つだけの場合でもよい．ただし，画像中に物体が写っていても，小さすぎるものは対象外とし，図 1 のステープラー程度の大きさで物体が写っているものとする．また，自然に撮影された画像とは，人が日常においてカメラで撮影した画像を指し，画像処理のために前提条件に従った背景が限定されたり，照明の角度が固定されたりしたようなものでなくてもよいということである．

研究の概要について述べる．本研究では BoF 法を用いて物体認識を行う．本研究の目的は，実用的な物体認識の実現である．従来の物体認識の対象は，1 つの物体が写っている画像であることが多かった．これは単独物体に対する物体認識自体が難しい課題であり，複数物体認識を行う段階になかったためである．しかし，今日では BoF 法の開発により，1 物体の認識に対して 6～7 割程度の精度で物体認識が行えるようになってきた[1]．BoF 開発以前も，物体認

識の研究は盛んにおこなわれていた．物体の形状特徴や色特徴に注目する手法が様々提案されたが，いずれも 2 割程度の精度しかだせていない[1]．それに対し，BoF 法の 6~7 割の精度は非常に良いといえる．今後は 1 物体に対しての物体認識だけでなく，実用的な場面での物体認識が求められる．そこで本研究では，実用的な場面として，複数物体が写った画像を対象とし，複数物体が写った画像からそれぞれの物体が何であるのかを認識する．

本研究では，後述する SURF 特徴点により画像領域を分割し，独自作成した BoF ヒストグラム DB を基に物体を認識する．詳細は後述するが，領域分割を行うことで複数物体が写った画像であっても，単独物体と同様に物体認識を行うことができ，SURF 特徴点により認識する分割領域部を定めることで，処理時間の増大を防ぐことができる．

## 3. 関連研究

### 3.1 局所特徴

局所特徴とは物体を構成する小さなパーツのことである．局所特徴による認識とは物体全体を見てそれが何かを認識するのではなく，物体は局所特徴，つまり細かなパーツの組み合わせによって構成されているため，局所特徴を抽出し比較すれば認識を行えるという考えに基づく手法[1]である．図 2 に局所特徴の例を示す．

図 2　局所特徴の例

人間は図 2 の左の画像を見た時，特に深く考えなくてもそれがバイクであると理解できる．しかし，コンピュータでは人のように簡単に画像を理解することはできず，長らく物体認識の研究は画像処理最大の課題の 1 つとされてきた．それのブレークスルーとなったのが局所特徴の考え方である．図 2 の右のように画像を分解する．図 2 では，画像中の物体を構成するパーツはタイヤ 2 個やハンドル，ヘッドライトなどである．そのようなパーツの組み合わせで出来る物体はバイクであると認識する．“タイヤ”や“ハンドル”のような物体を構成するパーツを局所特徴量(以降，特徴量)，それらが画像中にある位置を示した座標を局所特徴点(以降，特徴点)という．図 2 の例は概要を説明するために“タイヤ”などで分割したが，実際の局所特徴はこのように大まかなものではなく，Scale Invariant Feature Transform(以降，SIFT)[1][2][4][5]や近似計算を用いることで SIFT を高速化した scale Invariant Feature Transform(以降，SURF) [1][2][4][5]が用いられ，本研究では SURF を用いる．SURF は画像の回転，スケール変化，照明変化，オクルージョン(重なりなどにより物体の一部が隠れてしまうこと)に強い特徴を持つ．

### 3.2 BoF 法

BoF 法は Bag-of-Features 法の略で Bag-of-Keypoints 法や bag-of-Visual Word 法とも呼ばれる．これらの名前は Bag-of-Words 法と呼ばれる文書分類手法に基づいたもので，アルゴリズムもこの手法を基にしている．Bag-of-Words 法は文章を単語の集合とみなし，単語の語順を無視してその頻度で文章の分類を行う手法である．これと同様に BoF 法は“画像は局所特徴の集合であり，その位置関係を無視してもその頻度のみで何の画像か認識できる”というものである．具体例を示す．SURF で取得した特徴量の位置情報を放棄し，図 3 で示すように特徴量をいくつかの代表値に量子化する．

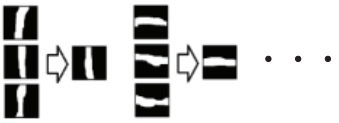


図 3　特徴点量子化のイメージ

入力画像から得られた特徴量を全て量子化した後，その頻度をヒストグラムとして作成する．図 4 に例を示す．

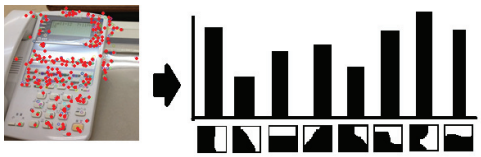


図 4　画像のヒストグラム化

図 4 の例では，理解しやすいようにヒストグラムの要素になる局所特徴を簡潔なパーツで示しているが，実際は SURF などの局所特徴を用いて分解する．ヒストグラムの要素数も例のような 8 要素ではなく数百から数千となる．

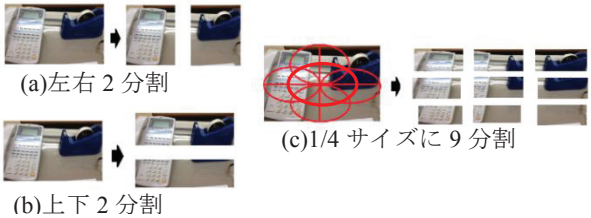
このように，入力画像に対し，それぞれの BoF ヒストグラムを作成することで，ヒストグラムの比較によって物体認識を行うことが出来るようになる．

## 4. 提案手法

複数物体が写った画像を認識するために，入力画像中の物体が写っている部分の領域検出方法と，ヒストグラムの比較方法の 2 つを提案する．

### 4.1 領域分割

画像中に複数の物体が写っていても認識を行えるようにするため，画像を分割し，分割後の画像を認識に利用する．画像の分割の仕方を図 5 に示す．



(a)左右 2 分割
(b)上下 2 分割
(c)1/4 サイズに 9 分割

図 5　分割方法

図 5 のように分割した上下 2 分割，左右 2 分割，1/4 サイズに 9 分割, 分割なしの計 14 枚を物体認識の対象画像とする．図 5(c)は 1/4 サイズへの分割であるが単純に 4 分割すると領域分割の切れ目に入ってしまい，物体が切れてしまう場合があるので図 5(c)中に示したように 9 領域に分割した．図 5(c)中で楕円で示している部分は直線にすると線の区別がつかないため楕円で示したが，実際は他領域と同様に長方形に分割している．

しかし，14 枚すべてに認識を行ってしまうと，処理時間の増大や何もない部分に物体があると認識されるなどの問題が生じる．そこで図 6 のように物体の周辺に特徴点が集中することを利用する．

図 6　特徴点が検出された位置

画像を図 5 のように分割し，その部分から得られた特徴点の数の割合が全体の特徴点の数に対して閾値以上であれば，そこに物体がある可能性が高いとして，その領域からヒストグラムを作成し，後述する物体認識のための比較を行う．物体が存在する可能性が高いとしてその領域を検出する閾値は，2 分割の場合は分割部の特徴点数が画像全体の特徴点数の 4 割以上，9 分割の場合は全体の 2 割以上とした．

### 4.2　ヒストグラムの比較による物体認識

入力画像から 4.1 節の処理で検出した領域(以降，候補領域)に対し，BoF ヒストグラムを作成しデータベースと比較を行う．

#### 4.2.1 データベースの構築

本研究では 30 種類の物体(クラス)に対し各々30 枚ずつ画像を用意し，その画像から得られた全てのヒストグラムを格納したデータベースを構築した．

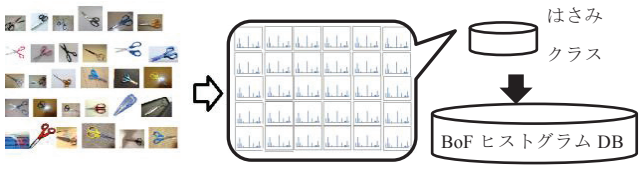



図 7　はさみクラス構築の例

図 7 ははさみクラス構築の例である．他の物体にも同様の処理を行い，30 クラス 30 枚ずつ計 900 個のヒストグラムが格納されたデータベースとした．データベース作成の際に用いたクラス(物体の種類)の一覧を表 1 に示す．表 1 の物体は，自然物や人工物など多くのカテゴリに分かれるように選択した．

表 1　データベースの作成に用いた物体

| | | | | |
|---|---|---|---|---|
| はさみ | ホッチキス | 地球儀 | 眼鏡 | 扇風機 |
| ピアノ | メトロノーム | ギター | 時計 | 電話 |
| マウス | パソコン | カメラ | 車 | 紙幣 |
| 野球ボール | バット | 盆栽 | 包丁 | 壺 |
| パンダ | ライオン | カモメ | キリン | シマウマ |
| 鳩 | 鷲 | イチゴ | リンゴ | バナナ |

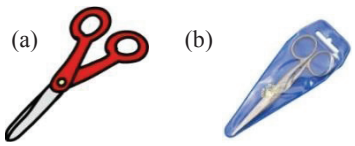



図 8　DB 作成画像のうち特徴的な画像

図 8 ははさみクラスのデータベースを構築するために用いた画像のうち，特徴的なものである．図 8(a)は，はさみのイラストである. このように画像がイラストであっても，ある程度正確に描かれていれば，実際の写真と同じような結果を得ることが出来る．(b)ははさみが外袋に入っている画像である．この場合，袋部分によって多少違った特徴量も検出されるが, 多くははさみの部分の特徴量であるため，オクル―ジョンに強い SURF では誤差の部分は問題にならない．しかし，(b)以上に他の物体などが写り込むと，はさみ以外の特徴が大きくなるため，今回はそのような画像は人手で判断し，学習画像から取り除いた．

#### 4.2.2 ヒストグラムの比較

候補領域のヒストグラムの比較にはヒストグラムインターセクション[3]を用いる．ヒストグラムインターセクションとはヒストグラムの各要素を比較し，小さい方を加算していくことによりヒストグラムの比較を行う処理である．これにより得られた値(以降，一致度)は 0.0 から 1.0 の範囲をとり，一致度が高い方がヒストグラムを得た画像同士は似ているということになる．

#### 4.2.3 結果出力

候補領域に対し，データベース全ての画像と一致度の計算を行った後，上位 30 位のものを出力する．上位 30 位以内に入ったということはその画像は入力画像と似ているということになるが, 何らかの誤差で違う画像が上位 30 位以内に入ってくることもある．そこで，上位 30 位中の出力回数に注目し，表 2 のように点数付けを行った．

表 2　出力回数による点数付け(それぞれ 1 枚に付き)

| | | | |
|---|---|---|---|
| 5 位まで | 7 点 | 6 位から 10 位 | 5 点 |
| 10 位から 20 位 | 3 点 | 20 位から 30 位 | 1 点 |

例えば，入力画像に対し，シマウマクラスの画像が 5 位までに 2 回，10 位までに 1 回，20 位までに 2 回，30 位までに 6 回出力されたときシマウマクラスの点数は(式 1)で求まる．

2 回×7 点＋1 回×5 点＋2 回×3 点＋6 回×1 点　(式 1)

(式 1)を計算すると 31 点になり，これがシマウマクラスの

点数である．この点数が高い順に物体の候補とし出力した．点数が 20 点以下になった場合はその領域には物体はないとした．

#### 4.2.4 同一物体の除去

同一物体名が複数領域から出力された際の処理について述べる．図 9 は画像全体からの出力も，左上の領域からの出力もシマウマとなるような例である．


図 9　同じ物体が複数回出力されてしまう例

図 9 の画像を処理し，出力だけを見た場合，シマウマが 2 頭写っているような結果にとれてしまう．このようなことを避けるために，物体が検出された領域に内包される領域で，同一物体名が 1 位に出力された場合は，その領域は外側の領域と同じ物体を指しているとして結果出力しない．

次に内包はしないが，領域の一部を共有するときについて述べる．図 10 に例を示す．

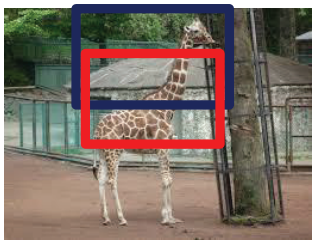


図 10　領域の一部を共有する例

図 10 の赤で示した領域も青で示した領域もキリンを 1 位に出力する．この場合も内包するとき同様，キリンが 2 頭写っているかのように出力されてしまう．このときは内包関係のときと異なり，領域の大きさ自体に差はなく，それに注目して片方の領域を採用することができない．そこで領域ごとの特徴点数に注目する．同一サイズの領域が一部を共有して 1 位に同じ結果を出した場合は，領域内の特徴点数が多い方を採用した．また，共有も内包もしない領域で 1 位に同じものが出力された場合は，同じ物体が 2 つ画像内にあるとして，どちらも結果として採用した．

## 5. 実験

1 物体が写った画像 50 枚，2 物体が写った画像 30 枚，3 物体が写った画像 10 枚，物体が写っていない画像 10 枚の計 100 枚に対して実験を行った．また，物体が写っていない画像とは BoF ヒストグラム DB に格納されているか否かに関わらず，物体が画像中に写っていない画像であり，風景などの画像を指す．

### 5.1 1 物体が写った画像

1 物体が写った画像に対して，提案手法の 1 つである複数物体認識のための領域分割は行わず，画像全体に対してのみ物体認識を行った．これにより今回実装した BoF 法を用いた物体認識が，どの程度の精度で認識できるのか確認した．

#### 5.1.1 結果

図 11 に示すような 1 物体が写った画像に対して表 3 の基準で評価を行った．評価結果を図 12 に示す．また，実験に使用した 50 枚の画像は BoF ヒストグラム DB に格納された各々のクラスの物体ごとに 1 枚か 2 枚用意し，使用した．

表 3　1 物体画像の評価基準

| |
|---|
| ○：正解の物体名が 1 位に出力された |
| △：正解の物体名が 3 位以内に出力された |
| ×：正解の物体名が出力されなかった |
| ※物体名が正解か否かは人手で判断 |

図 11　1 物体が写った画像の例

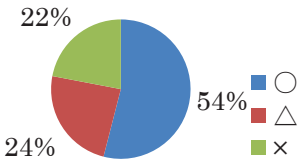



図 12　1 物体が写った画像

○が 27 枚，△が 12 枚，×が 11 枚であった．

#### 5.1.2 考察

BoF 法を用いた物体認識の既存研究において，物体が 1 つ写った画像に対する物体認識の精度は 2006 年時点で最高は 66.23%[1]である．そのため，本研究の結果の○だけに注目すると，高い精度が得られているとはいえない．しかし，○と△を合わせると 78%と既存研究よりも高い精度を得られた．つまり，約 8 割の確率で 3 位までに正解物体名が出力されたと言える．

成功例について述べる．特に得点が高かった例を図 13 に示す．


図 13　認識の際点数が高く出力された例

図 13 のように正しく認識できたものは物体特有の特徴がはっきりしていることがわかる．例えば，シマウマであればシマ模様，パソコンであればキーボードなどの他に似たものがない部分である．

失敗例について述べる．主な失敗として(1)背景が似ている画像(2)背景が複雑な画像(3)似ているクラスがあり区別が難しい画像の 3 つの場合を確認した．

(1) 背景が似ている画像(4 枚)

背景が似ていて物体名が 1 位に出力されなかった物体の例を図 14 に示す．

1 位　キリン　　25 点
2 位　ライオン　19 点
3 位　パンダ　　14 点

図 14　背景が似ている失敗例

図 14 では背景が草や木などである他の動物が出力されてしまった．象の特徴量よりも背景の特徴が多く取れ，それにより背景が重視された結果になってしまったと考えられる．このような失敗例は主に背景が草木になる動物系に多く見られた．このような画像は今後，背景を除去して物体部分のみを取り出す処理[6]や，象の背景としてよく出力される特徴量をあらかじめ学習しておき，それを取り除く処理[3]をすることで対処する必要があると考える．文献[6]は背景差分法という手法で画像からあらかじめ用意した背景を取り除いて物体のみを抽出している．本研究では，あらかじめ背景の画像を用意することは，ウェブ上から無作為に画像を収集するため難しい．そこで文献[3]のような背景の特徴量を除去したヒストグラムを作ることが重要になる．文献[3]では入力画像同士を比較することで近似する特徴量を除去しているが，背景差分法の考え方と合わせて，象と共によく出力される背景, 例えば草原などの画像から，あらかじめ BoF ヒストグラムを求めておき，それと入力画像の共通する部分を除去することで，背景の影響を抑えた物体認識が行えると考える．

(2)　背景が複雑な画像(4 枚)

背景が複雑で失敗した画像の例として図 15 がある．図 15 は絨毯の上に置いたはさみの画像である．この画像では絨毯から非常に多くの特徴点(図 15 の赤点)を抽出してしまい, はさみの特徴量の割合が非常に小さくなってしまった．

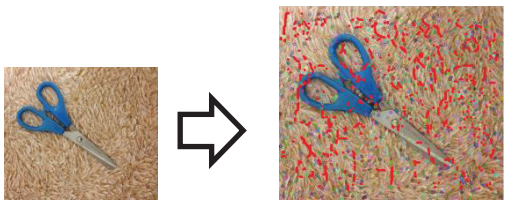


図 15　背景が複雑なものの例

背景により物体認識が失敗したという点では失敗例(1)と同様であるが，図 15 の場合，背景は象と草木のように，はさみと絨毯は直接関係のないものであり，あらかじめはさみと関連がありそうな画像を用意しておき除去することは難しい．このように背景から必要以上に特徴点を検出してしまう例として他に新聞の上に置いた物体の画像が確認された. 図 15 では絨毯の毛先 1 つ 1 つが特徴点として検出され，新聞の例では新聞に印刷された文字 1 つ 1 つが特徴点として検出されてしまった．このような背景の場合，SURF で処理を行ってもそれらの点が特徴点として検出されるため，SURF で処理を行う前に色情報などをもとに画像を領域分割し，不必要な部分を除去する必要があると考えられる．

(3) 似ているクラスがあり区別が難しい画像(3 枚)

似ているクラスがあり区別が難しい画像を図 16 に示す．

1 位　ハト　　23 点
2 位　ワシ　　18 点
3 位　キリン　12 点

図 16　似ているクラスがあり区別が難しいものの例

図 16 はワシの画像であるが, 1 位がハトとなってしまった．これはハトも飛んでいるときは同じような形になり，特徴が似かよるためであると考えられる．

このように，本手法ではハトとワシのように鳥という同一カテゴリ内の物体の詳細認識は困難となる．これは BoF が一般性を持たせるための手法であることため避けられない．一般性を持たせるとは，例えば，人が写った画像が入力されたときに“山田さん”と出力するのではなく，“人”と出力するということである．人が写った画像を入力したとき“山田さん”と出力するような手法は特定物体認識と呼ばれ，その物体は“山田さん”であるかどうかのみを認識する研究である．このような研究には SURF を BoF ヒストグラム化していない手法などが用いられる．BoF ヒストグラムは SURF の位置関係を放棄することにより一般性を持つが，それと引き換えに特定性を失っている．具体例として，SURF では，“目が少し離れていて，顎にヒゲが生えている”のが“山田さん”，それ以外は山田さんではないことはもちろん，人がどうかもわからない，という手法であるのに対し, BoF 法は目と鼻と口が顔のどこかにあれば人，でもそれが山田さんかどうかまではわからない，という手法である．失敗例(3)を改善するためには，本手法ではじめから“ハト”，“ワシ”などを認識するのではなく，“鳥”というクラスにしておき，その後 SURF などの特定物体認識手法でハトなのかワシなのかを認識することが必要であると考える．

さらに，このように，“鳥”クラスを作るということは BoF ヒストグラム DB の構築方法の検討が必要だと考える．例えば，本稿で作成したデータベースでは，“ワシ”“ハト”“カモメ”などを“鳥”クラスに統合すれば精度向上が期待されるが，逆に“ライオン”“パンダ”“シマウマ”などは高い精度を得られているため“動物”クラスに統合する必要はなく，むしろ動物は個々の特徴の差が大きいものが多いので統合すると精度低下すらも招きかねない．また，

本稿では“車”クラスとしているが，車クラスには，セダン，バン，SUV や広く見ればトラックまでも分類される．これも“動物”クラス同様，“車”クラスにまとめている方が精度が下がると考えられる．本研究では，BoF ヒストグラム DB の構築を研究の主題としておかなかったため改善の余地が残ったが，BoF ヒストグラム DB 構築方法を精錬することでより精度向上が見込めると考える．

### 5.2 2 物体以上が写った画像

#### 5.2.1 結果

図 17 に示すような 2 物体以上が写った画像における評価は，はじめに，画像内の個々の物体が正しく認識できたかを 1 物体のときと同様に表 3 の基準で調べた．2 物体が写った画像 30 枚の 60 物体，3 物体が写った画像 10 枚の 30 物体，合計 90 物体の結果を図 18 に示す．

図 17　2 物体が写った画像

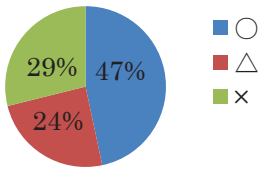



図 18　複数物体が写った画像(それぞれの物体)

○が 42 物体，△が 22 物体，×が 26 物体であった．

次に画像全体に対して評価を行った．評価基準を表 4 に示す．2 物体が写った画像 30 枚，3 物体が写った画像 10 枚の結果を図 19 に示す．

表 4　複数物体が写った画像の評価基準

| |
|---|
| ○：写っている物体名が全て 1 位に出力された |
| △：写っている物体のうち全てが 3 位以内には出力された |
| △⁻：写っている物体のうち 1 つは 3 位以内に出力された |
| ×：写っている物体全てが出力されなかった |

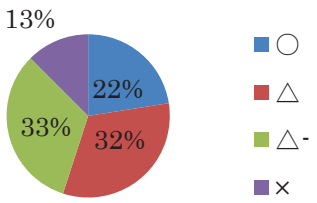



図 19　複数物体が写った画像(画像全体)

○が 9 枚，△が 13 枚，△⁻が 13 枚，×が 5 枚であった．

さらに，複数物体が写った画像の場合，何もない領域が誤認識により何かあると認識されてしまうことがある．図 20 はその時の例である．誤認識でパンダが出力されている．

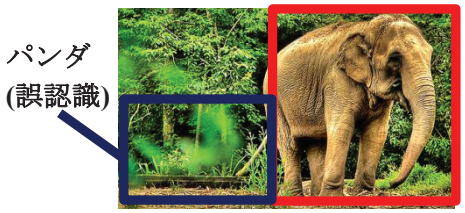



図 20　誤認識の例

複数物体が写った画像 40 枚についてこのような誤認識領域があったかどうかも調べ表 5 の基準で評価を行った．結果を図 21 に示す．

表 5　誤認識領域の評価基準

| |
|---|
| ○：誤認識領域なし |
| ×：誤認識領域あり |

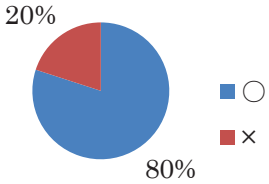



図 21　誤認識領域の有無

○が 32 枚，×が 8 枚であった．

図 19，図 21 の結果をまとめると図 22 のようになる．図 22 の項目の説明と件数を表 6 に示す．

表 6　複数物体が写った画像の結果のまとめ
(評価基準と件数(2 物体画像/3 物体画像))

| |
|---|
| A：物体全てが 1 位　誤認識領域なし　・・・・・7 枚/2 枚 |
| B：物体全てが 1 位　誤認識領域あり・・・・・・0 枚/0 枚 |
| C：物体全てが 3 位以内　誤認識領域なし・・・　11 枚/1 枚 |
| D：物体全てが 3 位以内　誤認識領域あり・・・・1 枚/0 枚 |
| E：物体いずれかが 3 位以内　誤認識領域なし・・7 枚/1 枚 |
| F：物体いずれかが 3 位以内　誤認識領域あり・・2 枚/3 枚 |
| G：物体いずれも 3 位外　誤認識領域なし・・・・2 枚/1 枚 |
| H：物体いずれも 3 位外　誤認識領域あり・・・・0 枚/2 枚 |

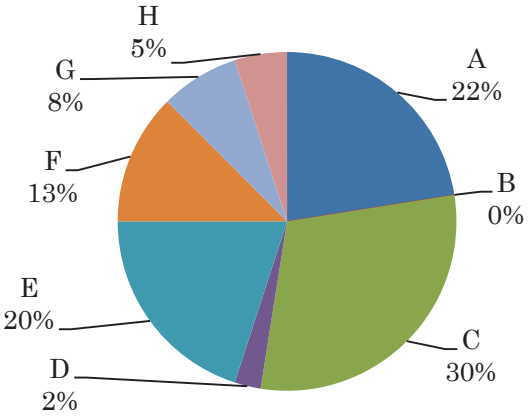



図 22　複数物体が写った画像のまとめ

5.2.2 考察

物体ごとの評価基準は 1 物体のときと同じである．多少の精度低下は見られたが，1 物体の時と比べ大きな精度低下は見られなかった．

1 物体のときと比べ，精度が下がっている原因となった失敗例について述べる．失敗例として(1)領域分割が失敗した画像(2)複数物体が写っていることでオクルージョンが大きくなりすぎて認識できなかった画像，があった．

(1) 領域分割が失敗した画像

領域分割の失敗について述べる．図 23 は扇風機とバットが写った画像である．

図 23 領域分割が上手くいかない画像

図 23 は画像全体の領域に認識を行うとバットと扇風機の特徴量が混合するため一致する物体がないという結果になる．本研究で提案した手法で分割すると図 23 の右のように扇風機の部分は特徴点が多く現れる領域に分割できるが，バットは全体が写る領域に分割出来ず，持つ部分と打つ部分が分断されてしまっている．提案研究ではこのような写り方をしている写真には対応できないとわかった．提案手法では，SURF 特徴点を用いて固定的に 13 領域に分割を行ったが，今後は動的に物体の輪郭に沿わせた領域分割が必要になると考える．

また，図 23 に限らず，図 24 のように物体が乱雑に散らかっている画像は特徴点が色々な部分に分散するので，本手法では対応が困難であると判明した．

図 24 乱雑に物体が散らかっている画像

本研究では，特徴点を用いておおよその物体の位置を判断することには成功しているため，本手法と合わせて，物体の形状に合わせた領域分割の手法を提案することで，より精度の高い認識が行えると考える．

(2) オクルージョンが大きすぎる画像

図 25 はマウスの影に，はさみが隠れて写った画像である．図 25 の右は，はさみが最も大きく現れる分割部である．図 25 では，はさみが最も大きく現れる領域であっても，はさみの持ち手のほとんどがマウスに隠されていることがわかる．


図 25 物体の大部分が隠されてしまっている例

BoF 法はある程度のオクルージョンまでは許容するが，一定以上に物体が隠れてしまった場合は対応することが困難となる．どこまで許容するのか，ということは物体によって異なるため定めることは難しいが，オクルージョンにより失敗する例もあると確認した．

このような画像への対処法として，隠れている物体部分を復元する手法や，もし撮影したカメラがロボットにインストールされたカメラなどで撮影位置の変更が可能な場合は撮影位置を変え撮り直しをする手法などが考えられる．

次に画像全体に対して述べる．画像全体に対して，写っている物体全てが 1 位に出力された画像は 40 枚中 9 枚であり精度が良いとは言えない．しかし，全ての物体の正解が 3 位までに入っているものが 13 枚，3 位までの候補中に正解物体のうち少なくとも 1 つは入っているものが 13 枚と，画像中の物体のヒントとなるものが少なくともなにか 1 つは得られたものは合計で 35 枚となり，全体の 9 割近くを占めている．今後の物体認識は画像処理手法のみでなく，画像が撮影された場所や条件，一緒に写りえるものなど，画像の周辺情報を利用することも重要になってくると言われている[1]．例えば，図 26 の入力画像に対し，図中の結果が得られたとする．


図 26 画像例

このとき，撮影場所が動物園だという情報があればシマウマの誤認識のピアノや車を取り除くことが出来るし，キリンと同時に写りえるものとして関連性を考慮出来た場合も，ピアノや車などを取り除くことが出来る．本研究では物体候補を一致度により順位づけしているため，今後重要になってくると考えられる画像の周辺情報を用いた認識に応用しやすいものになっている．

また，画像の周辺情報を用いる方法として，BoF ヒストグラム DB を状況ごとに作成しておくことも有効である．例えば，一般的には日本の街中にライオンやパンダはいない．しかし現在の全ての物体を 1 つにまとめたデータベースだと日本の街中で撮影された画像であってもライオンやパンダとの一致度も計算することになる．これにより計算

コストは増加し，誤認識率も上がると考えられる．今後，物体認識に画像処理ではなく画像の周辺情報を用いて取り組む際は，例えば，街中用，室内用，サバンナ用など画像の撮影地点ごとに対応できるような BoF ヒストグラム DB を作成し，適宜利用することも効果的である．

最後に領域分割時になにも写っていない領域にも物体があると認識してしまった失敗例について述べる．入力画像 40 枚中，何もない部分に物体があると認識したのは 8 枚と比較的少なかった．これは，特徴点が少ない部分は処理を行わないことと，処理を行った領域でも得点付けによる結果が 20 点をこえなかった場合は物体なしとしたことの 2 段階で物体なしの領域を除外する処理をしたことが効果的だったと考える．何もない部分に物体ありと出力した失敗例として，1 つの物体を認識するときの失敗例などと同様，背景に草木が多い動物の画像が，草だけが写っている領域に割り当てられたものがあった．

### 5.3 物体が写っていない画像

#### 5.3.1 結果

図 27 に示すような物体が写っていない画像 10 枚に対して表 7 の基準で評価を行った．評価結果を図 28 に示す．

図 27　物体が写っていない画像

表 7　物体が写っていない画像の評価基準

| |
|---|
| ○：物体なしと出力された |
| ×：何か物体があると出力された |

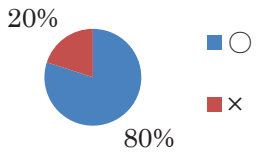



図 28　物体が写っていない画像

○が 8 枚，×が 2 枚であった．

#### 5.3.2 考察

何も映っていない画像に対しては物体なしと 8 割の精度で認識出来た．誤認識した画像は，草原をパンダとしたものと，空をワシとしたものであった．これは 5.1.2 項や 5.2.2 項で述べたことと同様に背景部分と物体部分を分けて処理すると改善されると考えた．もしくは，空，草原，などあらかじめ良くあるシーンで BoF ヒストグラムを構築しデータベースに格納しておいてもよいと考えた．

## 6. おわりに

BoF 法を用いた複数物体が写った画像の物体認識に取り組んだ．本研究では，物体認識自体に対し大きな精度向上が見られたわけではないが，複数物体を認識することができ，より実用的な物体認識の実現に近づけたと考える．また，今後は BoF ヒストグラム DB の構築を工夫し，5.1.2 項(3)のようなデータベース由来の誤認識率を下げることや，画像処理だけで物体認識を完結せず，それにより得られた情報と画像の撮影場所や関連性などを用いて，画像処理による特徴量比較などでは限界となる部分を補うことで精度向上が図れると考える．実際に物体認識の手法を画像処理的に改良し，大きく精度を上げることは難しいとされており[1]今後は物体認識は画像処理の範囲にとどまらず，撮影された場所や同時に写った物体との関連性を考慮したり，ウェブ上の情報を用いて推定したりすることが重要となる．

**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，pp.1-24 (2007)
2) 藤吉弘亘，山下隆義，”物体認識のための画像局所特徴量”，CVIM チュートリアルシリーズ：コンピュータビジョン最先端ガイド 2，pp.1-60(2010)
3)八木亮，“物体類似度知識ベースを用いた物体認識”，情報処理学会研究報告，2013-ICS-170，pp.1-5(2013)
4) 柳井啓司,“物体認識技術の進歩”,日本ロボット学会誌，Vol.28，No.3，pp.257-260(2010)
5) 藤吉弘亘，“Gradient ベースの特徴抽出-SIFT と HOG-”，情報処理学会研究報告，CVIM160，pp.211-224(2007)
6)波部斉，和田俊和，松山隆司，“照明変化に対して頑健な背景差分法”，情報処理学会研究報告，CVIM115-3，pp.14-23 (1999)