F-010

一般物体認識における画像知識ベースの雑音除去手法の提案
Proposal of a Method to Remove Noise in Image Knowledge Base for Generic Object Recognition

高岡 賢人† 芋野 美紗子‡ 土屋 誠司‡ 渡部 広一‡
Kento Takaoka Misako Imono Seiji Tsuchiya Hirokazu Watabe

1. はじめに

近年，人間の代わりとなるパートナーロボットは家事の手伝いや介護をこなして，人間の手助けをすることが期待されている．そのため，パートナーロボットには，自ら移動し，指示された物を取得するといった動作が必要となる．そこで，そういった動作を実現するには，ロボットが外界から取得した画像からその内容を理解する技術である物体認識が必要となる．

ここで，一般物体認識について説明する．一般物体認識とは，制限のない状態で撮影された画像から，“ライオン”や“リンゴ”のように，物体を一般的な名称で認識することである．また，人間の顔など特定の物体に注目した物体認識として特定物体認識といったものがあるが，本稿では一般物体認識を行う．一般物体認識では，画像に存在する雑音の影響を受けやすく，認識率が低くなるという問題がある．本稿で定義する雑音とは，画像中における認識したい物体以外のもの，例えば背景に写る草などである．そこで，本稿では画像中の雑音の影響を抑える手法を提案し，物体の認識率の向上を研究目的とする．

2. 研究概要

本稿では，一般物体認識における認識率低下の原因となる雑音を除去する手法を提案する．ここで，一般物体認識の認識手法を説明する．まず，入力画像から局所特徴量を取得し，それらを後述する Bag-of-Features[1]を用いてヒストグラム化する．そして，そのヒストグラムと画像知識ベースとを比較し，入力画像の物体名を出力する．画像知識ベースとは，用意した画像群に対して Bag-of-Features を行い，ヒストグラムとその物体名を格納した知識ベースである．そこで，その画像知識ベースに対して雑音除去を行うことで，一般物体認識の認識率の向上を図る．

3. 関連技術

3.1 SURF[2]

画像から局所特徴量を取得する手法として，Speeded Up Robust Features（SURF）[2]がある．この手法は，特徴点の検出，及び特徴量の記述を行う Scale Invariant Feature Transform（SIFT）[3]の処理時間を高速化したアルゴリズムとなる．SURF は画像の回転，スケール変化，照明変化，オクルージョン（物体が重なりなどによって一部が隠れていること）に強いという特徴をもつ．

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

3.2 Bag-of-Features[1]

Bag-of-Features は，画像を局所特徴量の集合と見做し，局所特徴量の出現頻度をヒストグラムとして表現する手法である．画像から抽出したすべての局所特徴量を用いてクラスタリングを行う．クラスタリングにより生成された各重心に属する局所特徴量（128 次元の特徴ベクトル）の数を正規化し，ヒストグラムとして表現する．ここで，生成された重心を Visual Word と呼ぶ．各 Visual Word は 0 から 1 の実数値で表現される．

3.3 Histgram Intersection

Histgram Intersection とは，比較する 2 つのヒストグラムを H_1，H_2 として，それらの i 番目の要素を $H_1[i]$，$H_2[i]$ と表すとき，（式 1）で求めることができる．

$$\sum_i \min(H_1[i], H_2[i]) \qquad \text{（式 1）}$$

（式 1）では 2 つのヒストグラムの各々の要素を比較して小さな方を加算していき，2 つのヒストグラムの類似度を算出する．本稿では，1 つの入力画像に対し，画像知識ベースのすべてのヒストグラムと比較し，それぞれの類似度を算出する．そして，その中から最も類似度の高かったヒストグラムに付随する物体名を出力する．

4. 提案手法

本稿では，対象物体以外のものはすべて雑音と定義する．例えば，図 1 に示す画像には対象物体として犬，雑音として草などがある．3.1 節で述べた SURF は画像中の対象物体に限らず，画像全体から特徴点を抽出してしまう．つまり，犬だけでなく，その周りの草などからも特徴点を抽出してしまう．その結果として，入力画像と画像知識ベースを比較する際に，誤認識を起こし物体の認識率を下げている．そこで，本稿では画像知識ベースに対して，雑音除去を行い，その知識ベースを用いた一般物体認識を提案する．

図 1　対象物体と雑音

4.1 入力画像と画像知識ベースの画像群

カリフォルニア工科大学が提供する画像のデータセットである Caltech 256[4]を入力画像と画像知識ベースの構築に使用する．このデータセットには，256 種類の物体の画像群と雑音の画像群があり，各物体の画像群はそれぞれ 80 枚以上の画像を保有する．その中から，50 種類の物体の画像

群中，各 60 枚を使用し，それぞれ各 10 枚を入力画像，残り 50 枚を画像知識ベースの構築に使用する．

4.2 画像知識ベースの構築

画像知識ベース構築には 50 物体各 50 枚，合計 2500 枚を用いる．まず，それらの 2500 枚の画像群に対して，3.3 節で述べた SURF 特徴量を抽出する．そして，得られたすべての SURF 特徴量を Bag-of-Features を用いて各画像の特徴量の出現頻度を作成する．なお，Visual Word の数は 2000 に設定した．画像知識ベースの例の一部を表 1 に示す．表 1 では，各物体 50 枚における，各 Visual Word(VW)1～2000 までの出現頻度を表す．

表 1　画像知識ベースの一部

	VW1	VW2	VW3	・・	VW2000
ライフル(1)	0	0	0.00282	・・	0
ライフル(2)	0.00699	0	0.00699	・・	0.00699
ライフル(3)	0	0.00794	0.00794	・・	0
・・・	・・・	・・・	・・・	・・	・・・
ライフル(50)	0.00224	0.00224	0.00224	・・	0
・・・	・・・	・・・	・・・	・・	・・・

4.3 画像知識ベースの雑音除去

4.2 節で構築した画像知識ベースに対して，提案する手法により雑音を除去する．Visual Word の出現頻度が 0 の値を示すとき，その Visual Word は物体の特徴を表現していないと考えられる．しかし，Bag-of-Features を用いた手法では，同じ物体の画像から特徴を抽出しても，物体の写る角度や隠れなどにより，必ず特定の Visual Word に出現頻度が存在するとはいえない．例えば表 1 のライフル（1），（2）において，VW3 に出現頻度が共通して存在するが，VW1 には共通して存在していない．このことから，各物体の複数枚の画像から取得した各 Visual Word の中で，出現頻度が共通に存在する枚数が少なければ，その Visual Word は雑音と見做せると考えた．そこで，そのような雑音を除去する手法として，まず VW1～VW2000 までの各 Visual Word の出現頻度が存在する枚数を数える．そして，それらの枚数の中央値を算出して，その中央値以下になった個数を示す VW の出現頻度を 0 にする．以下に示す表 2，3 を用いて具体的に説明する．

表 2　各 Visual Word の出現頻度の個数

	VW1	VW2	VW3	・・	VW2000
ライフル(1)	0	0	0.00282	・・	0
ライフル(2)	0.00699	0	0.00699	・・	0.00699
ライフル(3)	0	0.00794	0.00794	・・	0
・・・	・・・	・・・	・・・	・・	・・・
ライフル(50)	0.00224	0.00224	0.00224	・・	0
枚数	2	3	12	・・	10

表 3　中央値以下になった VW の処理結果

	VW1	VW2	VW3	・・	VW2000
ライフル(1)	0	0	0.00282	・・	0
ライフル(2)	0	0	0.00699	・・	0.00699
ライフル(3)	0	0	0.00794	・・	0
・・・	・・・	・・・	・・・	・・	・・・
ライフル(50)	0	0	0.00224	・・	0
枚数	0	0	12	・・	10

表 2 のライフル（1）～（50）における VW1～VW2000 の出現頻度が共通に存在する枚数を元に中央値を算出する．このとき，中央値が 6 だとすると，表 3 の枚数が 6 以下を示した VW の出現頻度を 0 にする．この処理を各物体で行っていき，雑音を取り除いた画像知識ベースを構築する．

5. 評価

入力画像を Caltech 256 から選んだ物体から 10 枚ずつ合計 500 枚を Caltech 256 からランダムに選択する．提案手法による画像知識ベースと提案手法を用いない画像知識ベースによる一般物体認識の精度の比較を行った．その評価結果を図 2 に示す．

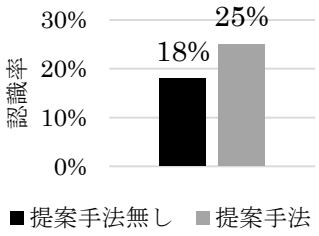

図 2　評価結果

6. 考察

図 2 の評価結果から提案手法による一般物体認識が提案手法を用いない場合に比べ 7%の精度向上が見られた．雑音と見做された Visual Word の出現頻度が 0 になり，Histgram Intersection で類似度を算出する際，雑音と見做された Visual Word が比較対象として扱われなくなった．その結果，誤認識を抑えることができたと考えられる．また，今回構築した画像知識ベースは，各物体 50 枚で構成されていたが，各物体の画像数をさらに増やすことで，物体の特徴を強く示す Visual Word と雑音を示す Visual Word がより鮮明になり，さらなる精度の向上が期待できると考えられる．

7. おわりに

Bag-of-Features を用いた一般物体認識において，入力画像の物体と比較する画像知識ベースの雑音除去について提案した．その結果，雑音と見做した Visual Word の出現頻度を除去し，認識率が向上した．今後は，画像知識ベースにおける雑音除去のみでなく，入力画像に対しても雑音除去を行う必要があると考えられる．

謝辞

本研究の一部は，科学研究費補助金（若手研究（B）247700215）の補助を受けて行った．

参考文献

[1] G. Csurka, C. Bray, C. Dance, L. Fan, “Visual Categorization with Bag of Keypoints”, Proc. ECCV Workshop onStatistical Learning in Computer Vision, pp59-74 (2004).
[2] D.Lowe , “Distinctive Image Features from Scale-Invariant Keypoints”, International Conference on Computer Vision, Vol.60, No.2, pp.91-110(2004)
[3] Lindeberg, Tony. "Scale invariant feature transform", *Scholarpedia* 7.5 (2012)
[4] Caltech 256 image dataset, http://www.vision.caltech.edu/Image_Datasets/Caltech256/