1R-04

ビッグデータからの能動的学習によるテキストと画像の双方向認識

棚橋　弘毅,　　水野　俊一郎,　　長谷川　修

東京工業大学

1. はじめに

近年，インターネット上のビッグデータを物体認識に活用する手法が多く提案されている．

その一例として，Fergus らによる Web 検索で得た画像を用いた手法[1]や，Samadi らによる実際にロボットを用いて Web 検索とユーザからの入力で学習させる手法[2]，Santosh らによる N-gram を用いて物体とそれに関連する単語を網羅的に学習させる手法[3]などが挙げられる．

しかし，[1][2]は Web 上のテキストと画像の双方を用いてはおらず，[3]はクエリ語に対する関連ワードの対象範囲が限られている．

そこで本研究では，Web 上のテキストおよび画像の双方を用いて，クエリ語に対する関連ワードとその画像を広く学習する手法を提案する(図1)．例えば，ロボットに調理させる事例では，食材を正しく認識させる必要がある．この時，あらゆる食材の名前とその画像をユーザが教示するのは大きな負担である．提案手法はこうした課題に対して有効と考えられる．

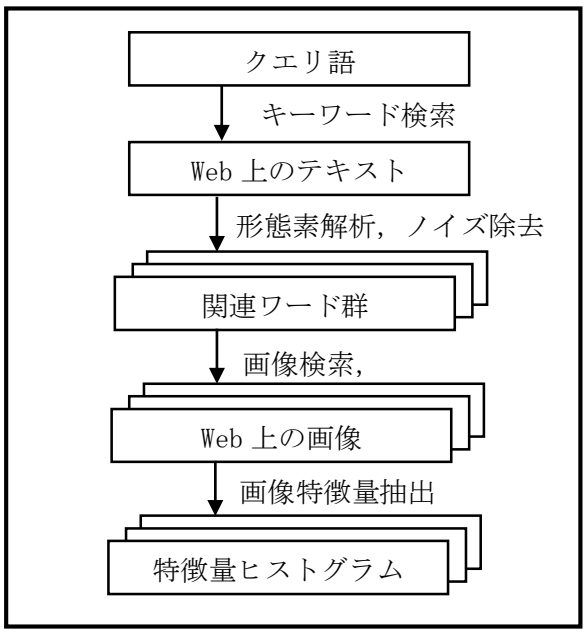

図 1： 提案手法

Automatic learning images by text analyses using the big-data

Kouki Tanahashi,　Tokyo Institute of Technology
Syunichiro Mizuno,　Tokyo Institute of Technology
Osamu Hasegawa,　Tokyo Institute of Technology

2. 提案手法

2.1. 概要

ユーザがクエリ語を入力すると，それを用いてキーワード検索を行い，Web 上から得られたテキストを形態素解析して，クエリ語に対する関連ワード群を抽出する．この時，相互参照によるノイズ除去（2.2.）も行う．

次に，得られた関連ワード群それぞれに対して Web 画像検索を行い，候補の上位画像をダウンロードし，それを代表する画像であるとみなす．

最後に，各単語に対応する画像特徴量を得るために，その単語の各画像から局所特徴量(SIFT や SURF など)を抽出し，Bag-of-Visual-Words を用いてそれぞれヒストグラム化する．その後，各単語の画像群の分割によるノイズ除去（2.3.）を行い，各画像のヒストグラムを合わせて，関連ワード群それぞれに対応する特徴量ヒストグラムを作成する．

2.2. 相互参照によるノイズ除去

一般に関連ワード群には,与えたクエリ語に関係なく Web 上での出現頻度が高い単語も含まれる．

そこで，まずクエリ語の関連ワード群 A を抽出する．その後，A を用いて更にそれぞれに対する関連ワード B を抽出し，B の中にクエリ語が含まれているものを関連ワード群として用いる．

2.3. 各単語の画像群の分割によるノイズ除去

通常，Web 画像検索で得られた画像には，検索象物体が写っていない画像も含まれる．また，物体が写っている画像であっても，多様な写り方が存在する．

これに対処するため，各単語に対して Visual-Words を用いて物体の写り方に対する画像群をそれぞれ作成する．この時のアルゴリズムを表 1 に記す．そして，用いる画像群の個数を変更することによりノイズの除去を行う．

表 1 ：画像群の分割手順

1	ランダムに画像を 1 枚選択する
2	選択した画像との距離が近い画像を複数（全画像数/群の数）-1 枚選択する
3	選択した画像をもとの画像群から取り除きすべての画像を分割出来たら終了.
4	手順 1 に戻る,

3. 評価実験

まず，ドメイン名をクエリ語に，その物体画像を入力する．次に，既に作成してある各単語に対応する特徴量ヒストグラムとの評価値を順にソートし，上位の単語を対象画像の物体名称と推定する(図 2)．

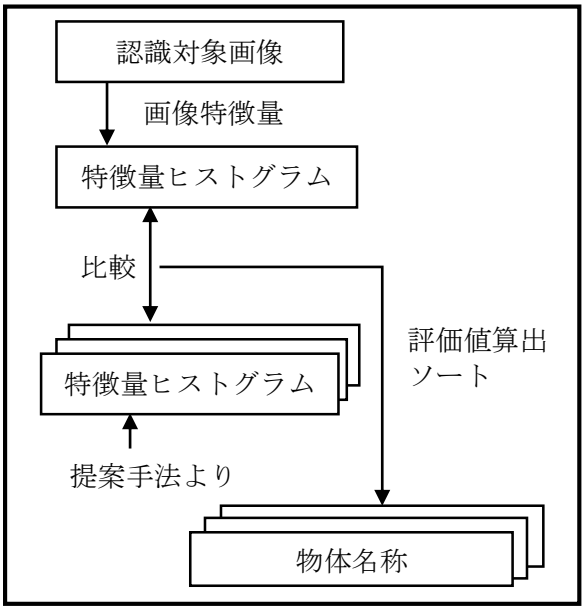

図 2 ：実験方法

3.1. 実験条件

表 2： 実験条件

クエリ語	“野菜”
検索エンジン	Google
検索結果ページ数	上位 20 ページ
形態素解析器	Mecab
関連ワード群数	上位 1000 個
相互参照	検索：上位 1 ページ 範囲：上位 100 件
学習画像数	上位 100 枚
画像特徴量	SURF(64 次元)
Visual-Words 数	2048
分割群数	5 群

実験条件を表 2 に示す．

認識対象画像として，入力したクエリ語のオブジェクト 10 種類に対して計 500 枚の画像(1 種類に対して 50 枚)を使用する．

今回は，評価値順にソートして結果上位 5 件に対象物体名が含まれれば正しく認識されたとする．また，特徴量ヒストグラムを構成するのに用いる分割群の数を 3, 4, 5 群で比較する．例えば，3 群を用いる場合は，学習画像が 100×3/5=60 枚を用いて，残りの 40 枚はノイズする．

3.2. 実験結果

実験結果を図 3 に示す．

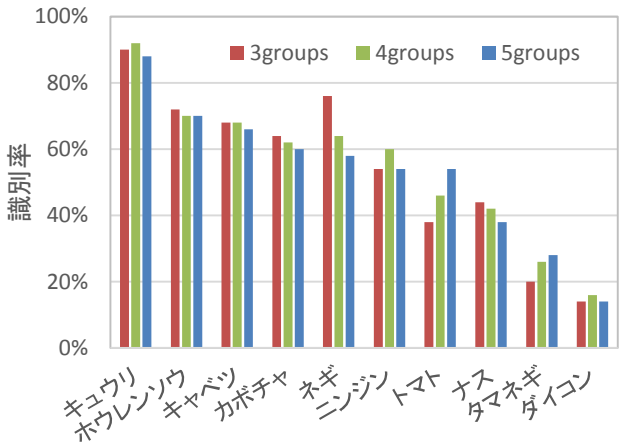

図 3 ：実験結果

4. 考察

キュウリ(図 4)に関しては，形が細長く光沢が少ないので，局所特徴量が万遍なく抽出された結果，良好な認識結果が得られたと推測される．これに対し，ダイコンでは，背景と同色の部分から局所特徴量として抽出されず識別率が低下したと推測される．

図 4： 左がキュウリ，右がダイコンの一例

また，物体によって用いる分割群の最適数は異なる．これは，その物体の写り方が多様である場合は群の数が多いほうが，識別率が高くなると考えられる．一方で，写り方が限られている物体の場合は群の数が少ないほうが結果は良くなっていると推測できる．

謝　辞

本研究は，JST CREST の支援を得て行われました．

文　献

[1] R. Fergus, et al.; “Leaning object categories from internet image searches”, Proceedings of the IEEE, 2010.

[2] M. Samadi, et al.; “Using the Web to Interactively Learn to Find Objects”, AAAI, 2012.

[3] S. Divvala, et al.; “Learning Everything about Anything: webly-supervised visual concept learning”, CVPR, 2014.