

Web上の大量画像を用いた特定物体認識手法による一般物体認識

電気通信大学大学院 情報理工学研究科 総合情報学専攻 秋山瑞樹 柳井啓司

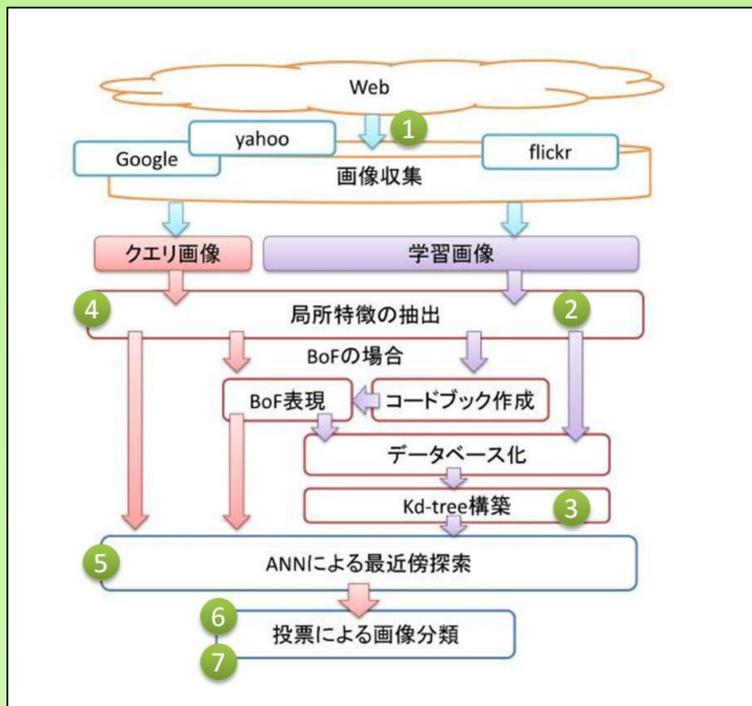
1. 研究の背景と目的

- ◆「一般物体認識」の実現は難しい
 - 「特定物体認識」はある程度可能

- ◆一般物体認識手法を特定物体認識手法で実現

- 特徴マッチングベースの手法

2. システム概要



(1) Web上から画像収集

- Google, yahoo!, Flickr から約15万枚
- 独自に決めた25種類

動物	ネコ	イヌ	ゾウ	ライオン	トラ
車	インプレッサ	レクサス	オデッセイ	パジェロ	プリウス
花	コスモス	タンポポ	ラベンダー	ユリ	バラ
食べ物	ケーキ	ハンバーガー	ピザ	ラーメン	スシ
楽器	ドラム	フルーツ	ギター	ピアノ	バイオリン

(2) 学習画像から局所特徴抽出

- SIFT, PCA-SIFT
- 特徴と抽出画像名のペアをデータベース化
 - 特徴から画像名が引けるようにする
- Bag of Features はSIFT特徴を使用し、visual words を作成コードブックとする。コードブックを用いて学習画像のBoF表現をデータベース化

(3) kd-tree を構築

- Approximate Nearest Neighbor(ANN)使用のための木構造
- 学習画像の特徴を葉ノードとする
- BoFではvisual wordsを葉ノードとする

(4) クエリ画像の局所特徴抽出

(5) ANNによる近似近傍探索

- クエリ画像特徴1つ1つに対して木構造探索により近傍特徴を探索

(6) 近傍特徴をもつ学習画像に投票を行う

- 近傍特徴 n 位までの学習画像に対して投票

(7) 投票結果による画像分類

- 上位k位まで許容した時、クラスに関しての k-Nearest Neighborによってクエリ画像のクラスを認識

3. 実験

◆実験内容

- 上位5クラス分類
 - 動物, 車, 花, 食べ物, 楽器
- 上位クラスに含まれる下位25クラス分類
 - イヌ, レクサス, コスモス, ラーメン, ピアノ...
- SIFT, PCA-SIFT, BoF について実験
- ANNの近傍特徴数 n とkNNの許容範囲k 位の範囲を変化
- 学習画像



	クラス辺りの画像数	総画像数	クラス辺りの特徴数	総特徴数
SIFT	1,050	26,250	600,000	15,000,000
PCA-SIFT	2,900	72,500	2,140,000	53,500,000
BoF	5,800	145,000	-	-

- クエリ画像 - 各下位クラス50枚、計1250枚で実験

◆評価方法

- 再現率: 正しく識別された画像数 / 正しいクエリ画像数
- 適合率: 正しく識別された画像数 / 正しいと識別された画像数
- 分類率: 正しく識別された画像総数 / クエリ画像総数

◆実験結果

手法による比較結果

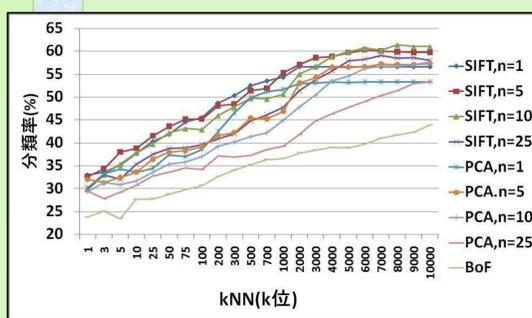
- ベースラインとしてBoF + サポートベクタマシンを使用

	上位クラス分類率	下位クラス分類率
SIFT (n=5, k=7000)	60.1	32.5
PCA (n=5, k=7000)	57.2	29.8
BoF (size=50000)	43.9	25.4
BoF+SVM (線形)	51.7	17.1
BoF+SVM (x^2)	66.9	36.2

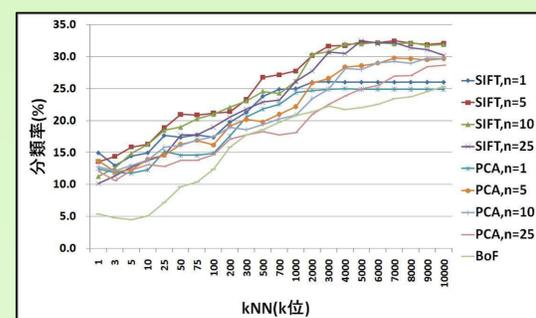
	SIFT 再現率	BoF+SVM 再現率	SIFT 適合率	BoF+SVM 適合率
ピザ	40	64	33	48
ラーメン	18	60	33	35
ドラム	0	22	0	42
トラ	78	70	63	63
ピアノ	70	46	60	46
ギター	58	34	76	35

- 一部のクラスでは良い結果が得られた

ANNの近傍数 n と kNNの範囲 k の変化による精度変化結果



上位5クラス分類



下位25クラス分類

SIFT実験(n=5, k=7000)における再現率と適合率

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	再現率			
動物	1:ネコ	7	4	5	6	3	2	0	2	1	0	0	1	0	5	3	3	1	2	3	0	0	0	1	0	1	14		
動物	2:イヌ	6	2	0	3	2	1	0	0	1	5	0	3	7	4	3	4	3	0	0	0	1	1	0	0	2	4		
動物	3:ゾウ	2	3	9	7	3	2	0	1	0	0	4	2	13	0	0	0	1	1	0	0	0	1	1	0	0	18		
動物	4:ライオン	2	1	1	21	6	0	0	0	0	0	2	3	8	0	0	1	1	1	3	0	0	0	0	0	0	42		
動物	5:トラ	3	0	0	4	39	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	78		
車	6:インプレッサ	0	1	0	2	0	21	3	6	6	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	2	40	
車	7:レクサス	0	0	0	0	0	13	12	7	6	7	0	0	0	0	0	1	0	1	1	1	1	0	0	0	0	24		
車	8:オデッセイ	0	0	0	0	1	10	8	7	7	12	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	14		
車	9:パジェロ	0	0	0	0	0	0	0	7	5	7	15	14	0	0	0	1	0	0	0	0	0	0	0	0	0	0	32	
車	10:プリウス	1	1	2	0	0	5	6	8	5	15	1	0	0	0	0	0	1	0	1	0	0	0	1	2	1	30		
花	11:コスモス	0	0	1	1	0	1	0	1	0	0	28	3	1	4	2	0	0	4	0	1	1	0	0	0	1	56		
花	12:タンポポ	1	0	0	4	0	0	0	1	0	0	9	22	9	0	0	1	0	1	0	0	0	2	0	0	0	44		
花	13:ラベンダー	0	0	2	4	2	0	0	0	0	0	6	30	0	0	2	0	0	0	0	0	0	0	0	0	0	60		
花	14:ユリ	4	1	0	1	0	2	0	1	0	0	6	2	2	12	7	3	0	4	1	2	0	0	0	0	0	24		
花	15:バラ	0	1	0	1	2	0	0	1	0	0	4	0	3	4	25	3	2	1	0	1	0	1	0	0	0	50		
食べ物	16:ケーキ	2	1	2	2	0	0	3	3	2	1	5	0	4	1	2	6	5	4	1	0	0	0	0	0	0	3	12	
食べ物	17:ハンバーガー	2	4	1	4	1	0	0	0	3	0	3	0	1	2	4	9	4	4	3	3	0	1	0	0	0	1	8	
食べ物	18:ピザ	3	3	0	5	1	0	0	0	0	0	3	0	4	0	3	3	2	20	2	0	0	0	0	0	0	1	40	
食べ物	19:ラーメン	1	0	3	1	1	0	2	2	0	1	2	1	2	2	3	9	3	5	9	2	0	0	0	0	0	1	18	
食べ物	20:スシ	0	4	0	1	1	1	1	1	0	1	3	0	1	3	4	13	3	2	1	6	0	1	0	0	0	3	12	
楽器	21:ドラム	3	1	1	1	0	3	10	3	2	1	1	0	3	2	1	1	0	3	2	4	3	2	0	0	2	1	2	0
楽器	22:フルーツ	0	0	0	1	0	7	11	4	4	2	0	0	0	0	0	2	1	2	0	3	0	1	5	0	0	10		
楽器	23:ギター	1	1	0	0	0	3	0	2	2	5	0	0	0	0	0	0	0	0	1	0	1	0	1	0	1	29	3	
楽器	24:ピアノ	0	0	0	0	0	4	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	35	0	
楽器	25:バイオリン	0	0	1	1	0	5	4	5	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	26	52	
適合率		18	7	32	30	63	25	17	11	27	19	35	59	32	33	40	10	11	33	33	32	0	38	76	60	55	32	5	

4. まとめ、今後の課題

◆まとめ

- ベースラインには少し及ばない結果となったが、一部のクラスに関してはよい分類精度が得られた
- 一般的な特定物体認識手法を利用しても、一般物体認識がある程度実現可能であると分かった

◆今後の課題

- 学習画像数を増やす
- BoFのコードブックサイズの変化