OBJECT CATEGORIZATION BY LOCAL FEATURE MATCHING WITH A LARGE NUMBER OF WEB IMAGES

Mizuki Akiyama, Yoshiyuki Kawano, Keiji Yanai
Department of Informatics
The University of Electroo-Communication,
Tokyo, JAPAN

Background

- Object Recognition
  - Generic object recognition - object, scene, face, …

TIGER

  - Specific object recognition - DB search, name search

Background

- Object Recognition
  - Generic object recognition – object, scene, face, …

Generic Object Recognition

  - Specific object recognition – DB search, name search

Specific Object Recognition

Background

- Specific Object Recognition
  - Large-scale Image database + local feature matching
  - High precision for recognition of specific-shaped objects
- Generic Object Recognition
  - Small-scale images database + machine learning
  - Low precision due to inter-class ambiguity
- Specific object recognition applies to generic object recognition.

Goal

- Generic object recognition
  - **Local feature matching based specific object recognition**
  - **Large-scale training and categorization**
    - 150,000 training images

Related Work

- Object recognition using a large amount of images
  - Specific object recognition method to landmark database creation[Zheng et al, ICCV2009].
  - 20 million geo-tagged images on the Web
    - Search a very similar image
  - SIFT-matching based image search

Related Work

- Generic object recognition using a large amount of image data
  - 80 million images [A. Torralba et al., PAMI (2008)]
  - Image categorization by k-NN with 32x32 tiny images
  - Comparable performance to the state-of-the-art method

  - K-NN method with a very large amount of image data was one of promising approaches for generic object recognition.

Related Work

- Our work is inspired by the “80 million images”
- 80 million images
  - Use the sum of squared differences (SAD) between 32x32 tiny images
- Our work:
  - Apply SIFT-based local feature matching and voting
  - Collect a large number of images from the Web
  - Use them for experiments on image categorization without excluding noise images

Overview

Image collection + Feature extraction	create a database	feature matching	recognition
▪Collect 150 ,000 from the Web ▪Extract local features	▪Local features of training images stored in the database	▪Feature matching for test images ▪Vote the k most similar samples stored in the database	▪Majority based categorization

Detail

- Image collection
  - 25 categories, total 150,000
  - Source: Google,Yahoo!,Flickr
  - noise images included




Detail

- Feature extraction
  - SIFT
    - 128 dimension, invariant to rotation, scale-change and illumination change
  - PCA-SIFT
    - 36 dimension, extension of SIFT
  - Bag-of-Features (BoF)
    - Many local patches, and vector-quantizing
    - Explore the best setting of the codebook size k

Detail

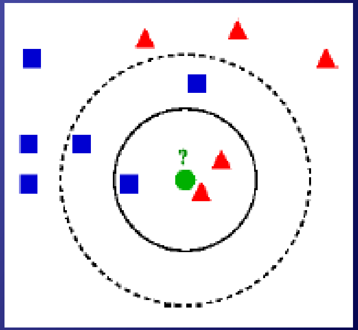
- Database
  - kd-tree
    - SIFT, PCA-SIFT
  - Inverted index
    - BoF

Detail

- Feature matching
  - Simple NN search is very costly
  - Approximate Nearest Neighbor(ANN)
    - Kd-tree based approximate nearest neighbor search method
    - Search the top n nearest points for each query local feature point
    - Vote on the image from which the nearest local features are extracted

Detail

- Recognition
  - Sort the image having votes in the descending order of the number of votes
  - Decide one of the given categories by applying k-Nearest Neighbor classification
  - k-Nearest Neighbor
    - Majority of the categories of the top k samples



Experiments

- 5 and 25 class categorization
- Feature representation
  - SIFT,PCA-SIFT,BoF
- parameters
  - ANN top n and k-NN top k
  - Codebook size

Dataset

- 5 and 25 categories Dataset
  - Examples of images

Experiments

- The number of training images
  - 32GB linux machine

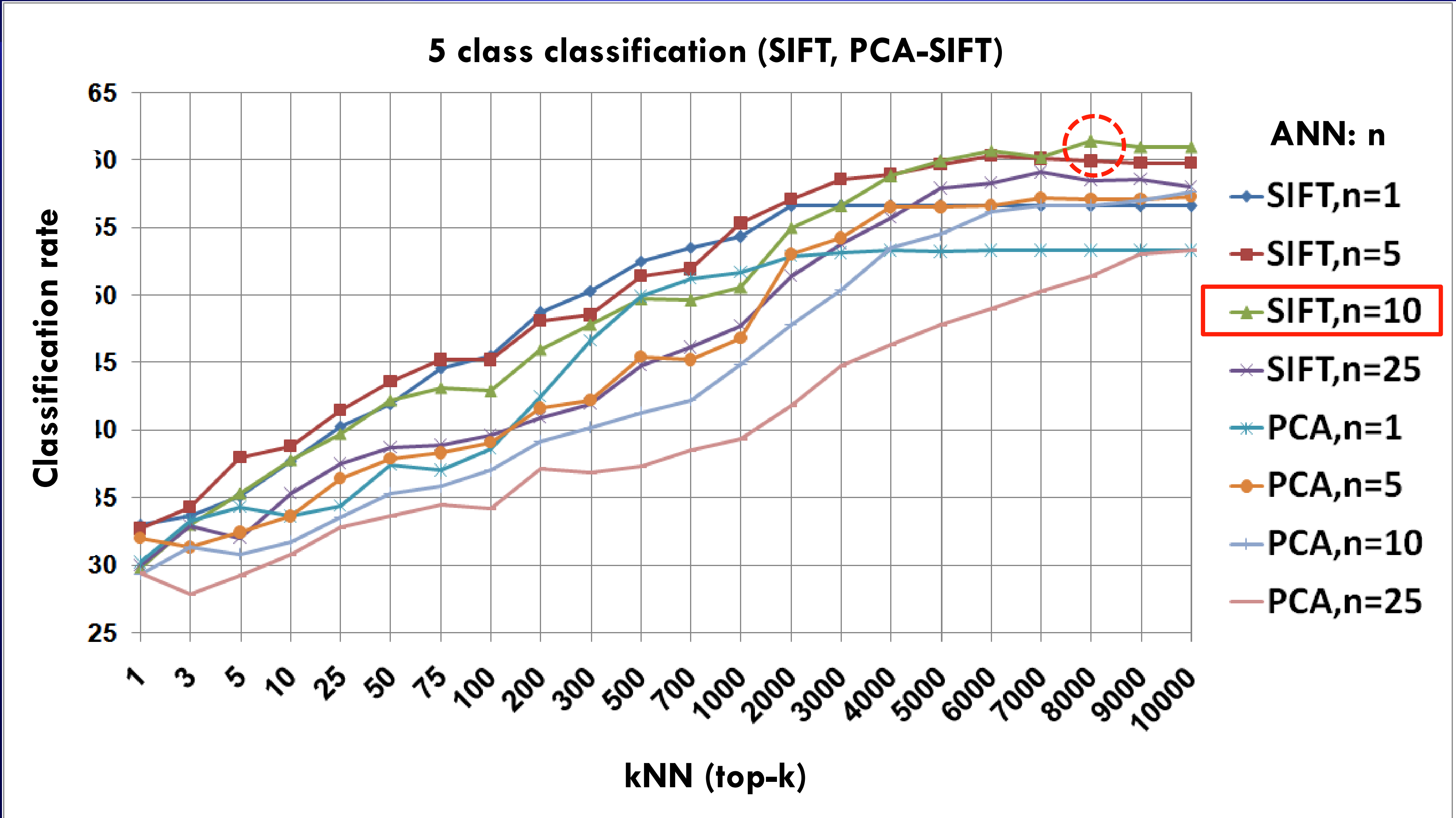
	# of training images	# of images per class	# of local features	memory
SIFT	26,250	1,050	15million	20GB
PCA-SIFT	73,500	2,940	53.50million	25GB
BoF	145,000	5,800	-	5GB

- Evaluation
  - Recall, precision, classification rate
  - Baseline: Bag-of-Features + SVM

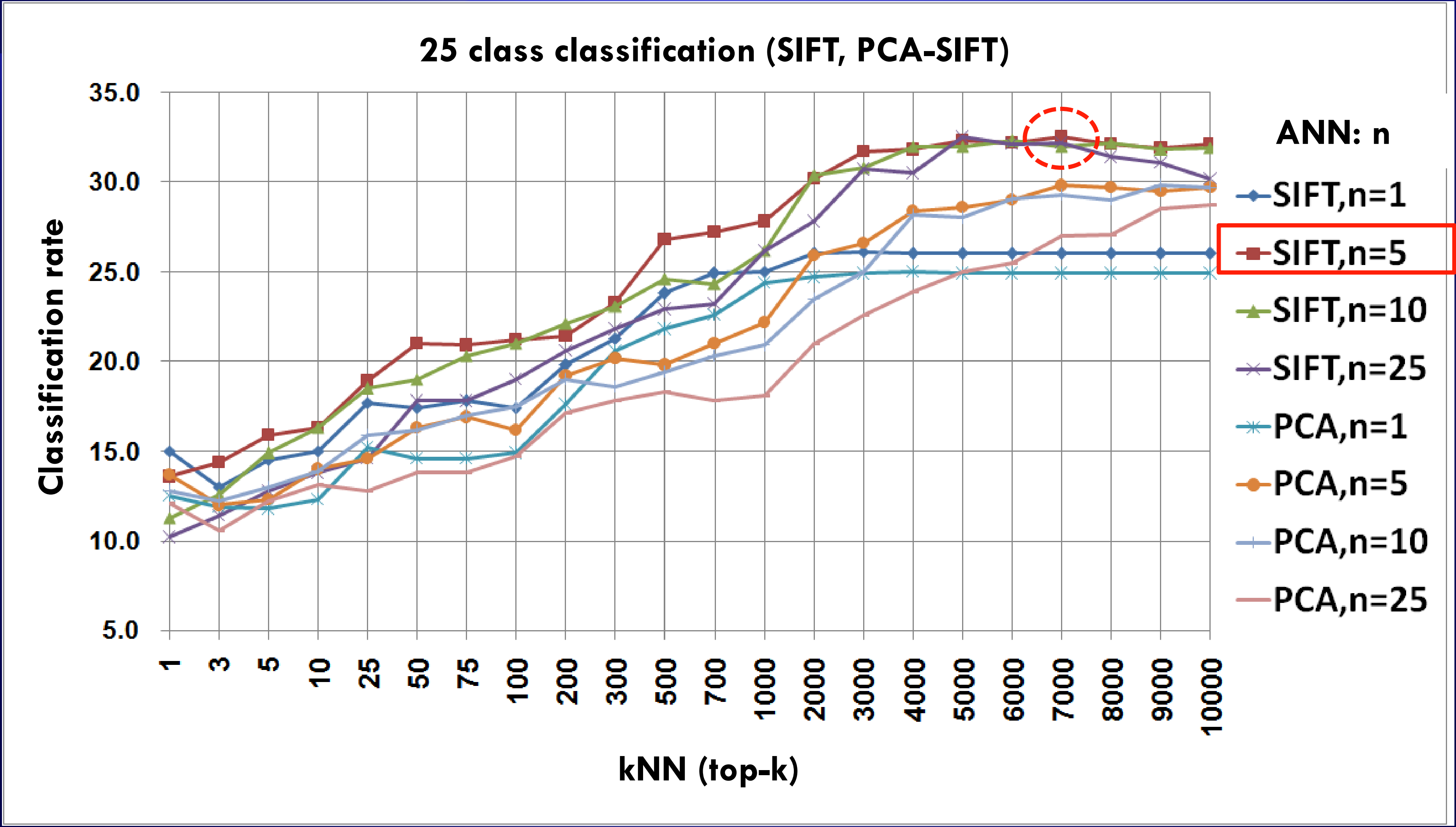
Experimental Results

	5 class Classification rate(%)	25 class Classification rate(%)
SIFT(n=5,k=7,000) Proposed	60.1	32.5
PCA(n=5,k=7,000) Proposed	57.2	29.8
BoF(size=200,000,k=20,000) Proposed	54.9	30.7
BoF+SVM(linear kernel) Baseline	51.7	17.1
BoF+SVM(χ^2 kernel) Baseline	66.9	36.2

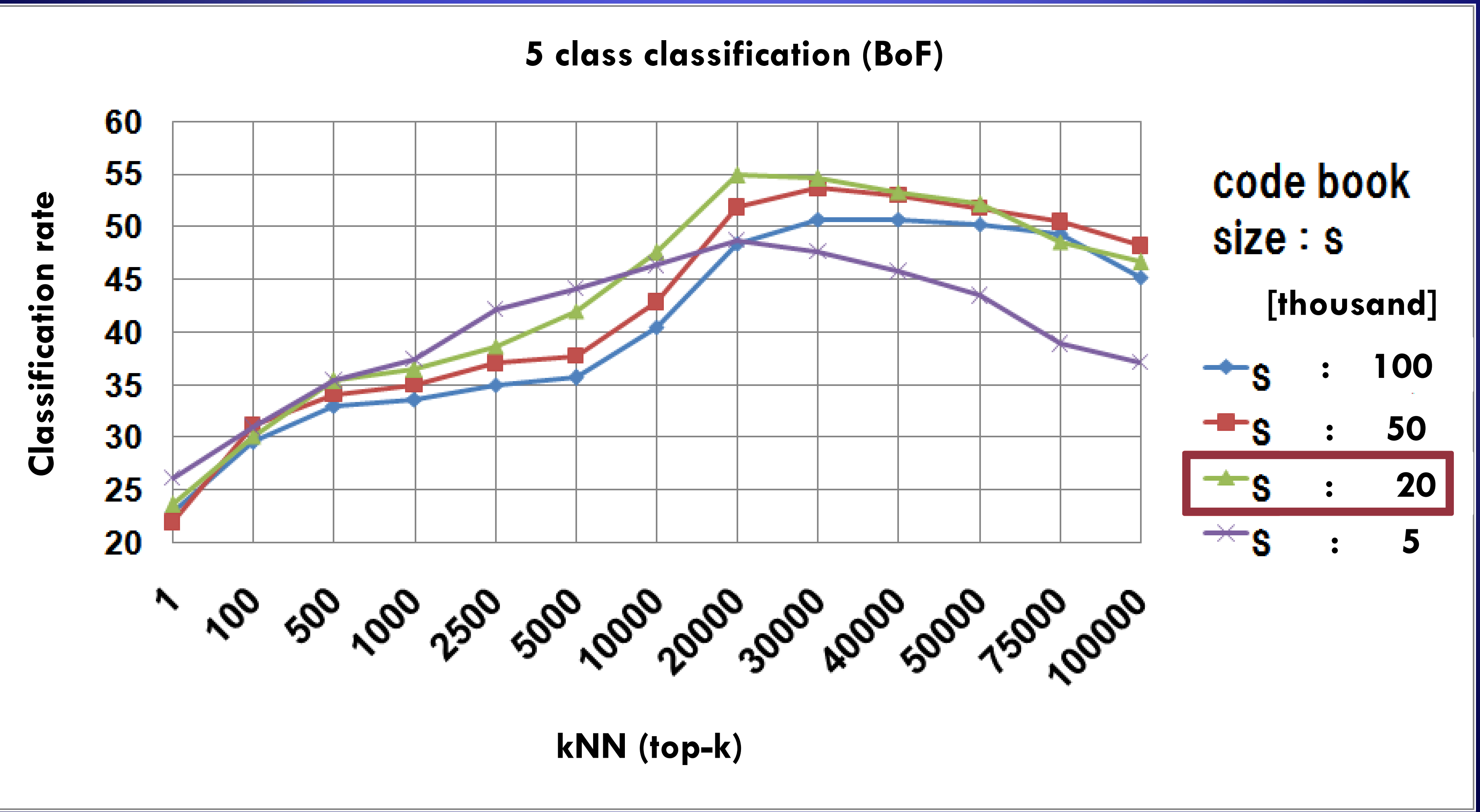
Experimental Results(SIFT,PCA,5class)



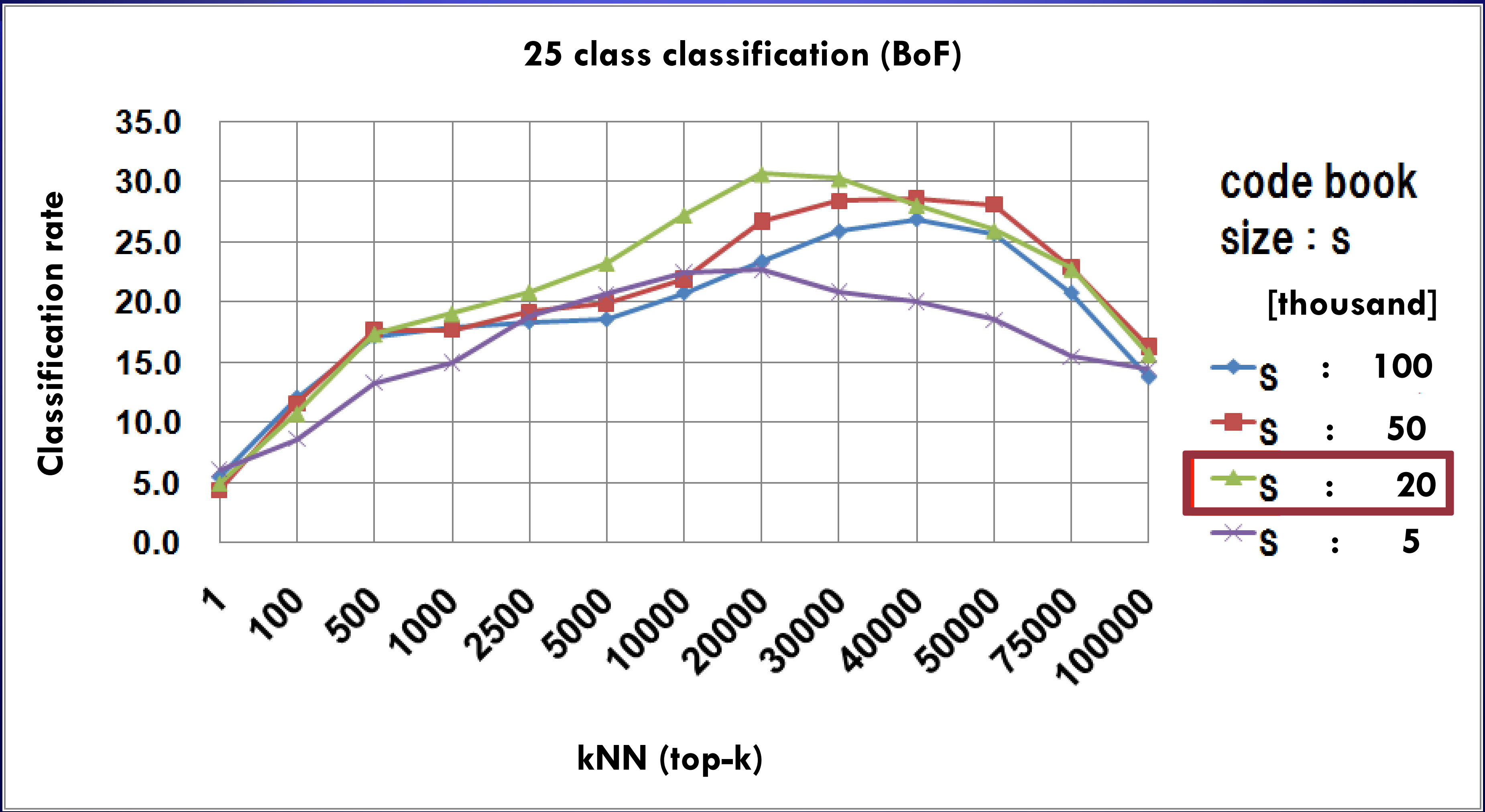
Experimental Results(SIFT,PCA,25class)



Experimental Results(BoF,5class)



Experimental Results(BoF,25class)



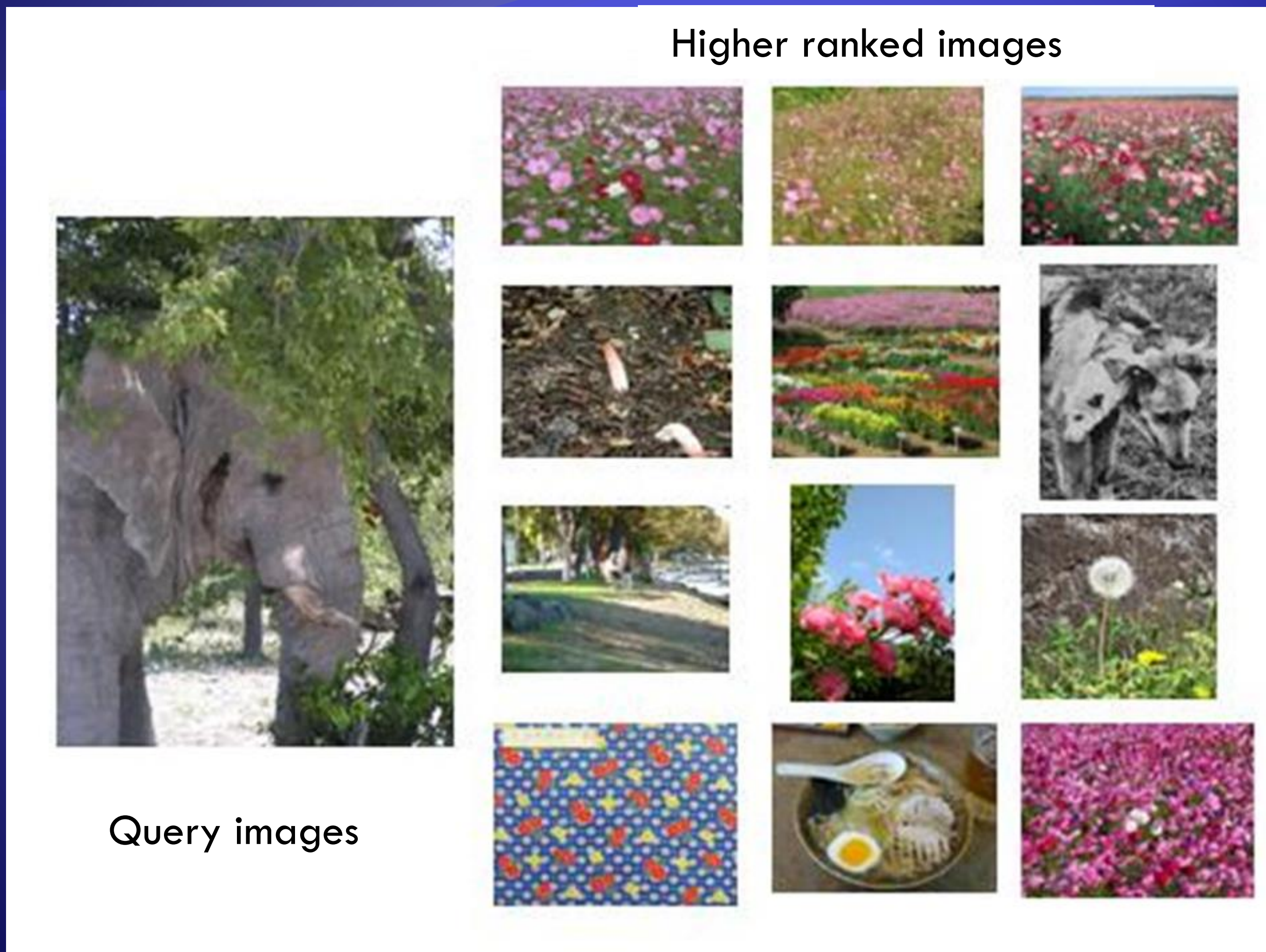
Experimental Result (SIFT,5class)

	1	2	3	4	5	Recall(%)
1: Animal	155	14	37	37	7	62
2: Car	10	228	3	4	5	91
3: Flower	35	15	150	41	9	60
4: Food	43	24	40	135	8	54
5: Musical instrument	12	128	10	14	86	34
Precision(%)	61	56	63	58	75	60.3

Experimental Results (SIFT,25class)

		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	Recall(%)
Animal	1: Cat	7	4	5	6	3	2	0	2	1	0	1	0	5	3	3	1	2	3	0	0	0	0	1	0	1	14
	2: Dog	6	2	0	3	2	1	0	0	1	5	0	3	7	4	3	4	3	0	0	2	0	1	1	0	2	4
	3: Elephant	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
	4: Lion	2	1	1	21	6	0	0	0	0	0	2	3	8	0	1	1	1	3	0	0	0	0	0	0	0	42
	5: Tiger	3	0	0	4	39	1	0	0	0	0	0	0	2	0	0	0	0	1	0	0	0	0	0	0	0	78
Car	6: Impreza	0	1	0	2	0	21	3	6	6	6	1	0	1	0	0	0	0	0	1	0	0	0	0	2	0	42
	7: Lexus	0	0	0	0	0	13	12	7	6	7	0	0	0	0	1	0	1	1	1	1	0	0	0	0	0	24
	8: Odyssey	0	0	0	0	1	10	8	7	7	12	0	0	0	0	0	1	1	0	1	0	0	0	1	1	0	14
	9: Pajero	0	0	0	0	0	7	5	7	16	14	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	32
	10: Prius	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
Flower	11: Cosmos	0	0	1	1	0	1	0	1	0	0	28	3	1	4	2	0	0	4	0	1	1	0	0	1	1	56
	12: Dandelion	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
	13: Lavender	0	0	2	4	2	0	0	0	0	0	6	0	30	0	0	2	0	2	0	0	0	0	0	2	0	60
	14: Lily	4	1	0	1	0	2	0	1	0	0	6	2	2	12	7	3	0	4	1	2	0	0	0	2	0	24
	15: Rose	0	1	0	1	2	0	0	1	0	0	4	0	3	4	25	3	2	1	0	1	0	1	0	0	1	50
Food	16: Cake	2	1	2	2	0	0	3	3	2	1	5	0	4	1	2	6	5	4	1	0	0	0	0	3	3	12
	17: Hamburger	2	4	1	4	1	0	0	0	3	0	3	0	1	2	4	9	4	4	3	3	0	1	0	0	1	8
	18: Pizza	3	3	0	5	1	0	0	0	0	0	3	0	4	0	3	3	2	20	2	0	0	0	0	0	1	40
	19: Ramen	1	0	3	1	1	0	2	2	0	1	2	1	2	2	3	9	3	5	9	2	0	0	0	1	0	18
	20: Sushi	0	4	0	1	1	1	1	1	0	1	3	0	1	3	4	13	3	2	1	6	0	1	0	0	3	12
Musical instrument	21: Dram	3	1	1	1	0	3	10	3	2	3	2	1	1	0	3	2	4	3	2	0	0	0	2	1	2	0
	22: Flute	0	0	0	1	0	7	11	4	4	2	0	0	0	0	2	1	2	0	3	0	1	5	0	5	2	10
	23: Guitar	1	1	0	0	0	3	0	2	2	5	0	0	0	0	0	0	0	1	0	1	0	1	29	1	3	58
	24: Piano	0	0	0	0	0	0	4	3	3	3	0	0	0	0	0	0	0	0	1	0	0	0	1	35	0	70
	25: Violin	0	0	1	1	0	5	4	5	1	2	0	0	0	0	0	2	0	0	0	0	0	0	1	2	26	52
	Precision (%)	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

Experimental Results(SIFT,25class)



Conclusions

- Generic object recognition by feature matching based specific object recognition

- It performed well with a large number of sample images
  - Almost equivalent to the result by generic recognition
    - 5class - 60.3%, 25class - 32.5%

Future works

- More large-scale experiments
  - Using parallel computing
- informative feature
  - Use only discriminative features for image categorization.

Experimental Results(SIFT,25class)