NLab UTokyo at ImageCLEF2013 Plant Identification Challenge

Augmenting Descriptors for Fine-grained Classification

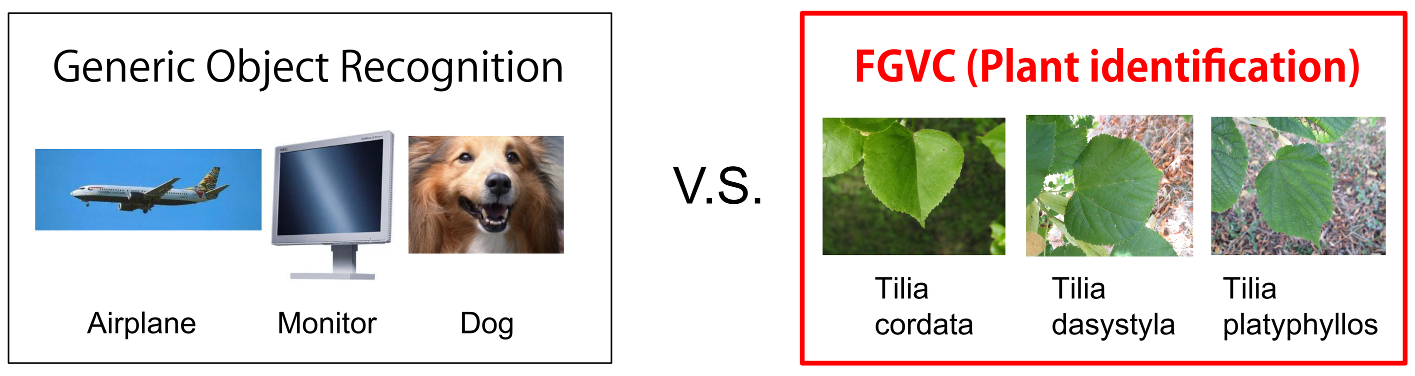
Hideki Nakayama The University of Tokyo
nakayama@ci.i.u-tokyo.ac.jp

THE UNIVERSITY OF TOKYO

Overview

Fine-grained Visual Categorization (FGVC)

- Distinguish visually very similar categories under a specific domain.
  Plant species identification is a typical example.
- Closely related to applications.
- Complementary to traditional generic object recognition.



We implemented our recently proposed FGVC method for this challenge.

Hideki Nakayama, "Augmenting descriptors for fine-grained visual categorization using polynomial embedding", *Proc. IEEE ICME*, 2013.

Motivation

We focus on developing discriminative image signatures.

① **Co-occurrence information of local features**

[Sabzmeydani et al., 2007] [Tuzel et al., 2006] [Harada et al., 2012] etc.

☺ Captures middle-level discriminative information

☹ Results in high-dimensional local representation

② **Fisher Vector** [Perronnin et al., 2010]

= bag-of-words encoding using higher-order statistics.

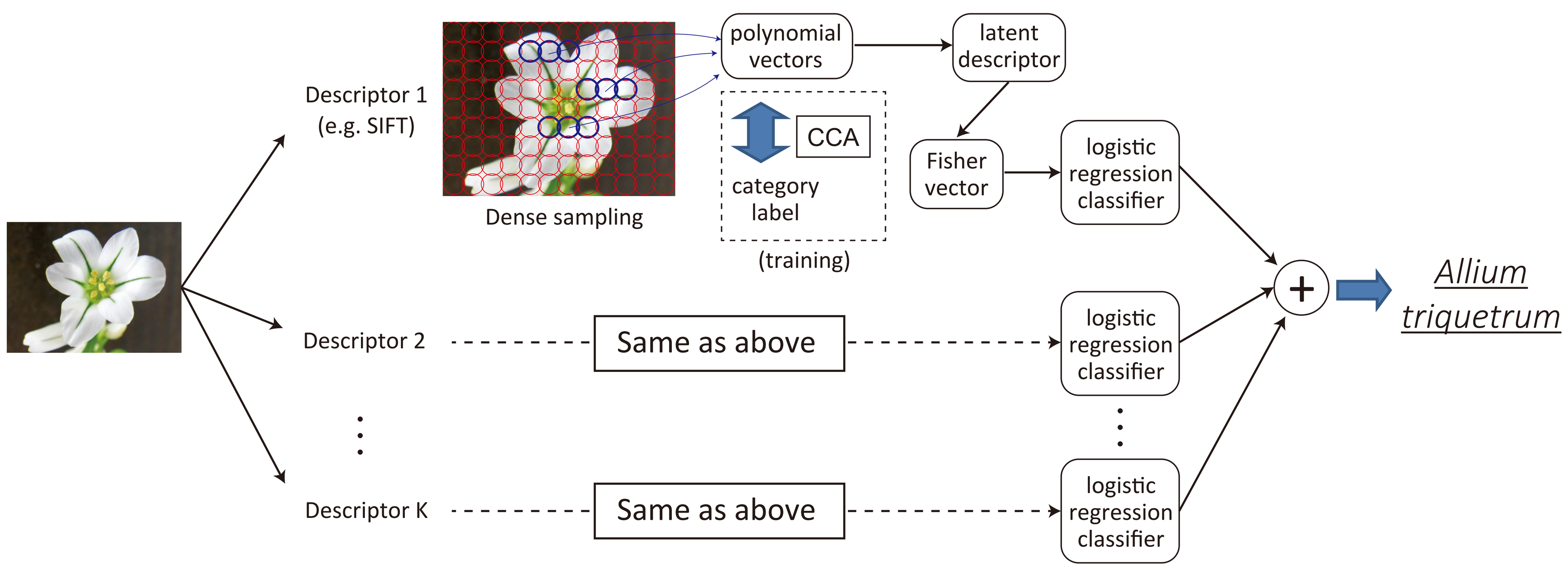
☺ Remarkably high-performance, enables linear classification

☹ Dimensionality increases in linear to the size of local representation

How to relax these problems?

Compress co-occurrence patterns of neighboring local descriptors into a discriminative latent descriptor.

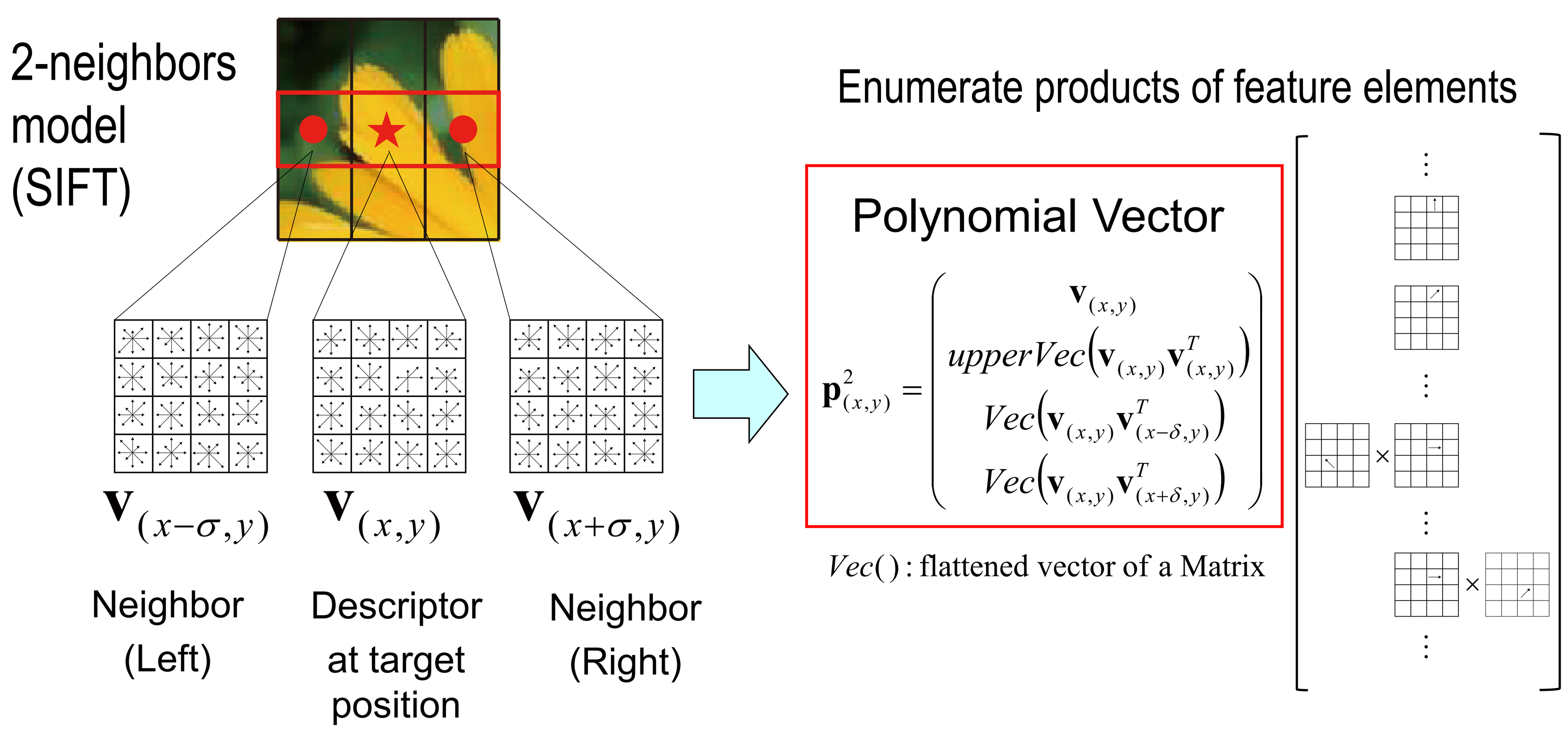
Proposed Method



Summary

1. Densely extract multiple types of local descriptors from images. They are compressed into 64 dimensions via PCA respectively. We use SIFT, C-SIFT, opponent-SIFT, HSV-SIFT and self-similarity (SSIM).
2. For each type of descriptors, take the polynomial vector using the co-occurrence patterns of neighboring descriptors, which is further embedded into a 64-dimensional latent descriptor via CCA.
3. Compute Fisher Vector using the latent descriptor.
4. Train logistic regression classifiers for each type of local descriptors. Classification is conducted by means of the average of scores output by multiple classifiers.

Taking co-occurrence patterns

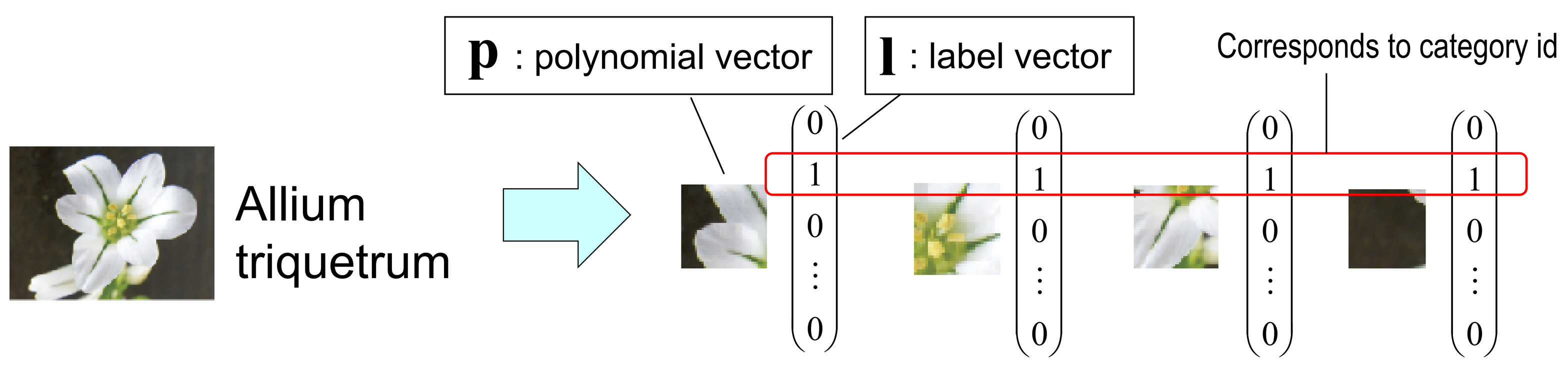


$\mathbf{v}_{(x-\sigma,y)}$ $\mathbf{v}_{(x,y)}$ $\mathbf{v}_{(x+\sigma,y)}$

$$\mathbf{p}^2_{(x,y)} = \begin{pmatrix} \mathbf{v}_{(x,y)} \\ upperVec(\mathbf{v}_{(x,y)}\mathbf{v}^T_{(x,y)}) \\ Vec(\mathbf{v}_{(x,y)}\mathbf{v}^T_{(x-\delta,y)}) \\ Vec(\mathbf{v}_{(x,y)}\mathbf{v}^T_{(x+\delta,y)}) \end{pmatrix}$$

$Vec()$: flattened vector of a Matrix

Supervised dimensionality reduction



- Image-level category label is shared by all patches within an image (Approximate approach).
- Somewhat reasonable for FGVC : Users will more or less target objects closely.

Canonical Correlation Analysis (CCA)

$$C_{pl}C_{ll}^{-1}C_{lp}A = C_{pp}A\Lambda^2 \quad (A^T C_{pp} A = I)$$

Latent descriptor

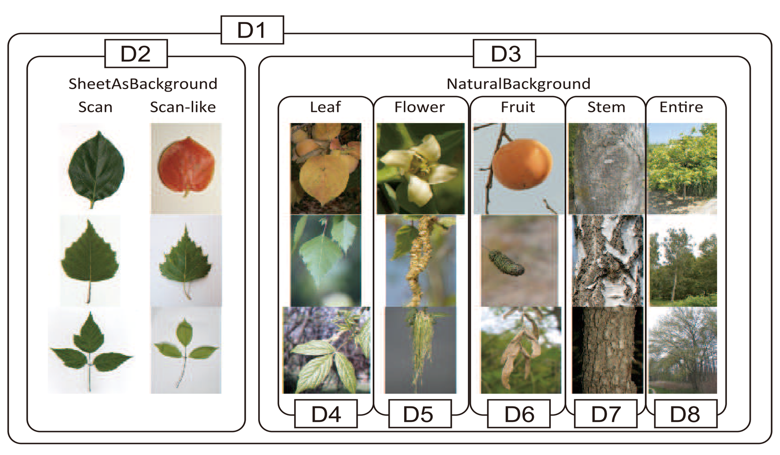
64 dim $\mathbf{s} = A^T(\mathbf{p} - \bar{\mathbf{p}})$ 1,000~10,000 dim

Plant Identification

Goal:
Classify 250 plant species using different views (organs).

Two main categories:
Sheet As Background (SAB)
Natural Background (NB)

Our runs



We trained multi-class classifiers for some specific domains independently in each run. (Domains for testing samples are provided)

Run 1: D1 (All)
Run 2: D2 (SAB), D3 (NB)
Run 3: D2 (SAB), D4~D8 (NB subcat.)

Natural Background category consists of generic natural image and was our primary target.

Validation

We roughly took 10% of training samples for validation to tune parameters and select appropriate combinations of descriptors.

Our features consistently improves performance from the baseline (i.e., standard Fisher Vector) for all types of descriptors and domains.

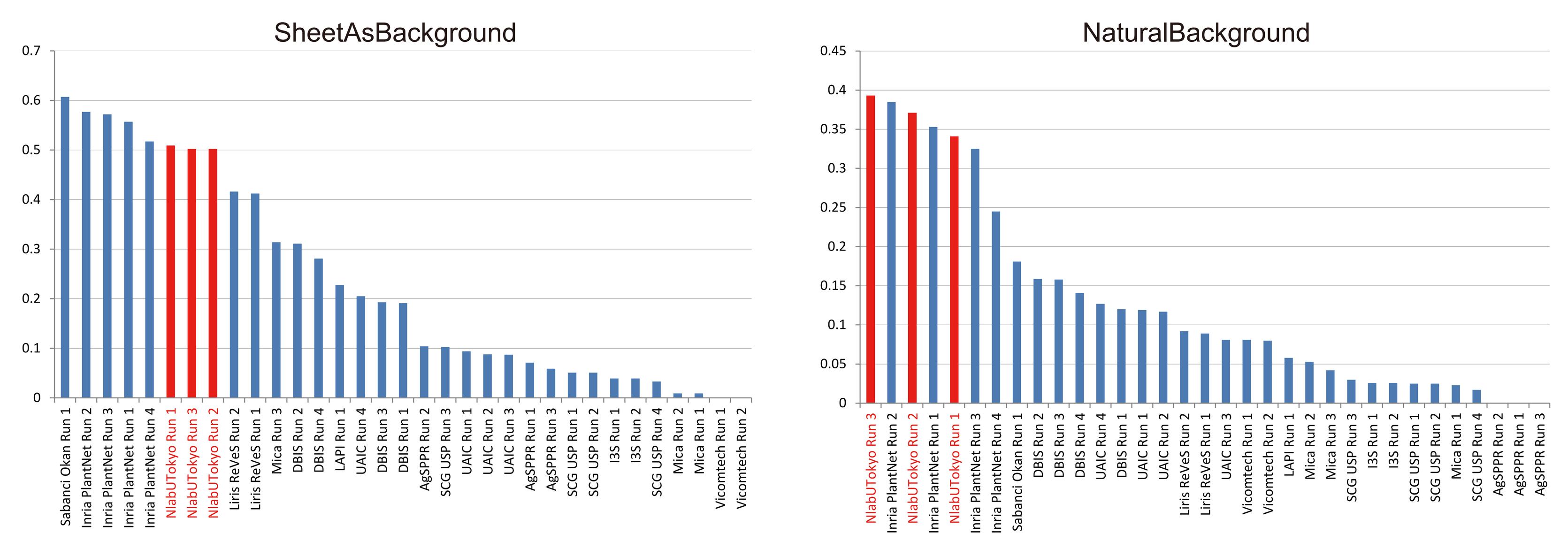
Used descriptor types and classification rates (%) on the validation set.

	SIFT	C-SIFT	Opp.-SIFT	HSV-SIFT	SSIM	Baseline	Ours
D1	✔	✔	✔	✔		38.2	38.8
D2	✔					50.8	52.5
D3		✔	✔	✔	✔	15.9	17.8
D4		✔	✔	✔	✔	15.2	17.3
D5		✔	✔	✔	✔	21.2	24.7
D6		✔	✔	✔	✔	7.4	11.1
D7		✔	✔	✔	✔	13.8	16.5
D8		✔	✔	✔	✔	8.2	8.5

Final results

We achieved:

- The 1st place in Natural Background category (and 4/5 subcategories).
- The 3rd place in Sheet As Background category.



Scores in Natural Background task

	Leaf	Flower	Fruit	Stem	Entire	All
Ours	0.275	0.472	0.311	0.253	0.297	0.393
INRIA PlantNet	0.272	0.494	0.260	0.240	0.274	0.385
Sabanci Okan	0.049	0.223	0.194	0.106	0.174	0.181

Discussion

During this challenge, we bet on implementing powerful image features and obtained promising results. On the other hand, our learning and classification algorithms are very simple and could be improved. Although some instances have multiple images of different organs, our system treats them independently and loses co-occurrence information. It would be interesting to develop classification methods utilizing them in an integrated manner.