

1. Aims of the Project

- To propose an innovative way that combines global and local point features to classify leaf images.
- To investigate the relationship between leaf shape contour and its vein pattern in the proposed system.
- To investigate the system using large volume training dataset.

2. Introduction

The recent development of digital camera and growth of data storage has led to a huge amount of leaf image accessible over internet.

And botanists need a computer-aided tool to study and identify leaves instead of holding a plant encyclopaedia. The machine learning techniques enable automatic leaf classification system which doesn't need human's interaction.

Previous approaches only focus on one local or one global feature.

My approach: extract contour and vein, local and global descriptor for each sample point.

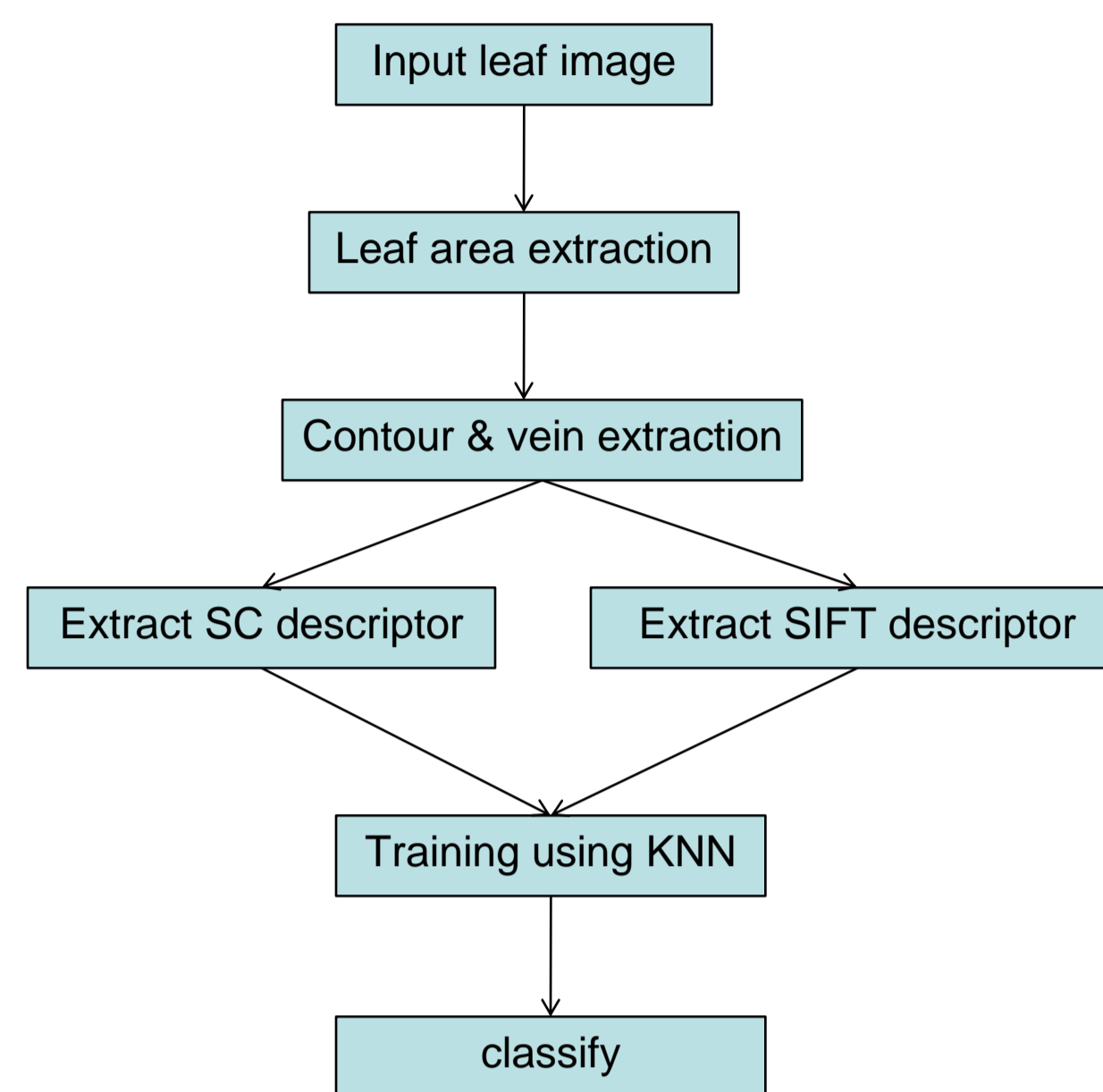
3. Methodology

Instead of using only one feature, my approach uses two features, local and global. Also, taking advantage of vein pattern. The purpose of doing this is to let both features complement with each other. This keeps the classifier from leaning to too local or too global.



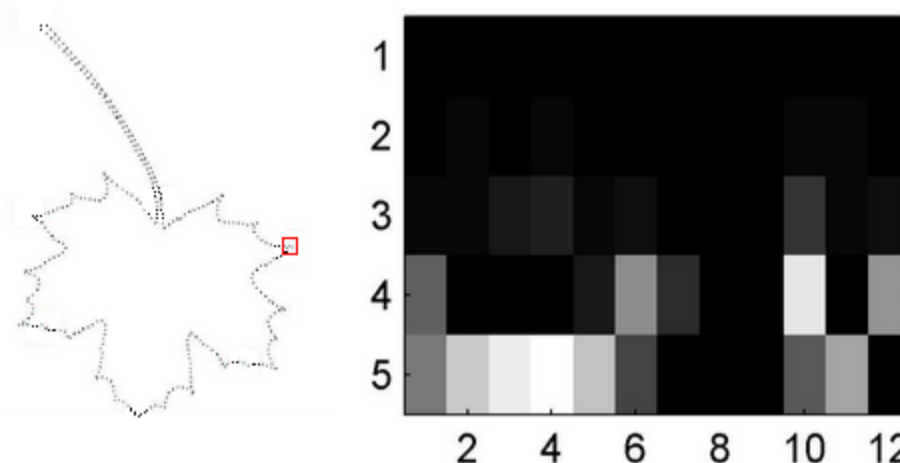
Example with extracted contour and vein

3.1 System overview



3.2 Global Shape Context descriptor

Shape context (SC) is a powerful tool for object recognition tasks. Shape context analysis begins by taking N samples from leaf contour or vein. It looks at a set of vectors from one point p_k to all other points p_i with $i \neq k$. This set of vectors is a rich discriminative descriptor of the shape for point p_k . The following picture shows the SC histogram of the red point.



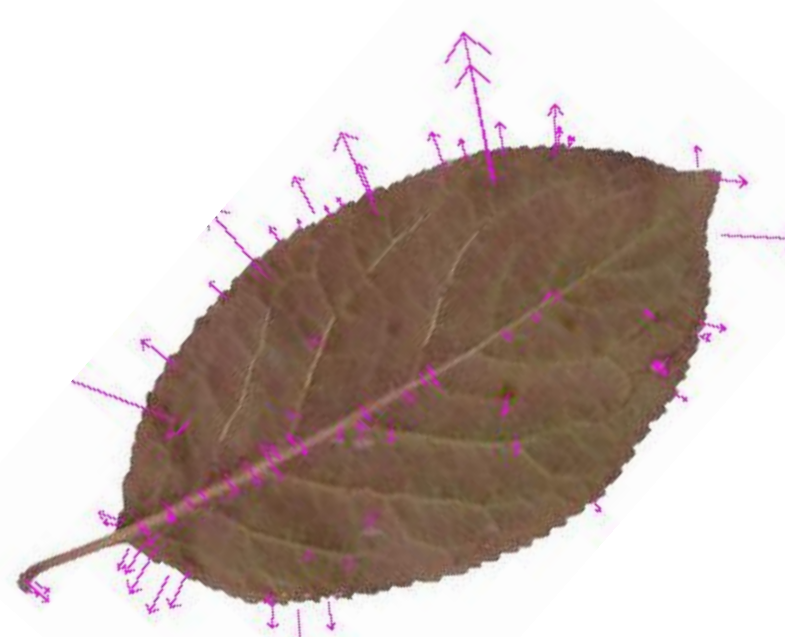
Leaf with SC histogram

3.3 Local SIFT descriptor

As proposed, we need a local descriptor for points to capture the leaf information. SIFT is our choice. SIFT

Properties of SIFT

- High discriminative
- Invariant to scale, orientation.



Leaf with SIFT

4. Experiment

Dataset:

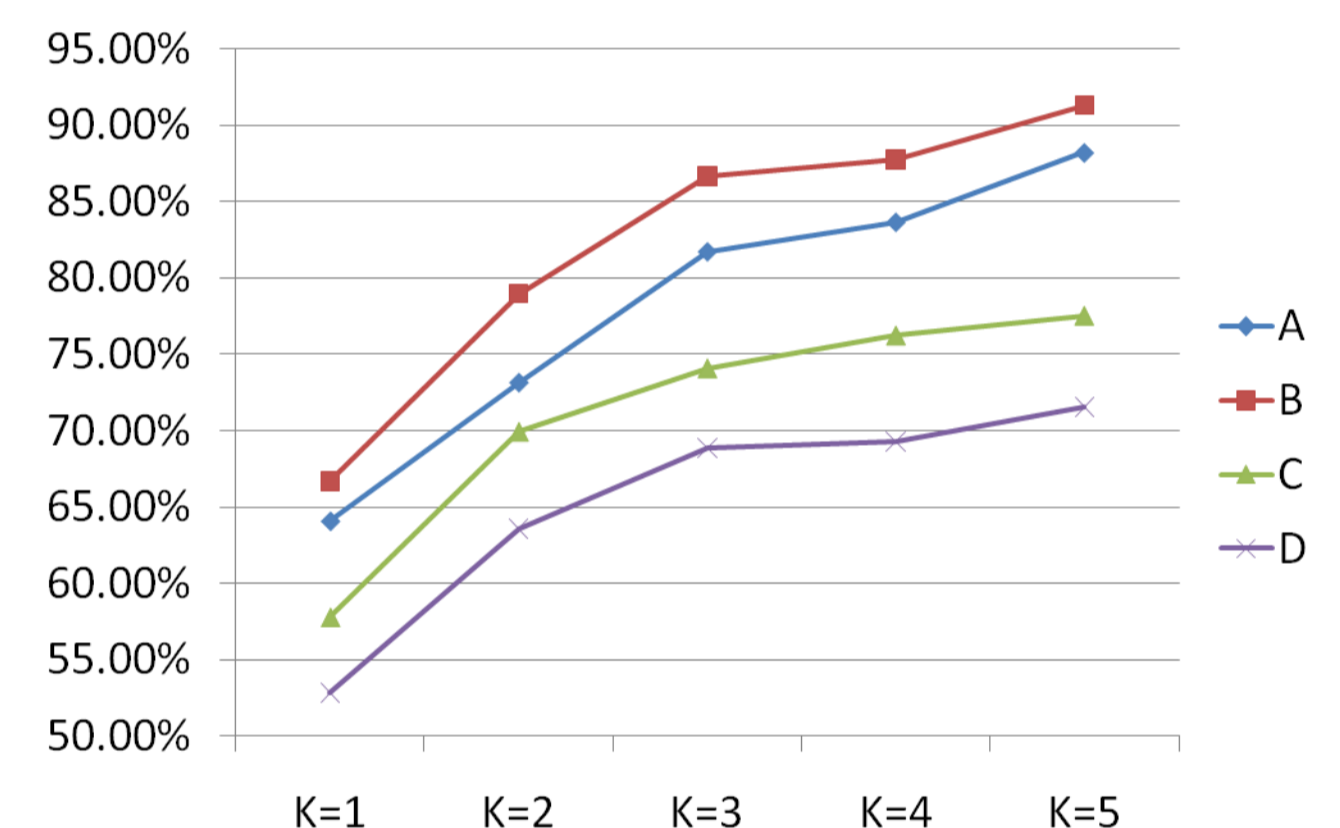
- 220 distinct leaf classes with a total 17073 images.
- For performance concern, 3300 images for training, while 2200 for testing.

Pre-processing: colour based EM algorithm are used to detect the leaf area.

The experiments are designed to investigate using SC and SIFT on leaf contour and contour with vein. The classification based on KNN.

- A 200 points contour + SC + SIFT
- B 200 points contour + vein + SC + SIFT
- C 100 points contour + SC + SIFT
- D 100 points contour + vein + SC + SIFT

5. Results



6. Major findings

- Utilizing SC and SIFT is helpful for leaf classification. (results A, B, C and D)
- Taking vein pattern into consideration normally boosts the accuracy of classifier unless the number of sampling point is inadequate.
- The combination of global and local point feature complement each other and yield promising result.

7. Future Work

- Use more accuracy techniques to extract leaf veins.
- Use colour to guard the classifier since neither SC or SIFT consider colour of leaves.
- Instead of random point sampling, pick up sampling points from corresponding positions of each leaf.