

Plant Leaf Recognition Based on Contourlet Transform and Support Vector Machine

Ze-Xue Li¹(✉), Xiao-Ping Zhang¹, Li Shang², Zhi-Kai Huang³,
Hao-Dong Zhu⁴, and Yong Gan⁴

¹ College of Electronics and Information Engineering, Tongji University,
4800 Cao'an Highway, Jiading, Shanghai, China
lizexue0@163.com

² Department of Communication Technology, College of Electronic Information
Engineering, Suzhou Vocational University, Suzhou 215104, Jiangsu, China

³ College of Mechanical and Electrical Engineering,
Nanchang Institute of Technology, Nanchang 330099, Jiangxi, China

⁴ College of Computer and Communication Engineering,
Zhengzhou University of Light Industry, Zhengzhou, China
ganyong@zzuli.edu.cn

Abstract. Plants are essential to the balance of nature and in people's lives as the fundamental provider for food, oxygen and energy. The study of plants is also essential for environmental protection and helping farmers increase the production of food. As a fundamental task in botanical study, plant leaf recognition has been a hot research topic in these years. In this paper, we propose a new method based on contourlet transform and Support Vector Machine (SVM) for leaf recognition. Contourlet Transform is a promising multi-resolution analysis technique, which provides image with a flexible anisotropy and directional expansion. By basing its constructive principle on a non-subsampled pyramid structure and related directional filter banks, contourlet transform decomposes input images into multi-scale factors which also enjoys additional advantages such as shift invariance and computational efficiency. Compared with one-dimensional transforms, such as the Fourier and wavelet transforms, Contourlet Transform can capture the intrinsic geometrical structure. In order to ameliorate the influence of unwanted artefacts such as illumination and translation variations, in this paper, the contourlet transform was firstly applied to extract feature with high discriminative power. Then the extracted features are classified by SVM. The experimental results show that the proposed method has high sensitivity of directionality and can better capture the rich features of natural images such as edges, curves and contours.

Keywords: Leaf recognition · Contourlet Transform (CT) · Support vector machine (SVM) · Contourlet Transform · Frames · Image denoising · Image enhancement · Multidimensional filter banks · Nonsubsampled filter banks

1 Introduction

Studying and exploiting the plant recognition has been one of the most important tasks in plant protection. As we know, the crucial stage of plant taxonomy is a genuine scientific and technical challenge, due not only to the huge number of plant species, but also to their highly specialized and diverse taxonomic properties. For this reason, the efficiency of manual plant recognition is too low and we should introduce pattern recognition technology to carry out this work. One key distinguishing feature for the identification of plant species is plant information obtained from leaf images [1–6]. Considering the different extracted features of leaf images, the recognition method can be roughly divided into 3 kinds - based on texture features, subspace projection and statistical features [7–10].

The method based on the structure feature need pre-processing and extracts the texture feature of leaf images. Such methods require complex pretreatment process, and the pretreatment results will influence the accuracy of recognition seriously. Although the texture features can obtain certain recognition accuracy, they are sensitive to the position and orientation changes during the collection process, which lacks stability and robustness [11, 12].

The subspace projection method applies Principal Component Analysis (PCA), Independent Component Analysis (ICA) or linear classification analysis in a certain transform domain of the leaf images. Then choose a proper classifier to take this recognition work by the projection coefficient as the feature. Compared with the method based on the structure, subspace projection method has anti-noise-interference ability without complex pretreatment process. But the changes of location and direction maybe interfered the ability of recognition.

This paper is organized as follows. We throw the concept and principle of Contourlet Transform and propose in Sect. 2 Contourlet Transform to decompose the leaf image, and then introduce the low frequency sub-band feature extraction and high frequency sub-band feature extraction. In Sect. 3, we provide a classifier - Support Vector Machine (SVM) Classifier. The logic of SVM is expatiated in Sect. 3. And then in Sect. 4, the experimental database and results are list. At last, we conclude this paper in Sect. 5.

2 Contourlet Transform

The research results by neural physiologist show that acceptant fields in the visual cortex are characterized as being localized, oriented, and bandpass for the visual system of human. There are experiments suggested that it will be efficient for a computational image representation, if it based on a local, directional, and multiresolution expansion.

As two-dimensional wavelets are constructed from tensor products of one-dimensional wavelets, with the finer resolution, we can clearly find the limitation of the wavelet that it needs to use many special dots to capture the contour, as show in the left of Fig. 1. The new scheme (in the right of Fig. 1.) shows that the support interval of baseband should behave as a long strip shape in order to make full use of the geometrical transformation of the original function and achieve with the least

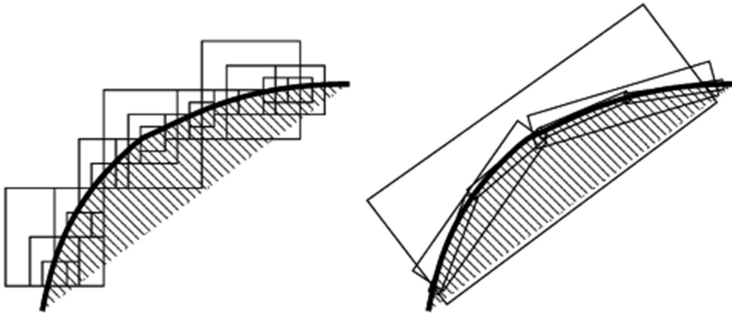


Fig. 1. Wavelet versus new scheme

coefficients to approximate the singular curve. In fact, the elongated support interval of baseband is a reflection of directionality, and this also called multi-scale geometric analysis.

2.1 Feature Extraction

The Contourlet Transform is implemented by a double filter bank named pyramidal directional filter bank (PDFB). PDFB can be seen as a cascade of two steps. Firstly the original image is multi-scaled decomposed into low frequency and high frequency subbands by Laplacian Pyramid (LP) transform. Then the Directional Filter Banks (DFB) decompose bandpass signal of each level in Pyramid into tree structure of L layer, and the band will be divided into two directions in each layer. The singular points distributed in the same direction will be synthesized as one coefficient. It achieved the image sparse representation more effectively that Contourlet Transform combined LP and DFB into the double filter group structure. The Contourlet Transform can be implemented iteratively applying PDFB on the coarse scale of image, as shown in Fig. 2 [20, 21].

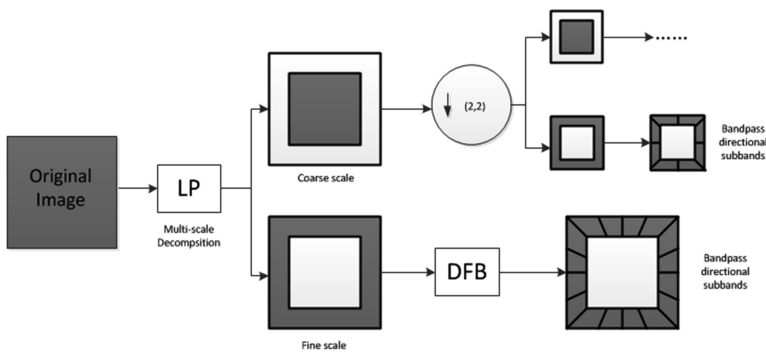


Fig. 2. Contourlet filter bank

2.2 The Low Frequency Sub-Band Feature Extraction

The low-frequency subband is the embodiment of coarse texture feature of image. In this paper, the uniformity of texture is reflected by Angular Second Moment (ASM), Contrast (CON), Correlation (COR) and Entropy (ENT) of the gray level co-occurrence matrix. We extracted these four features of gray level co-occurrence matrix after Contourlet Transform decomposed. Specifically the low frequency image coefficient matrix will be the gray level co-occurrence matrix transformed after the image is decomposed. We calculated Angular Second Moment (ASM), Contrast (CON), Correlation (COR) and Entropy (ENT) in $[0^\circ, 45^\circ, 90^\circ, 135^\circ]$ respectively. In order to reduce the dimension of feature vector and the computational complexity, we get a feature vector in 8 dimensions $f_1 = [a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8]$ through calculating the mean and variance of each parameter in four directions. The calculating formulae of the specific parameters are as following [22, 23].

Angular Second Moment (ASM):

$$ASM = \sum_i \sum_j P(i, j)^2 \quad (1)$$

Contrast (CON):

$$CON = \sum_i \sum_j (i, j)^2 P(i, j) \quad (2)$$

Correlation (COR):

$$COR = [\sum_i \sum_j i \times j \times P(i, j) - \mu_x \mu_y] / \sigma_x \sigma_y \quad (3)$$

Entropy (ENT):

$$ENT = - \sum_i \sum_j P(i, j) \text{lb}[P(i, j)] \quad (4)$$

where $P(i, j)$ is the elements whose coordinate is (i, j) that is in gray level co-occurrence matrix of coefficients of low frequency after the Contourlet transformation. μ_x and σ_x are the mean value and mean variance of $\{P_x(i) | i = 1, 2, \dots, N\}$. μ_y and σ_y are the mean value and mean variance of $\{P_y(i) | i = 1, 2, \dots, N\}$.

For further reflecting the extent of image texture, we composed the eigenvector $f_2 = [\mu, \sigma]$ by extracting the mean and variance of low frequency coefficient matrix after contourlet transformation.

The mean μ and variance σ can be calculated by:

$$\mu = \frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N P(i, j) \quad (5)$$

$$\sigma = \frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N (P(i, j) - \mu)^2 \quad (6)$$

In (5) and (6), $P(i, j)$ is the decomposition coefficient whose coordinate is (i, j) in $M \times N$ low frequency subband coefficient matrix after the contourlet transformation.

2.3 High Frequency Sub-Band Feature Extraction

The high-frequency directional subband of Contourlet Transform contains the image edges and fine texture feature. In this paper, the image is decomposed into 4 layers. From the first layer to the third layer are intermediate frequency band and the fourth layer is high frequency band.

The intermediate frequency band contains part texture information of image. The mean and variance can reflect not only the unevenness of gray image, but also the depth degree of the texture. Considering these factors, the mean and variance of intermediate frequency coefficient matrix are extracted as texture features of intermediate frequency sub-band. The three intermediate frequency sub-bands contain 3, 4 and 8 direction respectively, and we can compose a 30-dimensional feature vector $f_3 = [\mu_1, \mu_2, \dots, \mu_{15}, \sigma_1, \sigma_2, \dots, \sigma_{15}]$ by the mean and variance of sub-band coefficients in these 15 directions.

The energy distribution is sparser at the highest level sub-band. Because energy distribution at different scales and directions can effectively distinguish the texture, the energy of the coefficient matrix is extracted as high frequency characteristics. The high frequency image contains 16 directions after Contourlet Transform and we can extract the energy of sub-band coefficients in these 16 directions to form a 16-dimensional feature vector $f_4 = [b_1, b_2, \dots, b_{16}]$.

The energy can be calculated as following:

$$E = \sum_{i=1}^M \sum_{j=1}^N P(i, j)^2 \quad (7)$$

In (7), $P(i, j)$ is the decomposition coefficient whose coordinate is (i, j) in $M \times N$ high frequency sub-band coefficient matrix after the contourlet transformation.

In addition, the mean and variance of the high-frequency sub-band coefficient matrix can reflect the depth degree of the texture and it is also an important high-frequency image texture features. We extract the mean and variance of high frequency image sub-band decomposed to form a 32-dimensional feature vector $f_5 = [\mu_1, \mu_2, \dots, \mu_{16}, \sigma_1, \sigma_2, \dots, \sigma_{16}]$.

As we have pointed out above, we extracted 5 feature vectors from different frequency band. These feature vectors can fully represent the uniformity, depths of color and energy distribution and other characteristics of image texture. However, if we group the entire feature vector into vector set as the input of identification system, the speed of recognition is bound to be affected due to the dimension of the vector is too

large. It is a pretty important part that how to determine the optimal texture feature representation in Contourlet Transform, which means choosing as little feature vectors as possible to characterize the texture on the condition of ensuring the recognition accuracy. In this paper, we put the extracted feature vectors as input of the identification system, and finally determine the optimal feature vector after repeated recognition.

3 Support Vector Machine (SVM) Classifier

As mentioned above, the recognition process is as Fig. 3 showed. We choose Support Vector Machine (SVM) as classifier. Support vector machine is a machine learning method based on development of statistical learning theory [24]. It has very strong generalization ability and less depends on the quantity and quality of samples. It has the best generalization ability for the classification of unknown samples through constructing the optimal hyperplane. It promotes the problem to be processed in linear form that Support Vector Machine (SVM) maps the data from input space to a high-dimensional feature space by support vector (SV) kernel. As SVM usually tries to minimize a bound on the structural risk but not the empirical risk, it always can get a global minimum value.

Empirical Risk Minimization (ERM) is a formal term for a simple concept: find the function $f(x)$ that minimizes the average risk on the training set. Empirical risk is defined as bellow:

$$R_{emp}(f) = \frac{1}{N} \sum_{i=1}^N C(f(x_i), y_i) \quad (8)$$

where $C(f, y)$ is a suitable cost function, e.g., $C(f, y) = (f(x) - y)^2$.

Minimizing the empirical risk is not a bad thing to do, provided that sufficient training data is available, since the law of large numbers ensures that the empirical risk will asymptotically converge to the expected risk for $n \rightarrow \infty$. However, for small samples, one cannot guarantee that ERM will also minimize the expected risk. This is the all too familiar issue of generalization.

The Vapnik-Chervonenkis dimension (VC dimension) is a measure of the complexity (or capacity) of a class of functions $f(\alpha)$. The VC dimension measures the largest number of examples that can be explained by the family $f(\alpha)$. The basic argument is that high capacity and generalization properties are at odds. If the family $f(\alpha)$ has enough capacity to explain every possible dataset, we should not expect these functions to generalize very well. On the other hand, if functions $f(\alpha)$ have small capacity but they are able to explain our particular dataset, we have stronger reasons to



Fig. 3. Recognition Process

believe that they will also work well on unseen data. The VC dimension is the size of the largest dataset that can be shattered by the set of functions $f(\alpha)$. One may expect that models with a large number of parameters would have high VC dimension, whereas models with few parameters would have low VC dimensions. The VC dimension is a more “sophisticated” measure of model complexity than dimensionality or number of free parameters.

Because the VC dimension provides bounds on the expected risk as a function of the empirical risk and the number of available examples. It can be shown that the following bound holds with probability $1 - \eta$.

$$R(f) \leq R_{emp}(f) + \sqrt{\frac{h(\ln(\frac{2N}{h}) + 1) - \ln(\frac{\eta}{4})}{N}} \quad (9)$$

where h is the VC dimension of $f(\alpha)$, N is the number of training examples, and $N > h$.

Structural Risk Minimization (SRM) is another formal term for an intuitive concept: the optimal model is found by striking a balance between the empirical risk and the VC dimension. SVM achieves SRM by minimizing the following Lagrangian formulation:

$$L_P(\omega, b, \alpha) = \frac{1}{2} \|\omega\|^2 - \sum_{i=1}^N \alpha_i [y_i(\omega^T x_i + b) - 1] \quad (10)$$

where α_i is positive Lagrange multipliers [25, 26].

As the ratio N/h gets larger, the VC confidence becomes smaller and the actual risk becomes closer to the empirical risk. This and other results are part of the field known as Statistical Learning Theory or Vapnik-Chervonenkis Theory, from which Support Vector Machines originated.

4 Experimental Results

To evaluate the effectiveness of the proposed method, we carried out a series experiments on two large and comprehensive texture databases: the Sweden leaves database [27] and the ICL database¹ which is established by the Intelligent Computing Laboratory.

All images in the ICL database are taken by cameras or scanners on a white background paper under vary illumination conditions after the leaves are picked from plants. Two sides of every kind of leaf are respectively taken to images. Images are tended to be in an agreed size and colorful which are taken through this process. To guarantee the background smooth and clear, only one leaf is constructed an image. The ICL database includes 200 species of plants. Each species includes 30 samples of leaf images (15 per side). Hence there are totally 6000 images. Figure 4 shows some

¹ <http://www.intelengine.cn/dataset/index.html>

samples from the ICL database in which the top images are the front side of plant leaves and the bottom images are the opposite side.

The Sweden leaves database contains 15 the monolithic leaves pictures of different Swiss tree and each type has 75 image files. The original image data of Swiss plant leaves contains petiole. This does not have robust feature that the shape of the leaves should have, because the direction and length of petiole deeply depend on the process of the leaf collection when intercept the plant leaf image samples. While leafstalk will provide certain difference information, we can remove some kind of noise to build another data set.

In order to implement the proposed method, we choose 5 sets of data and set the relevant parameters empirically.

4.1 Experimental Results on the ICL Database and the Sweden Leaves Database

The experimental results of the ICL Database are as follows: (Table 1).

Table 1. The classification rates (%) on ICL database

Method	Recognition rate (%)					Equal error rates (EER)
	Set 1	Set 2	Set 3	Set 4	Set 5	
LBP	62.5774	62.4313	58.9223	61.5548	56.7119	0.3845
Curvelet	60.9754	62.7850	65.4688	69.5751	69.6489	0.3323
Contourlet	69.4189	72.2176	77.1574	75.9221	72.7874	0.2754

The experimental results of the Sweden Leaves Database are as follows: (Table 2)

Table 2. The classification rates (%) on Sweden leaves database

Method	Recognition rate (%)					Equal error rates (EER)
	Set 1	Set 2	Set 3	Set 4	Set 5	
LBP	79.0605	72.3183	74.7692	72.4716	72.9713	0.2932
Curvelet	84.2346	82.9483	79.1720	85.5022	76.3445	0.2254
Contourlet	86.3874	85.8156	89.6552	89.9520	83.8635	0.1723

5 Conclusions

In this paper, we studied a hybrid approach based on Contourlet Transform and Support Vector Machine (SVM) Classifiers for plant recognition. By decomposing input images into multi-scale factors which have attractive properties such as shift invariance and computational efficiency, we can extract discriminative features which are not sensitive to the variations of illumination and translation and capture the intrinsic geometry structure of images. By combining the crafted feature with large

margin classifiers (more specifically, SVM), the proposed recognition method has higher experimental performance and can better capture the rich features of natural images such as edges, curves and contours. In the future work, we plan to improve the efficiency of the proposed method, and implement it as recognition software which is suitable for real-word applications.

Acknowledgments. This work was supported by the grants of the National Science Foundation of China, Nos. 61133010, 61373105, 61303111, 61411140249, 61402334, 61472282, 61472280, 61472173, 61373098 and 61272333, China Postdoctoral Science Foundation Grant, Nos. 2014M561513, and partly supported by the National High-Tech R&D Program (863) (2014AA021502 & 2015AA020101), and the grant from the Ph.D. Programs Foundation of Ministry of Education of China (No. 20120072110040), and the grant from the Outstanding Innovative Talent Program Foundation of Henan Province, No. 134200510025.

References

1. Guyer, D., Miles, G., Schreiber, M., Mitchell, O., Vanderbilt, V.: Machine vision and image processing for plant identification. *Trans. ASAE* **29**, 1500–1507 (1986)
2. Wang, X.-F., Huang, D.-S., Xu, H.: An efficient local Chan-Vese model for image segmentation. *Pattern Recognit.* **43**, 603–618 (2010)
3. Huang, D.-S.: *Systematic Theory of Neural Networks for Pattern Recognition*, vol. 28, pp. 323–332. Publishing House of Electronic Industry of China, Beijing (1996)
4. Yu, H.-J., Huang, D.-S.: Normalized feature vectors: a novel alignment-free sequence comparison method based on the numbers of adjacent amino acids. *IEEE/ACM Trans. Comput. Biol. Bioinform. (TCBB)* **10**, 457–467 (2013)
5. Huang, D.-S., Jiang, W.: A general CPL-AdS methodology for fixing dynamic parameters in dual environments. *IEEE Trans. Syst. Man Cybern. Part B Cybern.* **42**, 1489–1500 (2012)
6. Huang, D.-S., Du, J.-X.: A constructive hybrid structure optimization methodology for radial basis probabilistic neural networks. *IEEE Trans. Neural Netw.* **19**, 2099–2115 (2008)
7. Wang, X.-F., Huang, D.-S.: A novel density-based clustering framework by using level set method. *IEEE Trans. Knowl. Data Eng.* **21**, 1515–1531 (2009)
8. Shang, L., Huang, D.-S., Du, J.-X., Zheng, C.-H.: Palmprint recognition using FastICA algorithm and radial basis probabilistic neural network. *Neurocomputing* **69**, 1782–1786 (2006)
9. Zhao, Z.-Q., Huang, D.-S., Sun, B.-Y.: Human face recognition based on multi-features using neural networks committee. *Pattern Recogn. Lett.* **25**, 1351–1358 (2004)
10. Huang, D., Ip, H., Chi, Z.: A neural root finder of polynomials based on root moments. *Neural Comput.* **16**, 1721–1762 (2004)
11. Huang, D.-S.: A constructive approach for finding arbitrary roots of polynomials by neural networks. *IEEE Trans. Neural Netw.* **15**, 477–491 (2004)
12. Huang, D.-S.: Radial basis probabilistic neural networks: model and application. *Int. J. Pattern Recognit Artif Intell.* **13**, 1083–1101 (1999)
13. Xiangbin, Z.: Texture classification based on contourlet and support vector machines. In: 2009 ISECS International Colloquium on Computing, Communication, Control, and Management, pp. 521–524 (2009)

14. Liu, Z., Fan, X., Lv, F.: SAR image segmentation using contourlet and support vector machine. In: Fifth International Conference on Natural Computation, 2009. ICNC 2009, pp. 250–254. IEEE (2009)
15. Wang, J., Ge, Y.: Texture feature recognition based on contourlet transform and support vector machine. *Jisuanji Yingyong J. Comput. Appl.* **33**, 677–679 (2013)
16. Haralick, R.M., Shanmugam, K., Dinstein, I.H.: Textural features for image classification. *IEEE Trans. Syst. Man Cybern.* **3**(6), 610–621 (1973)
17. Vapnik, V.: *The nature of statistical learning theory*. Springer Science & Business Media, New York (2000)
18. Burges, C.J.: A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Disc.* **2**, 121–167 (1998)
19. Schölkopf, B., Burges, C.J., Smola, A.J.: *Advances in kernel methods: support vector learning*. MIT press, Massachusetts (1999)
20. Soderkvist, O.J.O.: *Computer Vision Classification of Leaves from Swedish Trees* (2001)