Fan-Beam Binarization Difference Projection (FB-BDP): A Novel Local Object Descriptor for Fine-Grained Leaf Image Retrieval

Xin Chen[1], Bin Wang[2,1,*], Yongsheng Gao[1]
[1]Griffith University, Brisbane, Australia
[2]Nanjing University of Finance and Economics, Nanjing, China
xin.chen12@griffithuni.edu.au, wangbin@nufe.edu.cn, yongsheng.gao@griffith.edu.au

Abstract

Fine-grained leaf image retrieval (FGLIR) aims to search similar leaf images in subspecies level which involves very high interclass visual similarity and accordingly poses great challenges to leaf image description. In this study, we introduce a new concept, named fan-beam binarization difference projection (FB-BDP) to address this challenging issue. It is designed based on the theory of fan-beam projection (FBP) which is a mathematical tool originally used for computed tomographic reconstruction of objects and has the merits of capturing the inner structure information of objects in multiple directions and excellent ability to suppress image noise. However, few studies have been made to apply FBP to the description of texture patterns. Rather than calculating ray integrals over the whole object area, FB-BDP restricts its ray integrals calculated over local patches to guarantee the locality of the extracted features. By binarizing the intensity-differences between the off-center and center rays, FB-BDP enable its ray integrals insensitive to illumination change and more discriminative in the characterization of texture patterns. In additional, due to inheriting the merits of FBP, the proposed FB-BDP is superior over the existing local image descriptors by its invariance to scaling transformation, robustness to noise, and strong ability to capture direction and structure texture patterns. The results of extensive experiments on FGLIR show its higher retrieval accuracy over the benchmark methods, promising generalization power and strong complementarity to deep features.

1. Introduction

Leaf image patterns have been extensively researched and considerable leaf image descriptors [12, 15, 21, 28, 34, 35] have achieved encouraging performance on plant species recognition in the computer vision and image processing community. Recently, increasing concerns have been paid on whether leaf image patterns are also informative enough for plant cultivar identification which is essential for crops introduction, genetic improvement, breeders' intellectual property protection, and early identification of seedlings in modern agriculture systems [14]. As shown in Figure 1, leaf image patterns have great difference across coarse-grained categories (species), while have high similarity across fine-grained categories (cultivars).

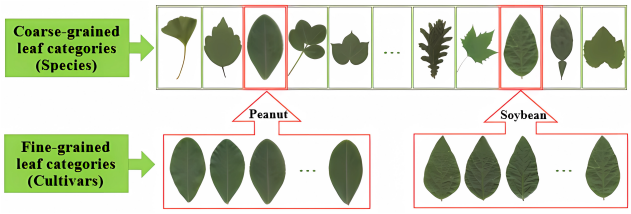

Figure 1. Coarse-grained (species) versus fine-grained (cultivars) interclass variation in leaf image patterns.

With the quick growth of new cultivars developed by human experts in modern agriculture, there is a growing need to develop effective fine-grained leaf image retrieval (FGLIR) systems for managing the existing cultivars and breeding new ones. In the content-based image retrieval (CBIR) research community, it is commonly known that characterizing image patterns to yield quantized feature representation are essential to the construction of a CBIR system. In this study, we focus on the description of leaf image patterns for addressing the challenging FGLIR issue.

In the past decades, many feature descriptors have been proposed for depicting leaf image patterns. Most of them are designed for coarse-grained (species level) leaf image identification. Considering the large varieties in leaf shape patterns across different species, many efforts have been made on the extraction of leaf shape features such as the leaf boundary curvatures [5, 8, 15, 21], the relative spatial distribution of leaf contour points [12, 19, 34], the geometrical

*Corresponding author

structure patterns [28, 35], etc. and achieved state-of-the-art performance on distinguishing plant species.

Another significant clue for leaf image pattern contributed to species recognition is its texture feature. Chaki et al. [7] used Gabor filter and gray level co-occurrence matrix (GLCM) to model leaf texture patterns for leaf species classification. Tang et al. [32] combined a non-overlap window LBP and GLCM for green tea leaf classification. Inspired by the idea that spatial co-occurrences among features have potential for boosting the discriminative power of descriptors, Qi et al. [27] developed a novel descriptor, named Pairwise Rotation Invariant Co-Occurrence LBP (PRICoLBP), which can not only effectively capture the spatial context co-occurrence information, but also be invariant to rotation. This method reported high classification accuracy on plant species recognition.

Besides handcrafted features, there are several recent studies applying deep learning to leaf image classification. Meet et al. [29] proposed a dual-path CNN to jointly learn complementary shape and texture features in each path and optimize them for leaf classification. Hu et al. [11] proposed a multiscale fusion CNN for plant leaf recognition. Grinblat et al. [10] applied CNN to the extraction of leaf vein features for classifying three different legume species. Tan et al. [38] applied three CNN models, pretrained AlexNet, fine-tuned AlexNet and D-leaf, to leaf vein feature extraction followed by several machine learning techniques for classification.

The aforementioned leaf image descriptors are all designed for leaf species identification. Recently, there are also several attempts focusing on FGLIR issue. Oleander (Nerium oleander L.) has many cultivars exhibiting high inter-class similarity. Baldi et al. [2] developed a back-propagation neural network for identifying oleander cultivars. Tavakoli et al. [33] employed CNN to distinguish 12 cultivars of common beans. Wang et al. [37] is the first to use joint leaf image patterns by combining the leaf features from the lower, middle and upper parts of plants towards more accurate soybean cultivar recognition. More recently, Local RsCoM [36] and SBT [9] both paid attention on the extraction of co-occurrence texture features of leaf images and achieved state-of-the-art performance on fine-grained leaf image identification. Despite the research progress of leaf image description on FGLIR issue, the reported performances are still unsatisfactory for the requirements of high recognition accuracy in modern agriculture which thus leaves us a large room for further research.

Local image descriptors have proven to be the leading model for many computer vision and image processing applications due to their excellent behaviors in the case of illumination change, translation, and occlusion [3]. Local binary pattern (LBP) [24] and its variants [4, 27, 39] are popular local image descriptors that have received wide applications. They work in a circular neighborhood structure and use the binarized grayscale difference information between the center pixel and its neighboring pixels to generate local image features. Although they can capture the spatial structure information of texture patterns and are insensitive to any monotonic transformation of the gray scale, they are not invariant to scaling transformation and sensitive to noise. In additional, the circular structure of LBP also hinder the characterization of directional texture patterns.

Fan-beam projection (FBP) is a mathematical tool originally used for computed tomographic (CT) reconstruction of objects [1]. It can capture the inner structure information of objects from multiple directions and has excellent ability to suppress image noise. However, few studies have been done to apply it to the description of texture patterns. This study is the first to introduce FBP to object description. To make FBP better adapt to the construction of local image descriptors, we introduce a new concept, named fan-beam binarization difference projection (FB-BDP). Rather than calculating ray integrals over the whole object image area, the proposed FB-BDP restricts the ray integrals over local patches to guarantee the locality of the extracted features. By binarizing the intensity-differences between the off-center rays and the center ray, the resulting ray integrals encode discriminative texture information and are insensitive to illumination change. In additional, due to inheriting the merits of FBP, the proposed FB-BDP outperforms the existing local image descriptors by its invariance to scaling transformation, robustness to noise, and strong ability to capture direction and structure texture patterns. Extensive experiments demonstrate its higher retrieval accuracy on FGLIR over state-of-the-arts, promising generalization power and strong complementarity with deep features to boost retrieval performance on FGLIR.

2. FB-BDP Local Descriptor for Leaf Image Representation

In this section, we begin with the definition of fan-shaped local patch (FLP) and use it to introduce a new concept of fan-beam binarization difference projection (FB-BDP) for the construction of local leaf descriptor. The collection of FB-BDP vectors is then used to learn a leaf codebook for quantifying the FB-BDP vectors. At last, we count the quantified FB-BDP vectors of a leaf image into histograms as the final image-level representation.

2.1. Fan-Shaped Local Patch

An image plane can be regarded as a complex plane with each pixel being denoted as complex number $\chi \in \mathbb{C}$ whose real and imaginary parts correspond to its horizontal and vertical coordinate, respectively. For a leaf image, its grayscale version can be mathematically expressed as a 2D function $f(\chi) : \mathbb{C} \rightarrow N$ describing the intensity value of

the pixel χ. Its contour Ω can be represented as an arch-length parameterization form $z(\tau) : [0, 1] \rightarrow \mathbb{C}$ with τ being the normalized arc length from the starting point of the contour in anticlockwise direction.

Given a point $z(\tau) \in \Omega$, let $R(\tau) = \max_{\tau' \in [0,1]} \parallel z(\tau') - z(\tau) \parallel_2$ be the maximum Euclidean distance between it and the other contour points. For a positive number $\rho \in (0, R(\tau))$, we take it as radius and $z(\tau)$ as center to draw a circle. Since the contour Ω is closed and $\rho < R(\tau)$, at least two intersection points can be obtained. We choose two of them (denoted by $z(\tau_l)$ and $z(\tau_r)$) that are nearest to the point $z(\tau)$ along the contour in clockwise and anticlockwise directions, respectively. By rotating the chord $\overline{z(\tau)z(\tau_r)}$ counterclockwise about the point $z(\tau)$ to coincide with the chord $\overline{z(\tau)z(\tau_l)}$, we obtain a fan-shaped local region (denoted by $FLP_{\tau,\rho}$). The rotation angle $\alpha_{\tau,\rho}$ is its opening angle defined as

$$\alpha_{\tau,\rho} = arg((z(\tau_l) - z(\tau))/(z(\tau_r) - z(\tau))), \quad (1)$$

where $arg(\cdot)$ returns the argument of a complex number in the interval $[0, 2\pi)$.

The patch $FLP_{\tau,\rho}$ has two parameters τ and ρ. Varying the former can change the position of the $FLP_{\tau,\rho}$. While the parameter $\rho \in (0, R(\tau))$ is related to the scale of the $FLP_{\tau,\rho}$. To facilitate the subsequent multiscale feature extraction, we generate multiscale $FLP_{\tau,\rho}$ by letting $\rho = R(\tau)/2^w, w = 1, \ldots, W$, where w is the index of scale level and W is the prespecified number of scale levels. Figure 2 (a) presents an example to show $FLP_{\tau,\rho}$ of different scale levels radiated from the same contour point. While Figure 2 (b) presents an example to show $FLP_{\tau,\rho}$ radiated from different contour points the same scale level.

2.2. FB-BDP Local Descriptor

In the following, we use the above defined fan-shaped local patch $FLP_{\tau,\rho}$ to introduce a new concept: fan-beam binarization-difference projection (FB-BDP), for the construction of local descriptor.

Fan Beam: For a fan-shaped local patch $FLP_{\tau,\rho}$, we treat its vertex $z(\tau)$ as source point to emit a ray to bisect its opening angle $\alpha_{\tau,\rho}$. We term it center ray and restrict it end at the point

$$p_0 = z(\tau) + (z(\tau_l) - z(\tau))e^{j \cdot \alpha_{\tau,\rho}/2}, \quad (2)$$

where $j = \sqrt{-1}$. From the above equation, we can see that p_0 is the midpoint of the circular arc between the points $z(\tau_l)$ and $z(\tau_r)$ of the $FLP_{\tau,\rho}$. Here, we denote the center ray as $\overrightarrow{p_0}$. By rotating the ray $\overrightarrow{p_0}$ about its source point $z(\tau)$ by the angles $\theta_u = u \cdot \alpha_{\tau,\rho}/(2U), u = \pm 1, \pm 2, \ldots, \pm U$, we obtain U pairs of rays whose end points can be calculated by

$$p_u = z(\tau) + (p_0 - z(\tau))e^{j \cdot \theta_u}, \quad (3)$$

where U is the parameter to prespecify the number of the generated rays. We name them off center rays and denote them by $\overrightarrow{p_u}$. Collecting the center ray $\overrightarrow{p_0}$ and all the off-center rays $\overrightarrow{p_u}$, we obtain a fan beam $FB_{\tau,\rho}$ generated from the $FLP_{\tau,\rho}$.

$$FB^U_{\tau,\rho} = \{\overrightarrow{p_0}, \overrightarrow{p_u}, u = \pm 1, \pm 2, \ldots, \pm U\}. \quad (4)$$

In Figure 3, the leftmost subfigure shows a fan beam of $FB^2_{\tau,\rho}$ having one center rays and two pairs of off-center rays.

Binarization-difference projection: We first use a function

$$q_0(l) = z(\tau) + (p_0 - z(\tau))(l/\rho), \quad (5)$$

to denote each point of the center ray $\overrightarrow{p_0}$, where $l \in [0, \rho]$ is the length between it and the source point $z(\tau)$ along the ray $\overrightarrow{p_0}$. Likewise, we define a function $q_u(l)$ to represent each point of the ray $\overrightarrow{p_u}$. Then for each off-center ray $\overrightarrow{p_u}$ in the fan beam $FB^U_{\tau,\rho}$, we define a binarization-difference integral over it as

$$B_u = \frac{1}{\rho} \int_0^\rho s(f(q_u(l)) - f(q_0(l)))dl. \quad (6)$$

$$s(y) = \begin{cases} 1, y >= 0 \\ 0, y < 0 \end{cases}. \quad (7)$$

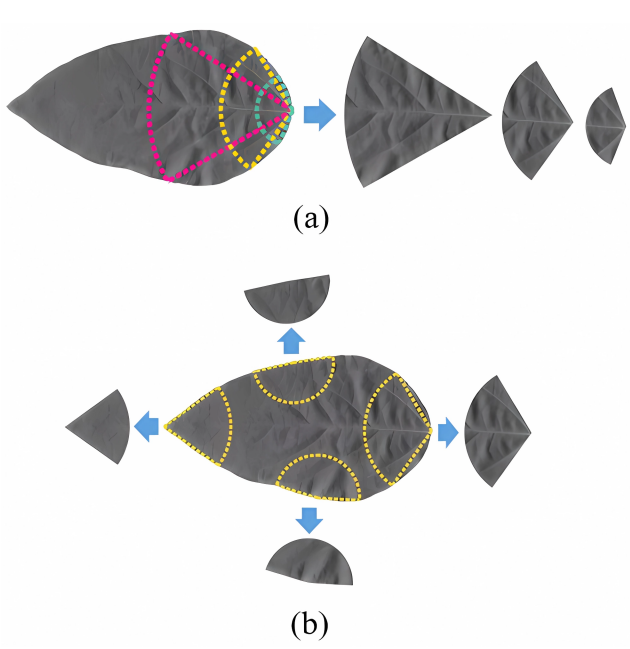

Figure 2. Examples to show the fan-shaped local patches $FLP_{\tau,\rho}$: (a) $FLP_{\tau,\rho}$ of different scale levels radiated from the same contour point. (b) $FLP_{\tau,\rho}$ radiated from different contour points at the same scale level.

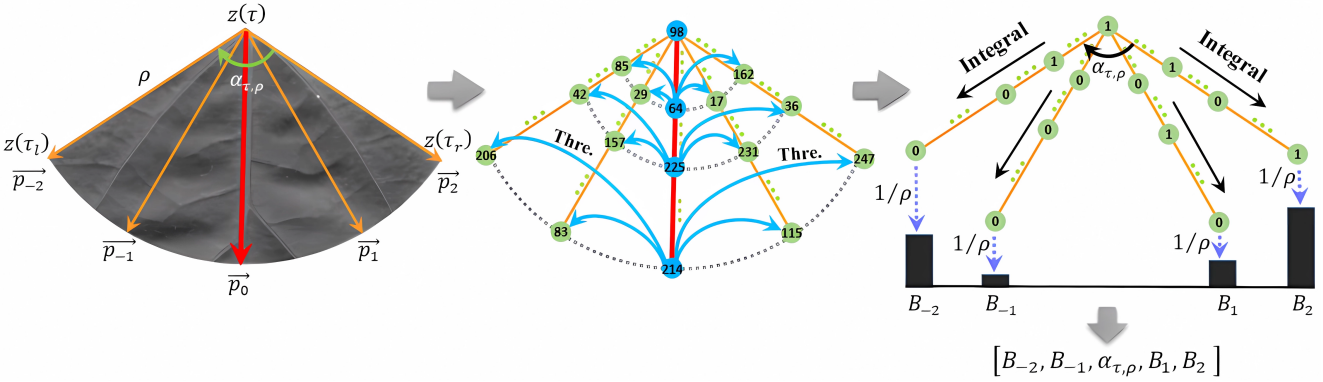

Figure 3. Illustration of the concept of our proposed fan-beam binarization-difference projection (FB-BDP) local descriptor. Left: A group of rays that are emitted from the same source point $z(\tau)$ and equally divide the fan-shaped local patch $FLP_{\tau,\rho}$ form a fan beam $FB_{\tau,\rho}$, where the ray colored by red is center ray and the remaining ones are off-center rays. Middle: The gray values of the pixels on the center ray $\overrightarrow{p_0}$ are used as threshold values to binarize the ones of the corresponding pixels on the off-center rays. Right: A texture feature measure B_u is derived by the conduction of an integral operation over each off-center ray $\overrightarrow{p_u}$ of binarized pixel values followed a scale normalization by the radius ρ. All the texture feature measures $B_u, u = 1, \ldots, U$ and the opening angle $\alpha_{\tau,\rho}$ are collected to form a local descriptor for depicting the fan-shaped local patch $FLP_{\tau,\rho}$.

The measure B_u is derived by first using the gray value of each pixel $q_0(l)$ of the center ray $\overrightarrow{p_0}$ as the threshold value to binarize the one of the corresponding pixel $q_u(l)$ of the off-center ray $\overrightarrow{p_u}$ and conducting integral operation against the binarized off-center ray $\overrightarrow{p_u}$ (See the two rightmost subfigures of Figure 3 for better understanding this concept). It encodes the local intensity differences between the off-center ray and the center ray.

The binarization-difference integral B_u has the following characteristics which enable it very suitable for texture feature depiction: (1) Since it is derived from the fan-shaped local patch $FLP_{\tau,\rho}$, it has the property of locality; (2) It is a geometrical invariant measure because the translation and rotation of the patch $FLP_{\tau,\rho}$ do not change its relative position in the patch $FLP_{\tau,\rho}$ and the normalization by the radius ρ in Eq. 6 makes it invariant to the scaling of the patch $FLP_{\tau,\rho}$; (3) Like the widely used local binary pattern (LBP), it is also a gray-scale invariant measure, i.e., insensitive to monotonic illumination changes due to the use of relative intensities instead of the exact intensities with respect to the center ray; (4) It is robust to noise because of the use of integral operation. In additional, by changing the length of the radius ρ of the patch $FLP_{\tau,\rho}$, multiscale measures B_u can be achieved.

FB-BDP local descriptor: After calculating the binarization-difference integral B_u for each off-center ray $\overrightarrow{p_u}$ in the fan beam $FB_{\tau,\rho}$, we use them to construct a local descriptor. Considering that the angle measure $\alpha_{\tau,\rho}$ characterizes the geometric property of the patch $FLP_{\tau,\rho}$ and the measure B_u depicts the texture character of the patch $FLP_{\tau,\rho}$, we combine them to construct a feature vector $V_{\tau,\rho}$ for making them complementary to contribute to the description of the patch $FLP_{\tau,\rho}$. To make it easier to handle the mirror transform of the object image, we organize the measures $\alpha_{\tau,\rho}$ and B_{θ_u} in the vector $V_{\tau,\rho}$ as follows:

$$V_{\tau,\rho} = [B_{\theta_u}, u = -U, \ldots, -1] \cup [\alpha_{\tau,\rho}] \cup [B_{\theta_u}, u = 1, \ldots, U] \tag{8}$$

We can see that the mirror transform of object image only makes the vector $V_{\tau,\rho}$ flipped which will be further addressed in the subsequent process of quantization of feature vector. We term the vector $V_{\tau,\rho}$ as fan-beam binarization-difference projection (FB-BDP) local descriptor.

2.3. Aggregating FB-BDP Local Descriptors into Histogram Representation

We have presented a FB-BDP local descriptor to describe the local patch $FLP_{\tau,\rho}$. In this subsection, we employ it to construct image-level representation. As discussed in Section 2.1, the parameters τ, ρ control the position and scale of the patch $FLP_{\tau,\rho}$, respectively. By uniformly sampling T values in the range $[0, 1]$ of the parameter τ, we can obtain T points, $z(\tau_i), i = 1, \ldots, T$, sampled from the contour Ω, where T is the parameter to prespecify the number of the sample points. Varying the index i of the contour point from 1 to T makes the local descriptor $V_{\tau_i,\rho}$ characterize the leaf image in multiple positions. For each contour point $z(\tau_i)$, we make the parameter ρ take W values from the range $(0, R(\tau_i)) : \rho_w = R(\tau_i)/2^w, w = 1, \ldots, W$. Letting the index w of scale level vary from 1 to W makes the local descriptor V_{τ_i,ρ_w} describe the leaf image from coarse to fine. However, directly using the set of the FB-BDP local descriptors to describe leaf image is not compact and is also very expensive for subsequent feature matching. In this subsection, we apply the widely used Bag-of-words model

[30] to aggregate them into a histogram representation for a compact description and efficient matching.

Codebook learning: For each scale level w, we collect all the FB-BDP local descriptors $V_{\tau_i,\rho_w}, i = 1,\ldots,T$, of the dataset images to build a training set of size $M \cdot T$ for learning a codebook, where M is the number of images in the dataset. We randomly select K FB-BDP local descriptors at the scale level w as codewords to construct an initial codebook $CB^w = v_k^w, k = 1,\ldots,K$, where K is the parameter to specify the number of the codewords in the codebook. We conduct the common K-means clustering algorithm to update each codeword in the CB^w and the algorithm is not terminated until the distortion error is less than 9×10^{-3}, where the distortion error is defined as the maximum of the absolute differences between the elements of the current version and last version of each codeword. Recall that the mirror transformation of a leaf image only makes the FB-BDP $V_{\tau,\rho}$ flipped. To make the encoded local descriptors invariant to mirror transformation, during the training stage, we define the distance between a training FB-BDP local descriptor V and a codeword v as

$$d(V,v) = min\left\{\|V-v\|_2, \left\|V^{'}-v\right\|_2\right\}, \tag{9}$$

where $V^{'}$ is the flipped version of V.

Vector quantization: We now have a total of W codebooks, $CB^w, w = 1,\ldots,W$ (one codebook per scale level). Then for each FB-BDP local descriptor $V_{i,w}$ of a leaf image we use the corresponding codebook CB^w to quantify it as

$$c_{i,w} = \underset{k=1,\ldots,K}{argmin}\,(d(V_{\tau_i,\rho_w}, v_k^w)). \tag{10}$$

For a leaf image, we count the codes $c_{i,w}$ of its FB-BDP vectors $V_{\tau_i,\rho_w}, i = 1,\ldots,T$ of the scale level w into a $K-$dimensional histogram H_w with each element defined as

$$h_w(k) = \sum_{i=0}^{T} \delta(c_{i,w} - k). \tag{11}$$

where $\delta(\cdot)$ is a binary function that returns 1 for an input of zero and 0, otherwise. Concatenating the histograms $H_w, w = 1,\ldots,W$, we obtain a vector of dimension $K \cdot W$ as the final representation for the leaf image.

3. Experiments

In this section, we first apply the proposed method to two challenging fine-grained leaf image retrieval (FGLIR) tasks: soybean and peanut leaf image retrievals, respectively for validating its effectiveness. Another group of experiments of cross-dataset test, i.e., applying the leaf codebook learned from one leaf dataset to the FGLIR on another leaf dataset, are then conducted to examine the generalization ability of our method. Finally, we perform additional group of experiments of fusing our FB-BDP image representation with state-of-the-art deep features to examine their complementarity on FGLIR.

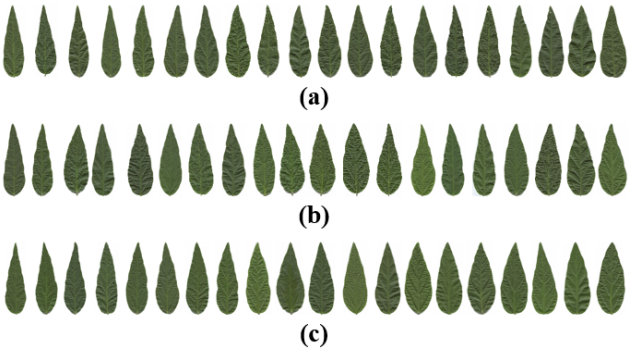

Figure 4. Parts of soybean leaf images of different cultivars from the three subsets of the SoyCultivar200 dataset [37]. (a) Soy-Up, (b) Soy-Mid, (c) Soy-Low.

3.1. Soybean Leaf Image Retrieval

Soybeans are commonly known as one of the most important economic crops and have many cultivars in the world. Several previous studies [16, 17, 33, 37] have made attempts to use leaf image patterns as clues for soybean cultivar identification. SoyCultivar200[1] [37] is a publicly available soybean leaf image dataset that has been recently used for evaluating the performance of an algorithm for cultivar recognition. It has 200 cultivars with each having 30 leaves collected from different parts of soybean plants: 10 samples from the upper part, 10 samples from the middle part, and 10 samples from the lower part. The SoyCultivar200 dataset is designed to test the performances of single leaf image pattern and joint leaf image pattern for cultivar recognition, respectively. For the former, all the samples from the same part of soybean plants are grouped into a subset which makes the dataset divided into three subsets, named as Soy-Up, Soy-Mid, and Soy-Low, respectively. Each subset accordingly consists of 200×10=2000 single leaf image patterns. While for the latter, all the 6000 leaves in the dataset are divided into 2000 groups with each containing three leaves of the same cultivar that are from the upper part, middle part and lower part of different soybean plants, respectively. Each group is treated as a joint leaf pattern and all of them form a set, named Soy-Joint, consisting of 200×10=2000 joint leaf patterns. Figure 4(a)-(c) show parts of leaf images of different cultivars from the three subsets, Soy-Up, Soy-Mid, and Soy-Low, respectively. More details about the SoyCultivar200 dataset refer to [37].

Two standard evaluation metrics, Bulls-eye test [12, 19,

[1] https://github.com/NUFE-AIAG/FGLIR

34] and precision-recall (PR) curves [5, 6, 9] are used to quantify the retrieval performance of the competing algorithms. Nine state-of-the-art image descriptors, PRICoLBP [27], LETRIST [31], Spatial Pyramid [18], Local RsCoM [36], SBT [9], DSFH [20], HeW-ResNet50 [25], ReSW [26], and siaMAC+ReSW [26], are used as baselines. Among them, PRICoLBP and LETRIST are local binary patterns designed for texture description. Spatial Pyramid is bag of SIFT descriptor for texture characterization. Local RsCoM and SBT both focus on extracting co-occurrence texture features for fine-grained leaf image identification. While DSFH, HeW-ResNet50, ReSW, and siaMAC+ReSW are all recently published deep feature representations particularly designed for image retrieval. The parameters for the proposed method are empirically set to $T = 600$, $W = 10$, $U = 10$ and $K = 400$. The parameters of the other competing methods follow their original setting.

The Bull-eye scores of all the competing methods are summarized in Table 1. It can be seen that on the use of single leaf image patterns (performing on the test cases, Soy-Up, Soy-Mid and Soy-Low, respectively), the proposed methods achieves the scores of 49.12%, 51.76% and 50.59%, respectively which are separately 1.05%, 2.63% and 2.67% higher than the other competing methods. While using joint leaf image patterns (Soy-Joint), the proposed method achieves an exciting score of 84.20% which outperforms the other competing methods by 2.51%. We also plot the PR curves for all the competing methods on the four test cases in Figure 5. It can be clearly observed that on all the test cases, the proposed method consistently achieves the best PR curves.

Algorithm	Up	Mid	Low	Joint
PRICoLBP	35.27	34.86	32.47	56.22
LETRIST	34.97	35.51	34.70	57.89
Spatial Pyramid	32.74	33.99	33.68	60.40
Local RsCoM	42.77	43.11	41.60	70.44
SBT	47.57	48.40	47.92	81.69
DSFH	41.52	41.82	40.35	69.90
ReSW	45.12	46.33	43.78	80.97
siaMAC+ReSW	47.32	48.35	44.25	78.58
HeW-ResNet50	48.07	49.13	45.23	80.42
Proposed	**49.12**	**51.76**	**50.59**	**84.20**

Table 1. The Bull-eye scores (%) of various methods on the four test cases, Soy-Up, Soy-Mid, Soy-Low and Soy-Joint of the SoyCultivar200 leaf image dataset [37].

3.2. Peanut Leaf Image Retrieval

Peanuts are another economic crop in the world which are widely grown in the tropics and subtropics, contributing to both small and large commercial producers. To examine the effectiveness of the proposed method for peanut leaf image retrieval, we collect 600 leaf images from 120 peanut cultivars with each consisting of 5 samples. Different from the SoyCultivar200, this dataset is designed only for testing the performance of single leaf image pattern. We name it PeanCultivar120[2]. Figure 6 shows an example sample for each peanut cultivar. It can be observed that they have very high inter-class similarity which makes it very difficult even for human experts to distinguish them. The same evaluation metric, benchmark methods and algorithm parameters as the former groups of experiments are used in this test.

The Bull-eye scores achieved by the proposed methods and the other competing methods are summarized in Table 2. As can be seen that the proposed method obtains the score of 51.80% which outperformers the other competing methods by 2.53%. In Figure 7, we plot the precision-recall curves of all the competing methods. As can be seen that the PR curve achieved by our method is obviously better than those of the other methods.

Algorithm	Bull-eye scores (%)
PRICoLBP	41.27
LETRIST	46.57
Spatial Pyramid	44.10
Local RsCoM	45.23
SBT	46.83
DSFH	37.10
ReSW	44.73
siaMAC+ReSW	41.03
HeW-ResNet50	49.27
Proposed	**51.80**

Table 2. The Bull-eye scores (%) of various methods on the peanut cultivar leaf image dataset.

3.3. Cross-Dataset Test for Generalization Capability Examination

Generally speaking, for a specific FGLIR task, e.g. soybean leaf image retrieval, if the used leaf codebooks are learned from other species, e.g. peanut or tree, the retrieval performance would be normally degraded in some extent due to the domain difference between distinct species. However, since they are all learned from leaf image patterns, the performance degradation should not be large. In

[2]https://github.com/NUFE-AIAG/ICCV2023-FB-BDP

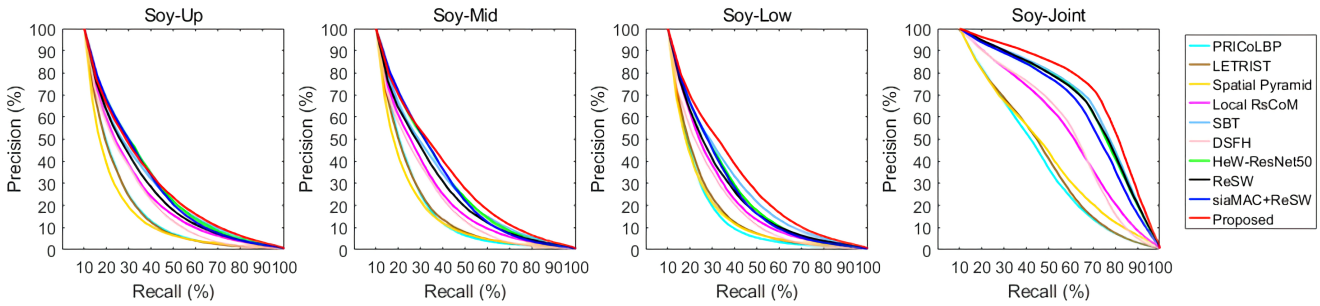

Figure 5. The precision-recall curves for all the ten competing methods on the four test cases, Soy-Up, Soy-Mid, Soy-Low and Soy-Joint of the SoyCultivar200 leaf image dataset[37].

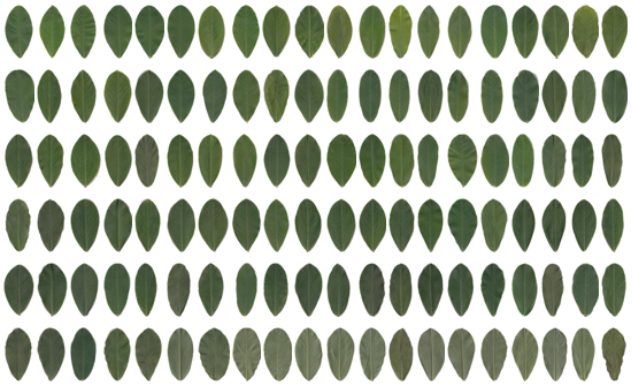

Figure 6. 120 example leaf images of different peanut cultivars from our collected PeanCultivar120 dataset.

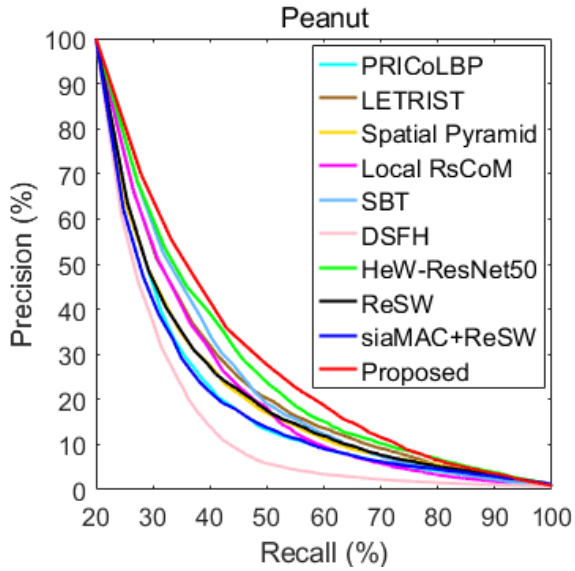

Figure 7. The precision-recall curves of all the competing methods on the PeanCultivar120 leaf image dataset.

Leaf image dataset used for codebook learning	FGLIR Task			
	Soy-Low	Soy-Mid	Soy-Up	PeanCultivar120
Soy-Low	50.59	51.40	48.51	48.73
Soy-Mid	50.38	51.76	48.40	48.93
Soy-Up	50.60	51.32	49.12	49.33
PeanCultivar120	47.23	47.63	45.64	51.80
MEW2012	49.10	48.88	46.53	48.97

Table 3. The Bull-eye scores (%) of cross-dataset test using the proposed method.

this subsection, we conduct a group of experiments of cross-dataset test to examine the generalization capability of the proposed FB-BDP image representation.

In our experiments, we use the following five leaf image datasets, Soy-Low, Soy-Mid, Soy-Up, PeanCultivar120 and MEW2012 [23], to learn codebooks, respectively. Among them, the first three are all soybean leaf image datasets as introduced in Section 3.1. The fourth dataset is a peanut leaf image dataset as introduced in Section 3.2. The last one is a publicly available leaf species image dataset. It consists of 9745 leaf images of 153 tree and shrubs species with at least 50 samples per species. For more details about MEW2012, refer to [23]. Since each species in MEW2012 has different number of samples, we randomly select 50 samples from each species to construct a dataset of 153×50=7650 samples for codebook leaning. Figure 8 shows parts of typical leaf images of different species in the MEW2012 dataset.

Based on the proposed FB-BDP image representation, we separately use the above mentioned five leaf image datasets to learn codebooks which are then applied to FGLIR tasks on the four datasets, Soy-Low, Soy-Mid, Soy-Up and PeanCultivar120, respectively. The retrieval scores are summarized in Table 3. By observing each column, we can find that for three soybean FGLIR tasks, Soy-Low Soy-

Figure 8. Parts of leaf images of different species from the MEW2012 leaf image dataset[23].

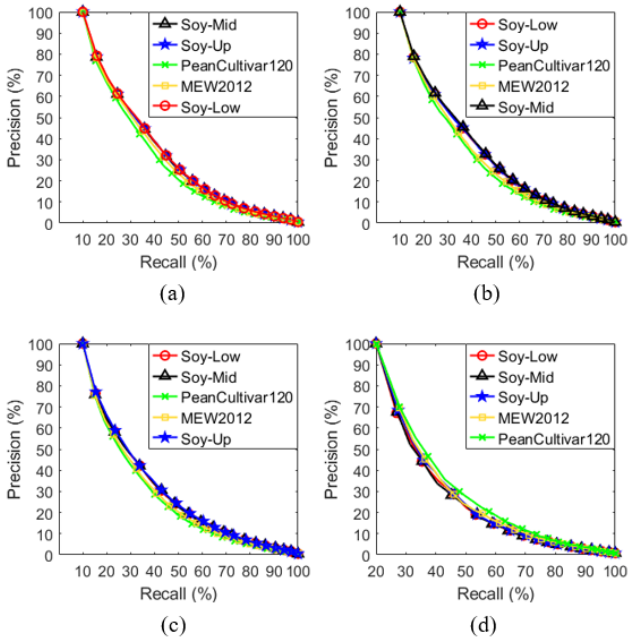

Figure 9. The PR curves of applying codebooks learned from Soy-Low, Soy-Mid, Soy-Up, PeanCultivar120 and MEW2012 leaf image datasets to the four FGLIR tasks: (a) Soy-Low, (b) Soy-Mid, (c) Soy-Up, and (d) PeanCultivar120.

Mid and Soy-Up, the codebooks learned from the different soybean leaf image datasets achieve very small decrease in retrieval scores (less than 0.72%). While using the codebooks learned from peanut or tree and shrubs species, although they have relatively large domain differences with the soybean plant, we also achieve an acceptable decrease in retrieval scores (less than 4.13%). While for the peanut FGLIR task, we achieve a decrease of less than 3.07% in retrieval score when using the codebooks learned from the soybean leaf images or tree and shrubs leaf images. In Figure 9 (a)-(d), we plot the PR curves of using five different codebooks on the four FGLIR tasks, Soy-Low, Soy-Mid, Soy-Up, PeanCultivar120, respectively. As can be seen that for each FGLIR task, the PR curves achieved from five different leaf codebooks are close to each other. These experimental results consistently indicate the promising generalization capability of the proposed method.

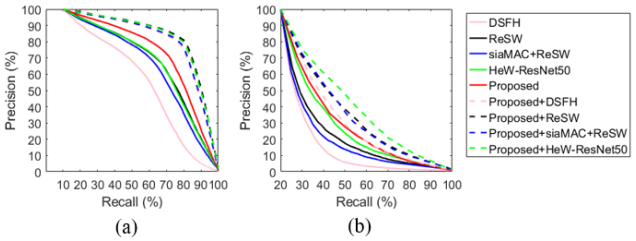

Figure 10. The PR curves achieved by the proposed FB-BDP, four deep feature representations, DSFH, ReSW, siaMAC+ReSW, HeW-ResNet50, and their fusions on the (a) Soy-Joint FGLIR task and (b) PeanCultivar120 FGLIR task.

3.4. Complementarity with Deep Features

Recently, there are increasing interests of fusing traditional handcrafted image descriptors with deep features to boost the accuracies of various image classification and retrieval tasks [9, 13, 22]. In this subsection, for examining the complementarity of our FB-BDP image representation with deep features, we conduct an additional group of experiments of applying the fusion of them to soybean and peanut FGLIR tasks. Since developing a novel fusion method is not our focus in this study, we directly use the recent feature fusion method, KNN-HDFF [9], which is a simple yet effective method of fusing deep features with handcrafted image descriptors. We separately fuse our FB-BDP image representation with four deep features, DSFH [20], HeW-ResNet50 [25], ReSW [26], and siaMAC+ReSW [26] for two FGLIR tasks, Soy-Joint and PeanCultivar120. The resulting PR curves are shown in Figure 10 (a) and (b), respectively. As can be observed that the fusions of our FB-BDP image representation with deep features greatly improve the retrieval accuracies of using their individual features which indicates the high complementarity of our method with deep features.

4. Conclusion

We have presented a novel method, named fan-beam binarization difference projection (FB-BDP), to construct leaf image representation for addressing the challenging FGLIR issue. It is designed based on the theory of fan-beam projection (FBP) which is a mathematical tool originally used for computed tomographic reconstruction of objects. The leaf image representation derived from the proposed FB-BDP has the following merits: (1) Locality in feature measures; (2) Invariance to geometrical transformations including scaling, translation, rotation and mirror; (3) Invariance to illumination change; (4) Robustness to noise and (5) Strong ability to capture direction and structure texture patterns which enable it more suitable for FGLIR tasks than the existing local image descriptors. Through using the

Bag-of-words model to aggregate the FB-BDP local feature descriptors into histograms, we obtain a compact image-level representation for efficient feature matching. Extensive experiments on two challenging FGLIR tasks, soybean and peanut leaf image retrievals, together with comprehensive evaluations including retrieval accuracy, generalization ability, complementarity to deep features prove its outstanding performance over state-of-the-art methods on FGLIR.

Acknowledgement

This work was supported in part by the National Natural Science Foundation of China under Grant No. 61372158 and the Natural Science Foundation of Jiangsu Province of China under Grant No. BK20221345.

References

[1] Amir Averbuch, Ilya Sedelnikov, and Yoel Shkolnisky. Ct reconstruction from parallel and fan-beam projections by a 2-d discrete radon transform. *IEEE Transactions on Image Processing*, 21(2):733–741, 2011. 2

[2] Ada Baldi, Camilla Pandolfi, Stefano Mancuso, and Anna Lenzi. A leaf-based back propagation neural network for oleander (nerium oleander l.) cultivar identification. *Computers and Electronics in Agriculture*, 142:515–520, 2017. 2

[3] Debotosh Bhattacharjee and Hiranmoy Roy. Pattern of local gravitational force (plgf): A novel local image descriptor. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(2):595–606, 2021. 2

[4] Xiuli Bi, Yuan Yuan, Bin Xiao, Weisheng Li, and Xinbo Gao. 2d-lcolbp: A learning two-dimensional co-occurrence local binary pattern for image recognition. *IEEE Transactions on Image Processing*, 30:7228–7240, 2021. 2

[5] Yongsheng Gao Bin Wang, Douglas Brown and John LaSalle. March: Multiscale-arch-height description for mobile retrieval of leaf images. *Information Sciences*, 302:132–148, 2015. 1, 6

[6] Darshan Bryner, Anuj Srivastava, and Eric Klassen. Affine-invariant, elastic shape analysis of planar contours. *In 2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 390–397, 2012. 6

[7] Jyotismita Chaki, Ranjan Parekh, and Samar Bhattacharya. Plant leaf recognition using texture and shape features with neural classifiers. *Pattern Recognition Letters*, 58:61–68, 2015. 2

[8] Xin Chen and Bin Wang. Invariant leaf image recognition with histogram of gaussian convolution vectors. *Computers and Electronics in Agriculture*, 178(105714), 2020. 1

[9] Xin Chen, Bin Wang, and Yongsheng Gao. Symmetric binary tree based co-occurrence texture pattern mining for fine-grained plant leaf image retrieval. *Pattern Recognition*, 129(108769), 2022. 2, 6, 8

[10] Guillermo L. Grinblat, Lucas C. Uzal, Mónica G. Larese, and Pablo M. Granitto. Deep learning for plant identification using vein morphological patterns. *Computers and Electronics in Agriculture*, 127:418–424, 2016. 2

[11] Jing Hu, Zhibo Chen, Meng Yang, Rongguo Zhang, and Yaji Cui. A multiscale fusion convolutional neural network for plant leaf recognition. *IEEE Signal Processing Letters*, 25(6):853–857, 2018. 2

[12] Rongxiang Hu, Wei Jia, Haibin Ling, and Deshuang Huang. Multiscale distance matrix for fast plant leaf recognition. *IEEE Transactions on Image Processing*, 21(11):4667–4672, 2012. 1, 6

[13] Shichao Kan, Yigang Cen, Zhihai He, Zhi Zhang, Linna Zhang, and Yanhong Wang. Supervised deep feature embedding with handcrafted feature. *IEEE Transactions on Image Processing*, 28(12):5809–5823, 2019. 8

[14] Nicholas Kibet Korir, Jian Han, Lingfei Shangguan, Chen Wang, Emrul Kayesh, Yanyi Zhang, and Jinggui Fang. Plant variety and cultivar identification: advances and prospects. *Critical reviews in biotechnology*, 33(2):111–125, 2013. 1

[15] Neeraj Kumar, Peter N. Belhumeur, Arijit Biswas, David W. Jacobs, W. John Kress, Ida C. Lopez, and João V. B. Soares. Leafsnap: A computer vision system for automatic plant species identification. *In European conference on computer vision*, pages 502–516, 2012. 1

[16] Mónica G. Larese, Ariel E. Bayá, Roque M. Craviotto, Miriam R. Arango, Carina Gallo, and Pablo M. Granitto. Multiscale recognition of legume varieties based on leaf venation images. *Expert Systems with Applications*, 41(10):4638–4647, 2014. 5

[17] Mónica G. Larese and Pablo M. Granitto. Finding local leaf vein patterns for legume characterization. *Machine Vision and Application*, 27:709–720, 2016. 5

[18] Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. *In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2169–2178, 2006. 6

[19] Haibin Ling and David W. Jacobs. Shape classification using the inner-distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(2):286–299, 2007. 1, 6

[20] Guang-Hai Liu and Jing-Yu Yang. Deep-seated features histogram: A novel image retrieval method. *Pattern Recognition*, 116(107926), 2021. 6, 8

[21] Farzin Mokhtarian and Sadegh Abbasi. Matching shapes with self-intersections:application to leaf classification. *IEEE Transactions on Image Processing*, 13(5):653–661, 2004. 1

[22] Loris Nanni, Stefano Ghidoni, and Sheryl Brahnam. Handcrafted vs. no-handcrafted features for computer vision classification. *Pattern Recognition*, 71:158–172, 2017. 8

[23] Petr Novotný and Tomáš Suk. Leaf recognition of woody species in central europe. *Biosystems Engineering*, 115(4):444—452, 2013. 7, 8

[24] Timo Ojala, Matti Pietikainen, and Topi Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002. 2

[25] Shanmin Pang, Jin Ma, Jianru Xue, Jihua Zhu, and Vicente Ordonez. Deep feature aggregation and image re-ranking with heat diffusion for image retrieval. *IEEE Transactions on Multimedia*, 21(6):1513–1523, 2019. 6, 8

[26] Shanmin Pang, Jihua Zhu, Jiaxing Wang, Vicente Ordonez, and Jianru Xue. Building discriminative cnn image representations for object retrieval using the replicator equation. *Pattern Recognition*, 83:150–160, 2018. 6, 8

[27] Xianbiao Qi, Rong Xiao, Chun-Guang Li, Yu Qiao, Jun Guo, and Xiaoou Tang. Pairwise rotation invariant co-occurrence local binary pattern. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(11):2199–2213, 2014. 2, 6

[28] Asma Rejeb Sfar, Nozha Boujemaa, and Donald Geman. Confidence sets for fine-grained categorization and plant species identification. *International Journal of Computer Vision*, 111(3):255–275, 2015. 1, 2

[29] Meet P. Shah, Sougata Singha, and Suyash P. Awate. Leaf classification using marginalized shape context and shape+ texture dual-path deep convolutional neural network. *In Proceedings of the IEEE International Conference on Image Processing*, pages 860–864, 2017. 2

[30] Josef Sivic and Andrew Zisserman. Video google: A text retrieval approach to object matching in videos. *In Proceedings of the 9th IEEE International Conference on Computer Vision*, pages 1470–1477, 2003. 5

[31] Tiecheng Song, Hongliang Li, Fanman Meng, Qingbo Wu, and Jianfei Cai. Letrist: Locally encoded transform feature histogram for rotation-invariant texture classification. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(7):1565–1579, 2018. 6

[32] Zhe Tang, Yuancheng Su, Meng Joo Er, Fang Qi, Li Zhang, and Jianyong Zhou. A local binary pattern based texture descriptors for classification of tea leaves. *Neurocomputing*, 168:1011–1023, 2015. 2

[33] H. Tavakoli, P. Alirezazadeh, A. Hedayatipour, A.H. Banijamali Nasib, and N. Landwehr. Leaf image-based classification of some common bean cultivars using discriminative convolutional neural networks. *Computers and Electronics in Agriculture*, 181(105935), 2021. 2, 5

[34] Bin Wang and Yongsheng Gao. Hierarchical string cuts: a translation, rotation, scale and mirror invariant descriptor for fast shape retrieval. *IEEE Transactions on Image Processing*, 23(9):4101–4111, 2014. 1, 6

[35] Bin Wang, Yongsheng Gao, Changming Sun, Michael Blumenstein, and John La Salle. Chord bunch walks for recognizing naturally self-overlapped and compound leaves. *IEEE Transactions on Image Processing*, 28(12):5963–5976, 2019. 1, 2

[36] Bin Wang, Yongsheng Gao, Xiaohui Yuan, and Shengwu Xiong. Local r-symmetry co-occurrence: Characterising leaf image patterns for identifying cultivars. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 19(2):1018–1031, 2022. 2, 6

[37] Bin Wang, Yongsheng Gao, Xiaohui Yuan, Shengwu Xiong, and Xianzhong Feng. From species to cultivar: Soybean cultivar recognition using joint leaf image patterns by multiscale sliding chord matching. *Biosystems Engineering*, 194:99–111, 2020. 2, 5, 6, 7

[38] Jing wei Tan, Siow-Wee Chang, Sameem Abdul-Kareem, Hwa Jen Yap, and Kien-Thai Yong. Deep learning for plant species classification using leaf vein morphometric. *IEEE/AMC Transcations on Computational Biology and Bioinformatics*, 17(1):82–90, 2020. 2

[39] Baochang Zhang, Yongsheng Gao, Sanqiang Zhao, and Jianzhuang Liu. Local derivative pattern versus local binary pattern: face recognition with high-order local pattern descriptor. *IEEE Transactions on Image Processing*, 19(2):533–544, 2010. 2