Pattern Recognition Letters, 32(15), 2011, pp2036-2046

Automatic Citrus Canker Detection from Leaf Images Captured in Field

Min Zhang[a], Qinggang Meng[b,*]

[a]College of Computer Science, ChongQing University, China.
[b]Department of Computer Science, Loughborough University,UK.

Abstract

Citrus canker, a bacterial disease of citrus tree leaves, causes significant damage to citrus production worldwide. Effective and fast disease detection methods must be undertaken to minimize the losses of citrus canker infection. In this paper, we present a new approach based on global features and zone-based local features to detect citrus canker from leaf images collected in field which is more difficult than the leaf images captured in labs. Firstly, an improved AdaBoost algorithm is used to select the most significant features of citrus lesions for the segmentation of the lesions from their background. Then a canker lesion descriptor is proposed which combines both color and local texture distribution of canker lesion zones suggested by plant phytopathologists. A two-level hierarchical detection structure is developed to identify canker lesions. Thirdly, we evaluate the proposed method and its comparison with other approaches, and the experimental results show that the proposed approach achieves similar classification accuracy as human experts.

Keywords:
Citrus canker detection, Zone-based texture distribution, Classification, Hierarchical detection, Feature learning, Hue-intensity-saturation.

1. Introduction

Citrus canker is a disease which gets worldwide concern as its potentially hazardous threat to citriculture. This disease can affect all types of citrus

*Corresponding author
Email addresses: zhangmin122@hotmail.com (Min Zhang), q.meng@lboro.ac.uk (Qinggang Meng)

crops, including oranges, sour oranges, grapefruit, tangerines, lemons and limes and presently it occurs in over thirty countries in Asia, Pacific and Indian Ocean islands, South America, Middle East and USA (Polek, 2007).

This disease is caused by the bacterium Xanthomonas axonopodis pv citri (Xac) (Vernière et al., 2003). The infection of citrus canker results in defoliation, die-back, premature leaf and fruit drop and at last the trees will produce no fruits at all. Citrus canker is highly contagious and can be spread rapidly by wind, rain, landscaping equipment, people work in field, moving infected or exposed plants or plant parts. Moreover, citrus canker is difficult to eradicate. Once it is introduced into an area, elimination of inoculum by removal and destruction of infected and exposed trees is the most accepted practice to quarantine the disease and stop further spread (Gottwald et al., 2001; Gottwald and Timmer, 1995). For example, U.S. Department of Agriculture established a regulation – the "1900-ft rule". The regulation requires the removal and destruction of diseased citrus trees and of all citrus trees within a 1900-ft radius. In United States, over 12 million US dollars per year are dedicated to citrus canker control program.

At present, there is no effective method to eradicate citrus canker, and the basic strategy is to reduce the effect of infection and to prevent the spread. Detecting citrus canker at the early stage is the key to control this disease. So far different technologies have been used to identify citrus canker, such as plant physiology, biochemistry, serological techniques, molecular biology and detection methods based on information technology (Gambley et al., 2009; Golmohammadi et al., 2007).

The most accurate methods of citrus canker identification are serological techniques, and molecular biology (for examples, enzyme-linked immunosorbent assay, protein profiles as determined by electrophoretic techniques and DNA analysis methods) (Park and Young, 2006; Park et al., 2006). These methods have to be carried out in laboratory and some of them are costly and time consuming, and they are mainly used by quarantine bureaus to confirm the disease.

The widely used method to identify canker in field is by plant phytopathologists' visual observation of each suspicious tree (Gottwald et al., 2002; Das, 2003). It is based on the fact that citrus canker is mainly a leaf-spotting disease. Leaf lesions become visible about 7 to 10 days after infection. As the lesions age, they change appearance in different phases, and they are easy to be confused with other citrus diseases, such as citrus scab disease. Identification of citrus canker needs experienced experts, otherwise the misjudgment can lose the best opportunity to prevent the spread of the disease. The lack of experts in this area limits the timely and wide identification of the disease.

As information technologies have been applied in more and more fields, new methods are now being investigated to identify citrus disease.

- **Fluorescence spectroscopy:** In Brazil, scientists proposed methods to detect citrus canker in citrus plants using laser induced fluorescence spectroscopy. They developed a new optical technique to detect citrus canker with a portable field spectrometer unit and showed that the laser induced fluroscence spectroscopy had the potential to be applied to citrus plan (Belasque et al., 2008).
- **Hyperspectral imaging:** hyperspectral imaging approach was developed by (Qin et al., 2009; Lins et al., 2009) to detect canker lesions on citrus fruits. They used spectral information divergence classification methods to detect the disease and obtained good classification results.
- **Machine vision technology:** Pydipati (Pydipati et al., 2006) used machine vision technology to identify the citrus canker on citrus leaves. All the sample leaves were preprocessed and their images were captured by an imaging station under the same angle and light. HIS color space and spatial gray-level dependency matrices were used to generate color texture features, then SAS stastical analysis were conducted to reduce feature set and classify four kind of citrus leaves, which are greasy spot, melanose, scab and normal citrus leaves. Dae (Dae et al., 2009) also used the similar methods to detect grapefruit peel diseases.

One limitation of the existing image-based citrus canker detection methods is that they are all based on images collected in a highly controlled environment under specific conditions. However in real world, it is often the planters who first find the symptom of disease in field. In comparison with the other two methods mentioned above, machine vision technology has advantages in detection citrus canker in field. It needs no specific equipments or chemical reagents, and images are easy to capture by digital cameras, mobile phones or other equipments and can be transferred by internet.

The objective of this paper is to present an approach based on computer vision to detecting citrus canker. The detection is based on citrus leaf images collected in field which is more difficult and challenge than those captured in labs. The main contributions of this paper are summarized as follows:

- Deal with citrus canker detection from real citrus leaf images captured in field rather than from labs.

- An improved AdaBoost algorithm was developed to segment citrus lesions from background.

- The whole leaf images were divided into several zones. Then the local features of each zone (distribution of color and texture information) were extracted and assembled to generate a citrus canker descriptor.

- A hierarchical and staged detection scheme was formulated to identify citrus canker based on images collected under various natural conditions.

- Several machine learning methods were investigated to construct the classifier and tested on real-world data. Furthermore, the proposed approach was also compared with human experts in this area to demonstrate the feasibility of machine vision and pattern recognition technology in citrus canker detection.

The rest of the paper is structured as follows. Section 2 proposes the hierarchical citrus canker detection method. Section 3 describes the citrus canker lesion descriptor. In this section, LBPH (Local Binary Pattern on Hue) features and the combined local feature are presented. Section 4 reports the experimental results. Finally section 5 concludes the paper.

2. Hierarchical Citrus Canker Detection

To detect citrus canker from the images collected in field is more difficult than the images captured in labs, one of the key reasons is because the background is sometimes similar to the specific part of a canker lesion. To deal with this problem, a hierarchical citrus canker detection algorithm is presented. Figure 1 illustrates this detection process including the global matching stage, and the local feature extraction and canker detection stage. The global matching stage aims to find suspicious citrus disease lesion areas from background and the canker detection stage is to identify canker lesions from other citrus disease lesions.

Due to the variety of canker lesions, in the global matching stage, we have to find all the possible areas and some of them may be other disease

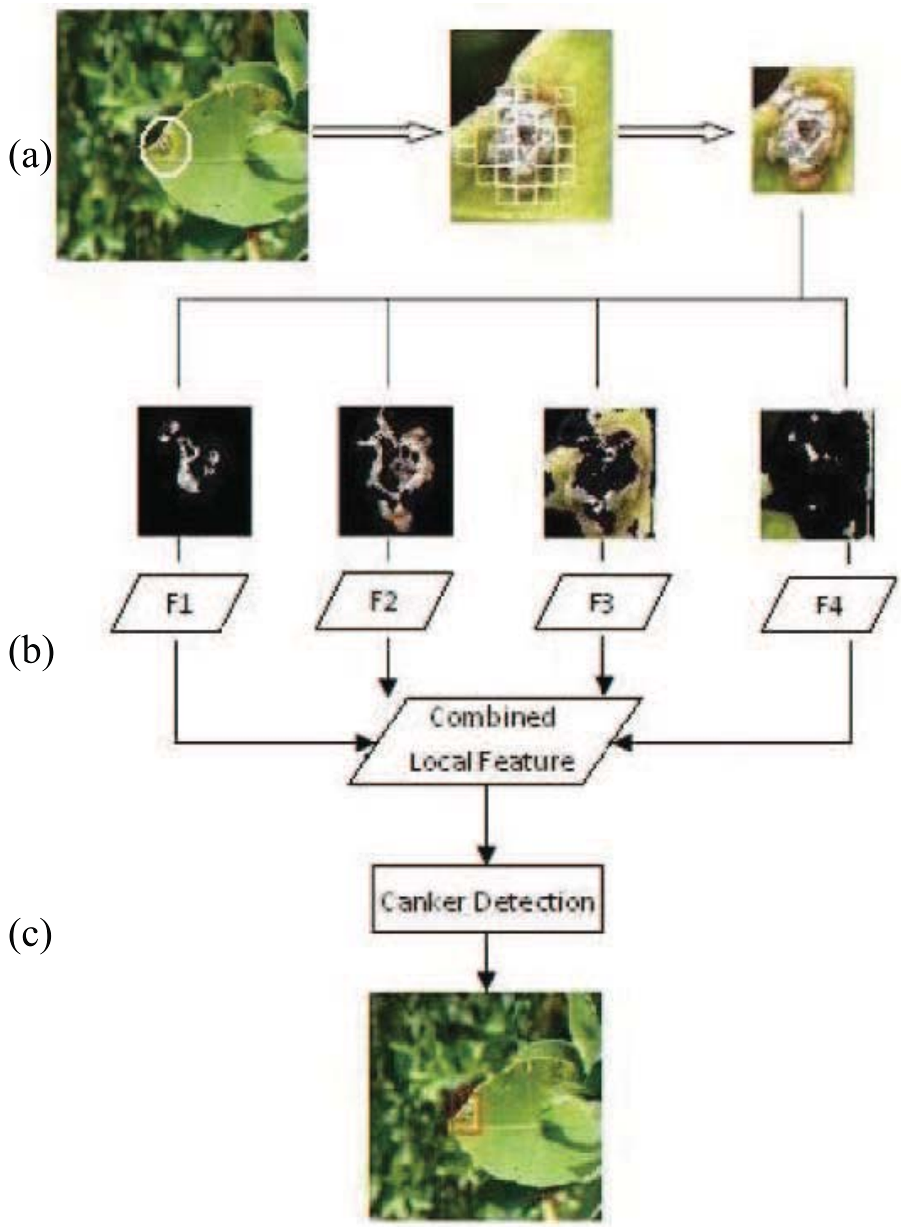


Figure 1: Hierarchical citrus canker detection.(a) global matching based on window union approach; (b) feature extraction based on zones; (c) canker detection and output. F1, F2, F3 and F4 are local feature vectors

infected lesions. To avoid missing the canker lesions and to search quickly, in this phase we use a bottom-up method: window union algorithm as shown in algorithm 1, for lesion area searching. Firstly the image is searched in a small window size and classified by classifier C_1 which was used for fast judging whether a small area is a part of any kind of disease lesion. Then the detected small windows are merged to form bigger areas. Finally the merged areas are judged by the classifier named C_2 which was trained with larger-size image samples than samples used by classifier C_1. Classifier C_1 and classifier C_2 use the same training method, but work on different window sizes. After the classification of C_2, the possible citrus lesion areas were located on the image. Figure 2 shows the procedure of global matching.

Then the merged area was quantized into four zones to extract the combined local features for canker detection. The whole set of citrus canker images was classified into six types by a clustering algorithm according to lesion color distribution. In the phase of canker detection, each of the six classifiers is trained on its corresponding type of citrus canker lesions (as shown in figure 3) and other disease (not citrus canker disease) lesion sample set.

The features used in this training and detection are the combined local features, which will be discussed in section 3.2. If the lesion is judged as any type of canker lesion described above, it is classified to be canker infected.

In our approach, a SceBoost algorithm was used to train the above threshold classifiers, the detailed description of SceBoost algorithm is in section 3.1. Our strategy is to include other disease samples we collected in negative sample set and take each type of canker lesion samples as positive sample set for the corresponding classifier. Then the obtained training sets are used to construct the six individual type canker classifiers.

3. Citrus Canker Lesion Descriptor

Citrus canker lesions' appearance can be described by phytopathologists as follows (Polek, 2007; Gottwald et al., 2002; Das, 2003): Leaf lesions develop first on the lower surface as tiny, slightly raised, blister like spots; At first they are circular in shape, then may become irregular; As the lesions age, they become tan or brown with water-soaked raised margins usually surrounded by a chlorotic or yellow halo or ring; At last the lesions change to be corky or spongy and the centers may become crater-like, old lesions may fall out, creating a shot-hole effect; Lesions' sizes depend on the cultivar

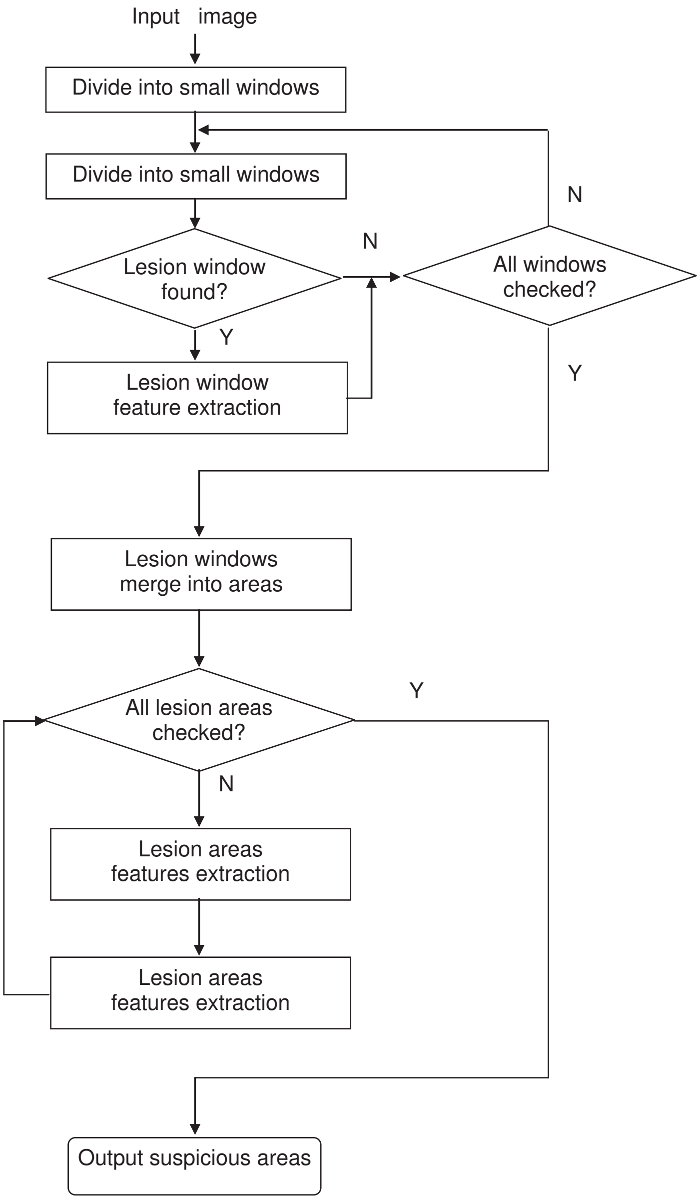


Figure 2: Flowchart of global matching stage

Algorithm 1 Window union algorithm for lesion area detection

Input:
The image, I;
The classifier of small size samples, C_1;
The classifier of area size samples, C_2;
The set of lesion windows, $Q = \emptyset$;
The set of merged windows, $P = \emptyset$;
The set of lesion area, $R = \emptyset$;
The threshold of merged area Th:
Output:
The set of merged lesion areas, R;

```
preprocess image I;
divide I into small windows W_ij which are in the same size, I = sum_{i=1}^{m} sum_{j=1}^{n} (W_ij),
  W_ij ∩ W_pq = ∅, if i ≠ p and j ≠ q, m is the number of lesion windows at
  vertical direction and n at horizontal direction;
for each W_ij, i = 1 ··· m, j = 1 ··· n, do
    extract features of W_ij;
    classify W_ij using classifier C_1;
    if W_ij is classified to be lesion, then
        add W_ij to Q;
    end if
end for
for each window Q_i, Q_i ∈ Q, do
    traverse every element in P,
    if Q_i is adjacent to any area in P, then
        if area P_k is adjacent to Q_i, then
            add Q_i to P_k and update P_k;
        else
            add Q_i to P as new element;
        end if
    end if
end for
traverse every element in P, if the size of P_k ⩾ Th, add P_k to R;
return R;
```

Where line 2 reads: divide I into small windows W_{ij} which are in the same size, $I = \sum_{i=1}^{m}\sum_{j=1}^{n}(W_{ij})$, $W_{ij} \cap W_{pq} = \emptyset$, if $i \neq p$ and $j \neq q$, m is the number of lesion windows at vertical direction and n at horizontal direction;

(a) type 1 (b) type 2 (c) type 3 (d) type 4 (e) type 5 (f) type 6

Figure 3: Examples of six types of citrus canker lesions

and the age of the host tissue at the time of infection. From the description we can find that the lesions vary in shape, size and color by the kind of citrus cultivar and the infection time. Rule-based citrus canker description was infeasible as it is hard to translate all the phytopathologist knowledge into digital image feature patterns. Instead, in this paper, machine learning algorithms were investigated to select the most significant features of citrus canker lesions. Two-level features are proposed to describe citrus canker lesions: the first level features named global features are extracted for detecting citrus lesion areas from the image background; and the second level features (named combined local features) are constructed from the lesion areas which are detected by global features to further identify canker lesions from other confusable citrus diseases lesions. The global lesion feature extraction is detailed in section 3.1 and followed by the description of combined local features in section 3.2.

3.1. Boosted Global Feature Selection

This first stage of citrus lesion detection from an image collected in field is to separate lesion areas from background. Figure 4 shows some examples of citrus canker images: image in 4(a) is collected in lab and others are collected in field. From figure 4 we can find that it is much more difficult to detect canker lesions from images collected in field than from those captured in lab: the background often includes grasses, citrus leaves and soil, and some of these objects are similar with canker lesions to some degree.

Because of the complexity of background and the fact that canker lesions have various appearances, it is hard to decide what features are the most distinguished ones to represent canker lesions. Several image process methods have been used to extract features from canker lesions and background, including each component's mean, standard deviation, variance and

(a) (b) (c) (d)

Figure 4: comparison of images captured in field and in lab. (a) Image captured in lab; (b)(c)(d) images captured in field

correlation coefficient in RGB color space and HIS color space; FFT texture features, Gabor features and gray level co-occurrence matrix, gray level difference features; the edge amount calculated by Prewitt operators, Canny operators and Sobel operators (Zhang, 2008).

Boosting algorithm (Freund, 1995; Xiao et al., 2003; Li and Zhang, 2004) is a statistical method and the motivation of this method is to integrate the results of a set of weak classifiers sequentially and vote them to form a more efficient and strong classifier using a weighted voting scheme. It was firstly proposed in (Kearns and Valiant, 1989), and (Freund and Schapire, 1997) presented Adaboost algorithm which has become a representative boosting algorithm.

In this study, our previously developed Adaboost algorithm, SceBoost, is used to select the most significant features and for constructing classifiers in algorithm 1. The selected features are combined into a global feature vector, which is tested to be efficient in detecting lesion areas from complicated natural background. we improve the original AdaBoost algorithm by using both adaptive symmetric cross entropy threshold and classification error to select a weak classifier at each range. The weak classifiers in our algorithm are linear classifiers using perception approach (Zhang et al., 2007). We can define the symmetric cross entropy of two weak classifiers h_i and h_j as:

$$SCE(h_i : h_j) = \sum_{k=1}^{N} |h_i^k - h_j^k| \cdot (\frac{{w_i}^k}{{w_j}^k})^{{w_i}^k} \cdot (\frac{{w_j}^k}{{w_i}^k})^{{w_j}^k} \quad (1)$$

Where h_i^k is the classification result of example X_k by weak classifier h_i, and w_i^k is the weight given to example X_k after the weak classifier h_i has been selected, N is the number of samples. $SCE(h_i : h_j)$ represents the information difference between h_i and h_j. For two class problem $h_j^k \in \{-1, 1\}$, we can use the weights to indicate the information of these random variables' distribution. If h_i^k was not equal to h_j^k, $SCE(h_i : h_j)$ can indicate the different amount of information carried by the two weak classifiers. The $SCE(h_i : h_j)$ value is large with big difference between h_i^k and h_j^k, and vice versa.

To determine whether a weak classifier h_i is redundant or not we can calculate $S(h_i)$ as:

$$S(h_i) = \max_t SCE(h_i : h_t); t = 1, 2, .., T \tag{2}$$

Where $h_1, h_2, ..., h_T$ are weak classifiers that have been selected at training round T. Before h_i is selected as the weak classifier for training round $T+1$, $S(h_i)$ will be compared with a threshold ATS. If value of $S(h_i)$ is less than ATS, then h_i is deleted from the candidate list. The value of ATS may change during learning period, if we can not find a weak classifier that the value $S(h_i)$ is less than ATS, then ATS is adjusted according to equation 3:

$$ATS = ATS * C; 0 < C < 1 \tag{3}$$

Where C is a coefficient which is selected based on experimental results (with different C). It can affect the search granularity and the computing time. The SceBoost algorithm is illustrated in algorithm 2, and more details can be found in (Zhang et al., 2007).

3.2. Local Canker Lesion Feature Description

To distinguish a citrus canker from other leaf diseases cannot be achieved easily by global features of the whole image only. As shown in figure 5, other disease lesions may have the similar shape or color or texture as canker lesions. Detailed information is needed for further identification. From the observations of phytopathologists it can be seen that the canker lesion may be divided into several specific zones. The combination of all zones and the fusion of different features of each zone can describe the subtle differences between canker lesions and lesions caused by other citrus diseases. A combined local feature descriptor is proposed in this research based on each zone's features.

Algorithm 2 Algorithm SceBoost-part 1/2

0. Input:

Training examples $E = (x_1, y_1), .., (x_N, y_N)$
The maximum number $Mmax$ of weak classifiers to be selected
The initial value of adaptive threshold ATS
The feature vector $F = (f_1, ..., fm)$;
The candidate classifiers set Ch;

1. Initialization:

$w_i = 1/N$; $H = \phi$; $h_0 = 0$;

2. Iteration:

for $t = 1, 2, ..., T$ **do**

(1)Using w_t to produce sample weights distribution D_t on E

$$D_t = \frac{w_t}{\sum_{i=1}^{N} w_i} \tag{4}$$

(2)On each feature vector $f_j, j = 1..m$, fit the weak classifiers $h_{j,t}$ on D_t ;
(3)Ch=($h_{j,t}, j = 1..m$)
(4)For $h_{j,t}, j = 1..m$, calculate classification error:

$$\varepsilon_i = \sum_i w_t^{(i)} |h_{j,t}(x_i) - y_i| \tag{5}$$

(5)
while Ch is nonempty **do**
Choose $h_{j,t}$ with lowest ε_j from the candidate classifiers
Calculate :

$$S(h_{j,t}) = \max_k SCE(h_{j,t} : h_k); k = 1, 2, .., t-1 \tag{6}$$

if $S(h_{j,t}) < ATS$ **then**
The classifier $h_{j,t}$ is selected, $h_t = h_{j,t}$, $\varepsilon_t = \varepsilon_j$
Goto (8)
else
Remove $h_{j,t}$ from Ch
end if
end while
(6) Adjust ATS according to Eq.(3)
(7) Goto (5)

Algorithm 2 Algorithm SceBoost-part 2/2

(8) Calculate :

$$\beta_t = \frac{1}{2}\ln(\frac{1-\varepsilon_t}{\varepsilon_t}) \tag{7}$$

(9)Update weights:

$$w_{t+1}(i) = w_t^i \beta_t^{1-|h_t(x_i)-y_i|} \tag{8}$$

end for

3. Return the strong hypothesis:

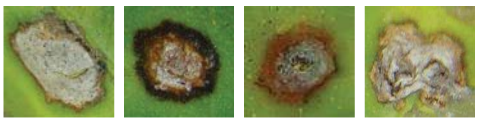
$$H = sign(\sum_{t=1}^{T} \beta_t h_t(x)),\ \textit{sign} \text{ is a signum function.} \tag{9}$$

3.2.1. Local Binary Patterns

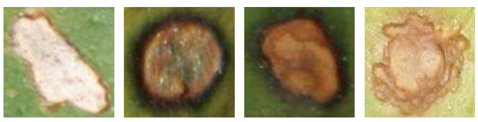
Local Binary Pattern (LBP) is a gray-scale texture description which was originally introduced by Ojala et al. (Ojala et al., 1996). The LBP operator defines a texture T for a central pixel in a local neighborhood area of radius R, which is sampled at P points:

$$T = t(g_c, g_0, ..., g_{P-1}) \tag{10}$$

where, g_c corresponds to the gray value of the central pixel, g_p is the value of its pth neighbor. The neighborhood is thresholded by the value of the



(a) samples of citrus canker lesions



(b) Samples of other citrus disease lesions

Figure 5: Citrus canker and other diseases lesions

central pixel and the thresholded pixels in the neighborhood are multiplied by a corresponding binomial coefficient weight. LBP is a unique P-bit pattern code by multiplying binomial coefficient 2^p with each S(g_p - g_c):

$$LBP_{P,R} = \sum_{p=0}^{P-1} S(g_p - g_c)2^p \tag{11}$$

where:

$$S(x) = \begin{cases} 1 & if \quad x \geq 0 \\ 0 & if \quad x < 0 \end{cases}$$

By definition, LBP describes the spatial structure of the local texture. However, LBP is normally derived from gray images, color texture images need to be transformed into gray images before calculating the LBP, therefore the color information is lost. In the following sections, we obtain the color-texture information of an image by deriving its LPB based on the Hue component.

3.2.2. Canker Lesion Zone Segmentation

A whole canker lesion includes several elements such as crater-like areas, water-soaked margins etc (Polek, 2007) as shown in figure 5(a). Canker lesions change with citrus types and the phase of the disease. Classifying canker lesions can be regarded as a multi-class classification problem. A new color-texture feature LBPH (LBP on Hue) and a feature combination method are proposed in order to describe canker lesions. This canker lesion description is based on the spatial structure of the canker lesion areas with several color quantized zones. The images of the citrus disease area are firstly transformed into HIS(Hue-Intensity-Saturation) color space from RGB. HIS color space is more related to human perception mechanism than RGB color space. Furthermore images collected in field are always under different light conditions, the hue component in HIS color space helps to reduce the effect of different lights.

Our approach is to divide the whole infected area into four zones based on the description of plant phytopathologists: the center area, the inner circular hue zone, the halo and the leaf background.

The quantization method is as follows: I is the image for segmentation, a global threshold algorithm is applied to find three optimized thresholds H_{t1}, H_{t2}, H_{t3} on hue component of I to segment image I into four zones Z_1, Z_2, Z_3, and Z_4. They may not be regularly segmented zones in shape, but the pixels with a similar hue value are labeled to be in the same zone. As shown in figure 6, after the partition, each zone mainly represents a relatively meaningful part of a canker lesion and the distribution of zones reflects the spatial structure of a canker lesion.

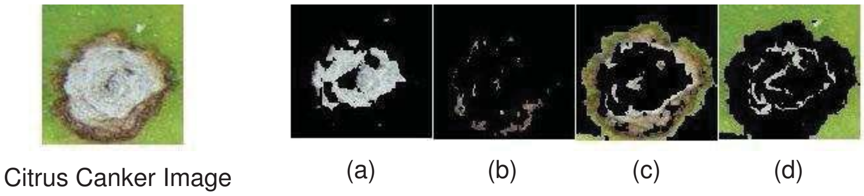


Figure 6: Citrus Canker zone segmentation; The hue-thresholds used are 0.1797, 0.2900 and 0.4036

3.2.3. Citrus Canker Local Feature Description

A measurement of the local color-texture feature of each zone can be defined as a LBPH descriptor. The proposed LBPH operator combines color and texture by simply deriving LBP based on hue component. It has been proved to be efficient (see comparison results in table 1) especially for color leaf images under various natural light conditions in field in our research.

example

h_3	h_2	h_1
h_4	h_c	h_0
h_5	h_6	h_7

$\longrightarrow$

thresholded

1	0	0
0		0
1	1	1

weights

8	4	2
16		1
32	64	128

Figure 7: Example of LBPH descriptor. (a) example of 8-neighborhood; (b) thresholded; (c) weights; $h_3, h_5, h_6, h_7 > h_c$; $h_0, h_1, h_2, h_4 < h_c$; $C = (h_3+h_5+h_6+h_7)/4-(h_0+h_1+h_2+h_4)/4$; $LBPH = (h_3*8+h_5*32+h_6*64+h_7*128)/C$

In figure 7, an image is firstly converted into HIS color space. For a local neighbored area, the central pixel h_c and its P neighbors h_p, $(p = 0, ..., P-1)$, we can calculate the joint difference texture T by subtracting h_c from h_p, where $t(h_i - h_j)$ is the difference distribution of color between neighbor pixels h_i and h_j.

$$T = t(h_0 - h_c, ..., h_{P-1} - h_c) \tag{12}$$

$$h_c - h_p = \begin{cases} 1 & if \quad h_p > h_c \\ 0 & if \quad h_p \leq h_c \end{cases} \tag{13}$$

Let the number of $h_p(h_p > h_c)$ be c_u and the number of $h_p(h_p \leq h_c)$ be c_l. Then contrast operator C can be calculated as:

$$C = \frac{S_u}{c_u} - \frac{S_l}{c_l} \tag{14}$$

where $S_u = \sum_{p=0}^{P-1} h_p,\ h_p > h_c$; and $S_l = \sum_{p=0}^{P-1} h_p,\ h_p \leq h_c$.

If c_u or c_l is zero, S_u or S_l is directly set to zero. Also from the definition 14, we can infer that C cannot be zero.

The LBPH value of a central pixel h_c is computed as:

$$LBPH_P = \frac{\sum_{p=0}^{P-1} s(h_p - h_c)2^p}{C} \tag{15}$$

where

$$s(x) = \begin{cases} 1 & if \quad x > 0 \\ 0 & if \quad x \leq 0 \end{cases}$$

3.3. Combined Local Feature

As shown in figure 6, the segmented zones may represent different parts of a canker lesion and the combination of zones can provide the spatial structure information of whole lesion. Color or texture vary in these zones, for example the texture may be water-soaked or halo. A zone-based combined local feature descriptor is proposed to integrate color and texture information. By using the segmentation methods mentioned in section 3.2.2, we can get hue-based segmented zones. The distribution of texture in a canker lesion can be computed by the mean of LBPH in each zone which is defined as formula 16:

$$Z_{kLBPH_P} = \frac{\sum_{i=1}^{N} \sum_{j=1}^{M} LBPH_{P_{(i,j)}}}{N_k}, (P(i,j) \in Z_k) \tag{16}$$

where Z_k is the mean of LBPH in zone k, N_k is the number of the pixels included in this zone. P is the number of the neighbors. N is the row number and M is the column number of this image.

Figure 8 shows an example of LBPH value distribution in each zone. The X and Y axes represent pixel position and the vertical axis represents the LBPH value. It can be seen that there are obvious differences between LBPH value distributions of the zones. To describe the color distribution we used the mean of hue components of pixels in each zone. Vector $[Z_{kLBPH_P}, Hm_k]$ is a combined feature which is used as the descriptor of a zone Z_k. For a lesion area with K zones, the combined local feature descriptor is $[Z_{1LBPH_P}, Hm_1, ..., Z_{K-1LBPH_P}, Hm_{K-1}]$, which covers all zones of a lesion and provides the structure information(by the sequence of zones), local color information and texture information of a lesion.

4. Experimental Results

The proposed method has been tested to evaluate its effectiveness [1]. All the experiments were carried out on a PC, with a Pentium 4 CPU of 3.4GHz and 1G RAM. The operating system is Microsoft Windows XP. The program was developed in Matlab version 7.0. The performance of different methods were evaluated in terms of classification rate.

The leaf images used in this research were collected from orange plants in winter in 2005 and 2006 from Guangdong province, China and in spring in 2007 from Guangxi province, China. We collaborated with a group of citrus phytopathologists from the Citrus Research Institute which is the national scientific research center of China for citrus fruits. All the images of citrus canker disease and other diseases in this paper were captured in field by the citrus phytopathologists from the citrus infected trees and they also provided the disease information so we could label each image with its relevant disease.

Different types of leaves were selected including normal leaves, citrus canker infected leaves, leaves infected by black spot of citrus, citrus melanose

[1] Some of the citrus canker datasets and source codes are available from this link `http://www-staff.lboro.ac.uk/~coqm/AdditionalInformationAboutCitrusCanker.htm`

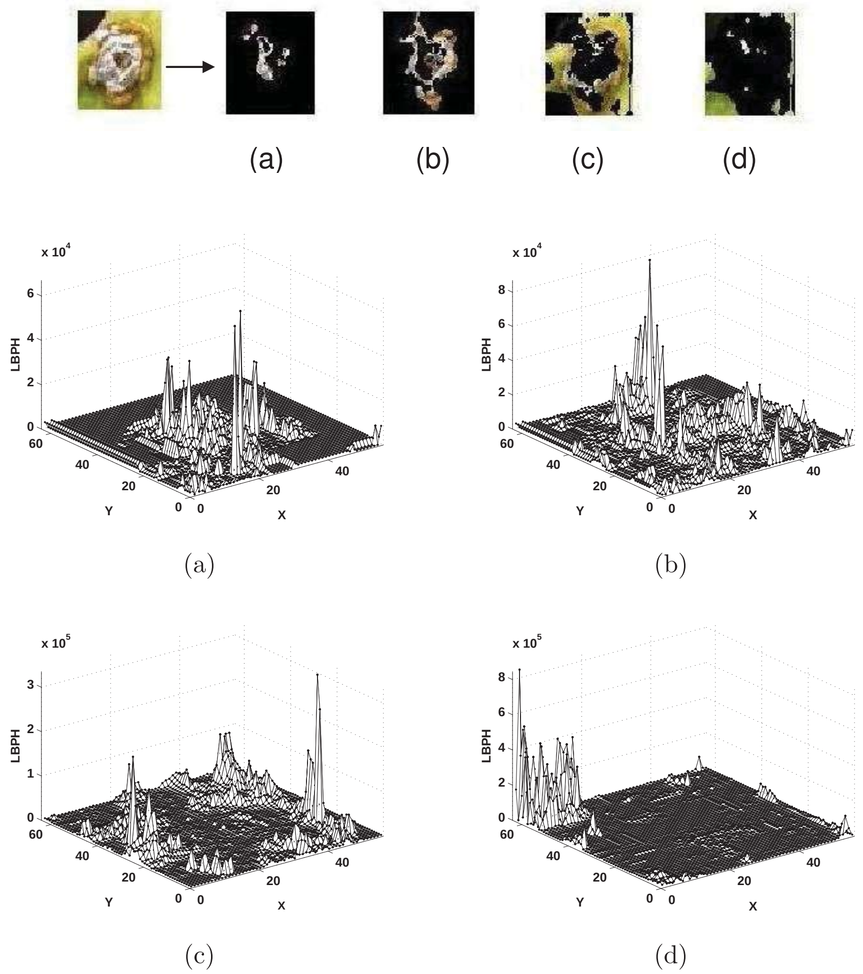


Figure 8: Example of LBPH value distribution in each zone

and citrus scab disease, they were classified into different diseases by experts. The images are at different phases of disease and taken under various environments. The original image size was between 1280×960 to 3456×2304 and the images were captured by digital camera Sony DSCP92 and Canon EOS350D.

4.1. Training Samples

The citrus canker samples were selected from more than 500 images from which the citrus phytopathologist labeled the canker lesions areas. 1000 canker samples were then obtained from the above 500 images (there might be more than one canker lesions in one image) and the lesions' length are from 60 pixels to 100 pixels, some of the citrus canker samples are shown in figure 9.

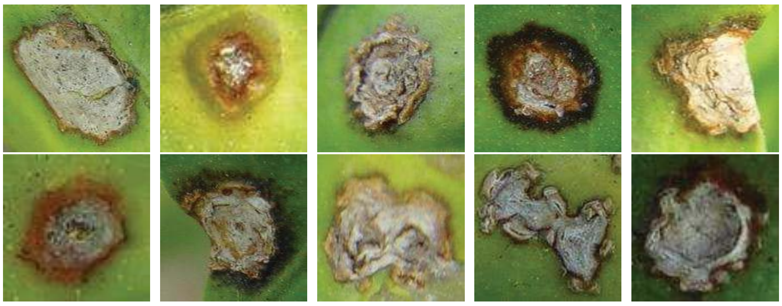


Figure 9: Samples of citrus canker lesions

The negative samples for citrus canker detection include normal leaves, leaves infected by other diseases and non-citrus leaves. We obtained the negative samples by three means: more than 2000 samples were from normal citrus leaf images as shown in 10(a); 1400 non-citrus leaf samples were searched and downloaded from web as shown in 10(b); 500 other samples were other disease lesions on citrus leaves.

After elimination of some images such as those with low image quality, we select 1000 positive citrus canker samples and 2000 negative samples. These samples were in different sizes depending on size of each lesion area. In the global matching period, the negative sample set includes normal leave samples without any lesions. As we need small window size (10×10 in this study) images to train the classifier C_1 in algorithm 1 at the first level, the

(a)

(b)

(c)

Figure 10: Negative samples. (a) Normal citrus leaves; (b) No-citrus leaves; (c) Other citrus disease lesions

original positive and negative samples were divided into 10×10 sub-images. The positive sample set with 7000 samples in 10×10 image size was created by the above process. Negative sample set with 10000 samples in the same size was simply set up by randomly selecting sub-images from the 2000 negative image samples.

The first level classifier C_1 was trained 100 rounds on the training sample set of *Set10000-10* which including 4000 positive 10×10 samples and 6000 negative 10×10 samples. At the second level of global matching, 600 positive samples from the above 1000 positive samples and 600 negative samples from the 2000 negative samples were randomly selected and normalized to 120×120 as *Set1200-120* to train the classifier C_2.

4.2. System Testing Samples

In the experiments, we chose two test sets in which samples are different from those in training. One set consists of 200 positive samples covering six canker lesion types and 200 negative samples including normal citrus leaves (figure 10(a)), non-lesion samples (figure 10(b)) and other citrus disease lesions (including those very similar to real citrus canker lesions and those relatively easy to distinguish, see figure 10(c)). The second test set has 891 randomly selected lesion samples including citrus canker and other citrus diseases which are very similar to the real citrus canker disease (e.g. blackspot, melanose, and citrus scab disease), therefore, it is more difficult to detect the true citrus canker than the first test set. This 891 data set is only used to compare the proposed approach with citrus human experts to test the system performance under this challenge situation. In the following, $Set400$ represents the first test set and $Set891$ represents the second test set.

4.3. Comparison of Different Texture Descriptors

This section reports the experimental results on $Set400$ using different texture descriptors: LBPH feature, original LBP operator and Gabor operator in the second stage of hierarchical detection procedure, in which the classifier C_2 were trained using different features on the *Set1200-120* as mentioned in 4.1. Table 1 shows the comparison results of the three texture descriptors on $Set400$ during conducting the hierarchical detection. In the figure, "$LBPH_8$" represents the features proposed in section 3.2 at canker detection phase; while "$Gabor_{6,8}$" represents Gabor features on six scales and eight directions; and "LBP_8" represents the original $LBP_{8,1}$ operator to describe the texture. We can find that the classification performance is 88% for $LBPH_8$ and it is higher than the original LBP_8 whose classification rate is 85.25%. Also $LBPH_8$ obtained a better classification result than $Gabor_{6,8}$ which has high-dimension features than $LBPH_8$.

Table 1: comparison of different texture descriptors

	Classification Rate	canker	non-disease	other disease
LBP_8	0.8525	0.98	0.64	0.81
$LBPH_8$	0.88	0.975	0.67	0.9
$Gabor_{6,8}$	0.86	0.975	0.64	0.85

4.4. Zone-based Features vs. Whole-image-based Features

In section 3.2.2 we proposed a color-quantized method to divide a lesion area into four zones and extract features from each zone, we keep classifier C_1 and retrain C_2 using *Set1200-120* using two different features. The test set is $Set400$. Table 2 lists the experimental results of zone-based and whole-image-based methods in the canker detection using $LBPH_8$ feature descriptor on $Set400$ data set. Because it contains some spatial and more detailed information than area-based features, the zone-based method provides better results with the same type of features. More importantly, zone-based features have their obvious advantages on distinguishing canker lesions from other disease lesions. Especially for the similar diseases identification, the zone-based method obtained 90% classification correct rate while the whole-image-based method only had 20%.

Table 2: comparison of zone-based and whole-image-based features

	Classification Rate	canker	non-disease	other disease
$Zone-based$	0.88	0.975	0.67	0.9
$Whole-image-based$	0.6725	0.895	0.70	0.2

4.5. Comparison of Different Classifiers

Neural Networks such as Radial Basis Network(RBN), Support Vector Machine(SVM) and k-nearest neighbors algorithms have been successfully exploited in various pattern recognition problems. In this research, we train these classifiers on *Set1200-120* at canker detection stage as a single type canker classifier and compare their performance with AdaBoost classifier on $Set400$. RBF is used as the kernel function of SVM and the number of nearest neighbors is set to be 4 shown as KNN_4 in table 3. In this table, TPR means true positive rate and FPR means false positive rate. It can be seen Adaboost classifier outperformed the other classifiers in this problem on both TPR and FPR, and RBN worked better than KNN_4 and SVM.

Table 3: comparison of different classifiers

	Classification Rate	TPR	FPR
$AdaBoost$	0.88	0.975	0.785
RBN	0.7325	0.88	0.585
KNN_4	0.6925	0.92	0.465
SVM	0.63	0.6375	0.6825

4.6. Subclasses Classifiers vs. All-against-all Detection

In section 2, subclasses classifiers are trained for each type of citrus canker lesion at canker detection stage and these classifiers are combined to conduct the classification task. We selected 600 samples canker lesions which were divided into six types, and each type canker lesion classifier was trained for 50 rounds on the set of 100 positive samples and 100 other similar disease lesions to train the classifiers. Another strategy is to train an all-against-all classifier that covers 600 all types of canker lesions and all types of negative samples. The two types of classifiers are all based on AdaBoost and the number of samples for training all-against-all classifiers are six times of each subclass classifier. Figure 11 shows the classification rate of six-subclass classifiers and all-against-all classifier during training. It is shown that the all-against-all classifier needed more rounds of training to reach stable classification accuracy than subclass classifiers did.

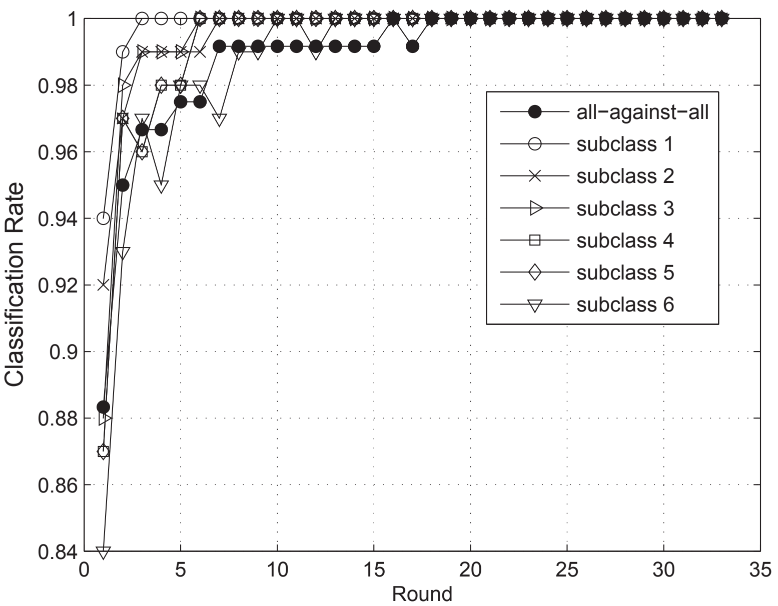
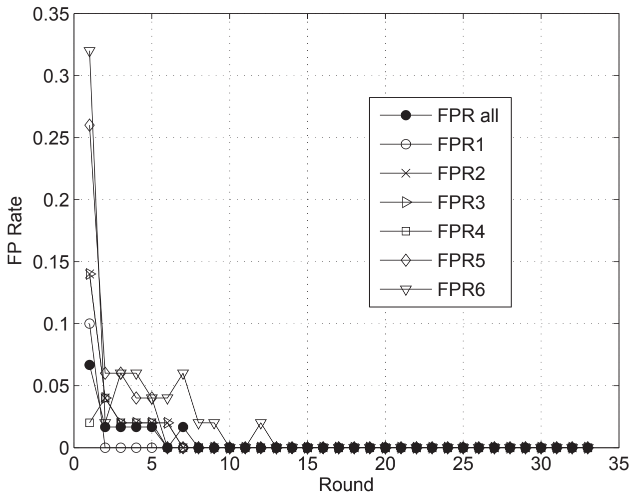
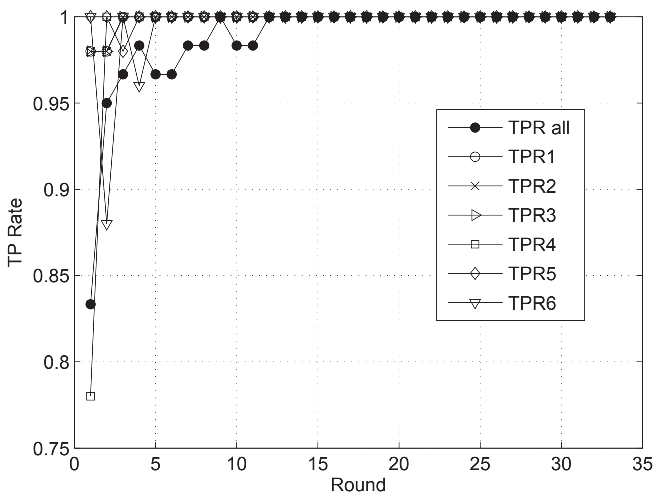


Figure 11: Training classification rates comparison of subclass classifiers vs.all-against-all classifier.

Figure 12 detailed the comparison of TPR and FPR during training between two methods. Table 4 compares the experimental results for subclass classifiers and the all-against-all classifier. It can be seen that the subclass classifiers can identify the canker lesions more accurately; while the all-against-all classifier performs better on non-lesion samples. Considering



(a) FP rates



(b) TP rates

Figure 12: FP and TP comparison of subclass classifiers vs.all-against-all classifier.

the harm of the citrus canker, the miss of canker in detection is more dangerous than the non-lesion, therefore subclass strategy is more reasonable for this research.

Table 4: results from subclasses classifier and all-against-all

	Classification Rate	canker	non-disease	other disease
$Subclasses$	0.88	0.975	0.67	0.9
$All - againt - all$	0.8475	0.83	0.80	0.93

4.7. Machine Vision vs. Human Vision

In our experiments, we chose $Set891$ (in which each sample's citrus canker type was determined by a plant expert in field) to compare the performance of the proposed approach with human experts. We randomly changed the order of the $Set891$ samples and then sent them to other experienced plant experts who never saw them before. The experts were required to classify each sample image on PC screen. We compared the expert's classification results with the results gained by the proposed approach. We used hierarchical detection method, zone-based combined features and AdaBoost classifier as mentioned in previous sections. Table 5 shows the comparison results. It can be seen that the proposed approach achieves a quite similar result as the experts.

In this experiment, a few factors might affect the detection success rate of human experts. Detecting lesion images on screen is quiet different from the way in field. Plant experts use several modalities when working in field including vision and touch etc., while in above comparison, only one modality, vision, was used. In field, experts make judgments by observing the leaves/lesions from different angles. Especially on the late stage of canker disease, the lesions' center bulges on the leaf surface and experts usually observe lesions from each side of the leaves and sometimes they will make the decision by touching the leaves as well. By discussing with some plant experts we found that when experts work in field, the types of lesions are usually less than in $Set891$, they usually need to distinguish one or two diseases at one site. The $Set891$ combines true citrus canker samples and several other very similar citrus disease samples to test the performance of the proposed approach under this more challenging situation. In this dataset, for some citrus leave images, even human experts cannot be quite sure whether it is true citrus canker or not by just looking at one image on computer screen.

Also in field the experts can check several leaves on the same tree, thus even they are not quite sure about one or two lesions they can still make the right decision eventually; while on computer screen, they need to make decision for each lesion image. When required to judge several hundreds pictures on screen, some experts said their emotional instability changed during this process and they had different feel from in field. Furthermore, the quality of the pictures in datasets varies, partial details of some pictures are not clear. All the above factors cause the relative lower success rate of human experts on screen than in field.

The camera-based canker detection system can not replace plant experts in field or in dedicated labs. However, the proposed method aims to work from a remote place and to quickly obtain an initial detection result. It can be used as an early detection/warning system to detect canker disease at their very early stage or as a server-based remote pre-detection method using images transmitted through internet. Since the citrus plants are widely distributed and we do not have enough plant experts, camera-based systems can be used to select the suspicious canker samples and then experts can make further confirmation/final diagnosis or go to the field to make further checks.

Table 5: machine vision vs. human vision

	classification rate
Machine vision	0.8799
Human vision	0.8687

5. Conclusions

This paper presented an approach to automatically detecting citrus canker from citrus leaf images captured in field. A hierarchical detection strategy was introduced to segment lesion leaf images captured in field from background, which is different from previous research based on images collected in a laboratory environment. Then a citrus canker feature descriptor was proposed by combining leaf image color and texture information to model citrus canker lesions. Local LBPH descriptors were used in order to reveal the spatial properties of citrus canker in each lesion zone. A modified AdaBoost algorithm (SceBoost) which we developed before was used to select the most significant features.

Different feature operators and classification techniques were evaluated and compared based on citrus leaf samples in this research including several kinds of citrus diseases and normal citrus leaves in different environments. The experimental results demonstrated that the proposed approach leaded to a higher classification accuracy than other methods. Meanwhile the experiment compared the proposed approach with human expert classification, and the results showed that the classification accuracy of the proposed approach is similar to citrus plant's experts who examined the image of each citrus leaf on computer screen. It proves that the proposed approach in this paper has great potential to be applied in real world. Future study will simulate the experts' observation to combine multi-angle images of a citrus leaf for identification and extend the proposed approach to other plants' disease detection and quality management.

Acknowledgements

The first author thanks Citrus Research Institute of China for providing pre-classified citrus disease datasets. The authors also wish to thank anonymous reviewers for their constructive and detailed comments.

References

Belasque, J., Gasparoto, M., Marcassa, L., 2008. Detection of mechanical and disease stresses in citrus plants by fluorescence spectroscopy. Applied Optics 47 (11), 1922–1926.

Dae, G., Tomas, F., Qin, J., Duke, M., 2009. Classification of grapefruit peel diseases using color texture feature analysis. Int J. Agric. and Biol. Eng. 2 (3), 41–50.

Das, A., 2003. Citrus canker - a review. Journal of Applied Horticulture, 52–60.

Freund, Y., 1995. Boosting a weak learning algorithm by majority. Information and computation, 256 –285.

Freund, Y., Schapire, R., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences 55 (1), 119–139.

Gambley, C., Miles, A., Ramsden, M., Doogan, V., Thomas, J., Parmenter, K., Whittle, P., 2009. The distribution and spread of citrus canker in emerald, australia. Australasian Plant Pathology 38 (6), 547–557.

Golmohammadi, M., Cubero, J., Penalver, J., Quesada, J., Lopez, M., Llop, P., 2007. Diagnosis of xanthomonas axonopodis pv. citri, causal agent of citrus canker, in commercial fruits by isolation and pcr-based methods. Journal of Applied Microbiology 103 (6), 2309–2315.

Gottwald, T., Graham, J., Schubert, T., August 2002. Citrus canker: The pathogen and its impact. Plant Management Network. http://www.plantmanagementnetwork.org/pub/php/review/citruscanker/.

Gottwald, T., Hughes, G., Graham, J. H., Sun, X., Riley, T., 2001. The citrus canker epidemic in florida: The scientific basis of regulatory eradication policy for an invasive species. Phytopathology 91 (1), 30–34.

Gottwald, T., Timmer, L., 1995. The efficacy of windbreaks in reducing the spread of citrus canker caused by xanthomonas campestris pv citri. Tropical Agriculture 72 (3), 194–201.

Kearns, M., Valiant, L., 1989. Crytographic limitation on learning boolean formulae and finite automata. In: Proc. of the 21st annual ACM Symposium on Theory of Computing. New York. NY:ACM press. pp. 433–444.

Li, S. Z., Zhang, Z., 2004. Float boost learning and statistical face detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1112–1123.

Lins, E., Belasque, J., Marcassa, L., 2009. Detection of citrus canker in citrus plants using laser induced fluorescence spectroscopy. Precision Agriculture 10 (4), 319–330.

Ojala, T., Pietikainen, M., Harwood, D., 1996. A comparative study of texture measures with classification based on feature distributions. Pattern Recognition 29 (1), 51–59.

Park, D., Hyun, J., Park, Y., Kim, J., Kang, H., Hahn, J., Go, S., 2006. Sensitive and specific detection of xanthomonas axonopodis pv. citri by pcr using pathovar specific primers based on hrpw gene sequences. Microbiological Research 161 (2), 145–149.

Park, D., Young, J., 2006. Rapid and efficient identification protocols for xanthomonas campestris pv. citri for quarantine investigations. New Zealand Journal of Crop and Horticultural Science 34 (3), 195–205.

Polek, M., 2007. Citrus bacterial canker disease and Huanglongbing (Citrus greening). University of California, Agriculture and Natural Resources, ANR Publications. http://anrcatalog.ucdavis.edu/IntegratedPestManagement/8218.aspx, accessed in 2010.

Pydipati, R., Burks, T. F., Lee, W. S., 2006. Identification of citrus disease using color texture features and discriminant analysis. Computers and Electronics in Agriculture 52 (1), 49–59.

Qin, J., Burks, T. F.and Ritenour, M. A., Bonn, W. G., 2009. Detection of citrus canker using hyperspectral reflectance imaging with spectral information divergence. Journal of Food Engineering 93 (3), 183–191.

Vernière, C. J., Gottwald, T., Pruvost, O., 2003. Disease development and symptom expression of xanthomonas axonopodis pv. citri in various citrus plant tissues. Phytopatgology, 832–843.

Xiao, R., Zhu, L., Zhang, H., 2003. Boosting chain learning for object detection. In: IEEE International Conference on Computer Vision.

Zhang, M., 2008. Research on key technologies of citrus canker intelligent detection. PhD thesis, Chongqing University.

Zhang, M., Zhu, Q., Liu, F., 2007. Sceboost learning algorithm for feature selection. In: Proceedings of ICNC 2007: Third International Conference on Natural Computation, Vol 1, 285-289.