Journal of Information Hiding and Privacy Protection
DOI:10.32604/jihpp.2020.010114
Article

Research on Prevention of Citrus Anthracnose Based on Image Retrieval Technology

Xuefei Du* and Xuyu Xiang

College of Computer and Information Engineering, Central South University of Forestry and Technology, Changsha, China
*Corresponding Author: Xuefei Du. Email: oandx@foxmail.com
Received: 10 May 2020; Accepted: 28 May 2020

Abstract: Citrus anthracnose is a common fungal disease in citrus-growing areas in China, which causes very serious damage. At present, the manual management method is time-consuming and labor-consuming, which reduces the control effect of citrus anthracnose. Therefore, by designing and running the image retrieval system of citrus anthracnose, the automatic recognition and analysis of citrus anthracnose control were realized, and the control effect of citrus anthracnose was improved. In this paper, based on the self-collected and collated citrus anthracnose image database, we use three image features to realize an image retrieval system based on citrus anthracnose through SMV, AP clustering optimization. The results show that: 1) In the accuracy of image feature retrieval, Gist feature extraction > cumulative color histogram > Gabor texture feature; 2) In the maximum divergence diversity retrieval, semi-supervised AP clustering retrieval > AP clustering retrieval > SVM relevance feedback results > initial retrieval. 3) Practice shows that this technology can reduce the workload of monitoring and management in the control process of citrus planting area, and promote the intelligent and efficient control of citrus anthracnose, which has high practical application value.

Keywords: Citrus; anthracnose; control; image retrieval technology

1 Introduction

At present, there are numerous studies based on image retrieval technology in the field of plants, but most studies are still committed to the classification and discrimination of some plant species, and only a few studies are applied in the management of agricultural information. In the field of the citrus anthracnose control based on image retrieval technology, Pydipati et al. used the co-occurrence method to extract image characteristics of three citrus diseases and healthy leaves which are classified and identified in combination with a statistical classification algorithm [1]. Liu proposed an improved cascade Ada Boost classification mode and a search mechanism merged with window for citrus canker [2]. Zhao et al. constructed two classifiers with 95% recognition rate [3] (partial least square discriminant analysis and BP network), by Fourier spectrum analysis for the 7 typical citrus diseases images.

In the early development and application of image retrieval system, scholars tended to traditional linear feature extraction methods. Such operation sacrificed the generalization learning ability of samples and limited the development and application of the system [4–5]. Especially in some complex applications with distinguishing subtle features, the system will make many errors, which is more common in studies on plant image retrieval. To provide users with accurate information, many scholars focused on diversity retrieval method and developed methods based on support vector machine (SVM) relevance feedback. During each feedback, the samples of positive and negative examples marked by users are learned; retrieval is conducted through SVM classifier model. However, it is difficult to use

SVM in kernel function selection [6].

On this basis, this study realized the automatic control of citrus anthracnose by designing and simulating the image retrieval system for citrus anthracnose, and improved the control efficiency of citrus anthracnose.

2 Design of Image Retrieval System

In combination with the status of citrus anthracnose control based on image retrieval technology, the traditional initial search is more likely to have too many wrong cases in the check results. Thus it cannot meet the design requirements. In order to reduce wrong cases in retrieval and improve the accuracy of retrieval results, this study optimized on the basis of SVM relevance feedback, AP class and AP multi-clustering algorithm, so as to strive to realize the automatic control of citrus anthracnose through simulation and design of image retrieval system of citrus anthrax, thus improving the control efficiency of citrus anthracnose.

2.1 Framework of Image Retrieval System

To eliminate the impacts of numerous duplicate contents in the image library on the retrieval, the image retrieval system adopted diversity retrieval technique to remove duplicate or overly similar images by processing preliminary retrieval results. Meanwhile, to solve the problem of "semantic gap", the data feedback technique was introduced in this system to improve the relevance of preliminary retrieval results [7]. The specific steps of the algorithm are shown in Fig. 1.

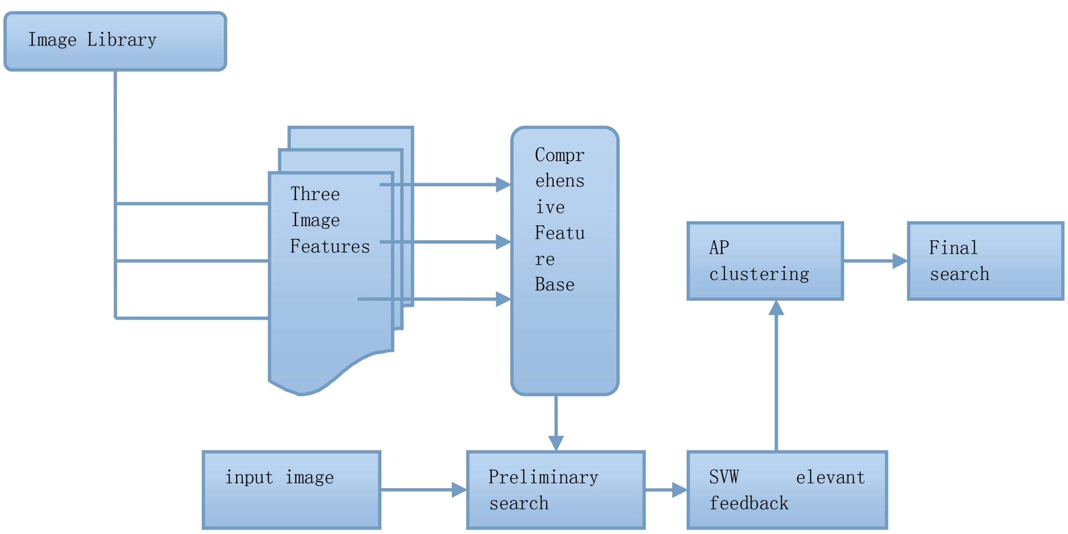

Figure 1: General algorithm framework

2.2 Preparation of Image Library

The image library for citrus anthracnose is mainly records the disease characteristics of citrus anthrax. In this study, a large number of physical photos were collected according to such characteristics and the main lesion characteristics and representative images were summarized.

Figure 2: Partial schematic diagram for image library

As shown in Fig. 2, the main lesion areas of citrus anthracnose are the flowers, branches, leaves, and fruits of citrus trees; and some are in fruit stem [8–9]. Leaf disease characteristics are divided into two kinds: one is the leaf margin and leaf tip were light grayish brown; the other starts at the tip of the leaf, which is dark green at first and may have vermilion spots in wet weather. Branch disease characteristics are divided into two kinds: A top-down part which is initially brown; the other is in the middle of leaf tip and is light brown at the early stage. The two types are grayish-white after death. There is clear boundary between lesion area and healthy part. There are four types of fruit lesions: brown spots may appear on or near the pedicle of the stiff fruit, with white mildew or vermilion spots on the lesion; the fruit waist of scabbed fruit is yellow-brown and in leathery shape; small tear-like reddish-brown bumps on the skin surface of tear-like fruit; the lesions of rotten fruit mainly occur during storage and transportation, and there are pale mildew spots in the lesion area.

2.3 Feature Extraction for Citrus Anthracnose Images

2.3.1 Extraction of Cumulative Color Histogram

The feature extracted by cumulative color histogram has strong robustness and is insensitive to changes in quantization parameters. Thus it is superior to general color histogram in color feature extraction. The cumulative color histogram for image $\overline{H}$ is defined as [10]:

$$\overline{\mathrm{H}}=\{<\overline{h}[c_1],\overline{h}[c_2],\overline{h}[c_3],\cdots,\overline{h}[c_n]>|\,0\le\overline{h}[c_k]\le 1\} \quad (1)$$

where: $\overline{h}[c_k]=\sum_{c_i\le c_k} h[c_i]$ represents the cumulative estimation of pixels from colors c_i to c_k; n represents the pixel value range of image color quantization. There was a significant difference in leaf and fruit colors between lesion area and surrounding area for images of citrus anthracnose collected in this study. The randomly-taken photos of citrus fruit lesions were examined and analyzed. The retrieval accuracy in the first 20 photos is 85%.

2.3.2 Gabor Texture Feature Extraction

The wavelet feature of Gabor is very similar to the visual stimulus-response of human visual system to objects. The diseased organs of anthracnose citrus have significant visual sensory changes [11]. The sensory comparison is shown in Fig. 3.

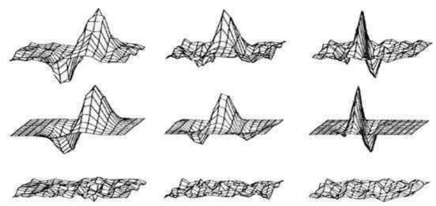

Figure 3: Gabor filters

Since the local features of different scales need to be grasped, 20 filters composed of 4 center frequencies ($w_0 = 0, 1, 2, 3$) and 5 directions ($\theta = 0, 1, 2, 3, 4$) was added for wavelet transformation, so as to ensure the detection of images with higher characteristic dimensions and avoid too long processing time. In this study, the first 20 images were retrieved by Gabor texture features; the retrieval accuracy was 70%. This analysis showed that the significant structural changes in lesion areas caused significant changes in texture features. Therefore, this method is basically applicable.

2.3.3 Gist Feature Extraction

In Fig. 4, images were filtrated through the Gabor filters set with different directions and dimensions. The filtrated images were divided into 4 × 4 windows, and each window is averaged. The window averages in different directions and dimensions were taken as the Gist feature vectors of images. Finally, the detection showed that the detection accuracy of the 20 images was 90%, and the retrieval effect is good.

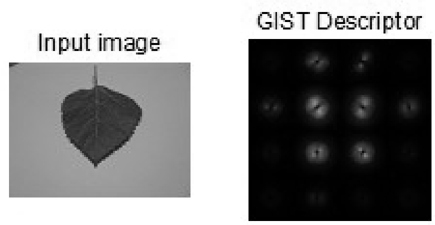

Figure 4: GIST descriptor

The above-mentioned three methods of image feature extraction and retrieval were integrated to retrieve the three feature regions of citrus lesions and calculated the retrieval accuracy of the first 20 images.

Table 1: Comparison of retrieval accuracy (%) of image features of disease sites of different citrus anthracnose

Image feature	Leaf		Branch		Fruit			
	Leaf spot	Leaf blight	Top	Middle	Stiff fruit	Scabbed fruit	Tear-like fruit	Rotten fruit
Cumulative color histogram	80%	75%	70%	65%	60%	75%	65%	75%
Gabor texture feature	70%	70%	55%	60%	55%	75%	60%	85%
Gist feature extraction	85%	90%	70%	70%	80%	85%	70%	90%

Tab. 1 showed the retrieval accuracy of the first 20 images under the 8 conditions in the image library. When analyzing different image features, the same example images of citrus anthracnose were selected. According to the retrieval results, the accuracy of different feature retrieval was more than 50%. However, compared with the accuracy of the three search results, Gist feature extraction was superior to cumulative color histogram, while the effect of Gabor texture was the worst among the three. The main reason is that the color, shape and texture of citrus anthrax are obviously changed, which reduces the difficulty of system differentiation. However, in the position of the branch shoot, there was difficulty in regularity due to the characteristics of its relative shape, so the accuracy of capturing information by texture feature decreased. On the contrary, because of the chromatic aberration after the color change in the disease area, the retrieval accuracy of cumulative color histogram in this position is relatively higher.

2.4 SVM Feedback

To solve problems of the low accuracy and the appearance of irrelevant images, this study used LIBSVM toolkit [12] to retrieve the positive and negative example images in the preliminary results, and the search results were marked. Also, SVM training was carried out according to positive and negative example images. After the training, the trained SVM classifier was used to classify the image library and sort out positive sample images in turn.

2.5 Clustering

To make the search results more diverse, similar images should be deleted. This study adopted AP clustering algorithm. Each data point L L sets its bias parameter $S(l,l)$. The greater the $S(l,l)$, the more possible the corresponding point l will be selected as the center point. First, AP algorithm assumed that the $S(l,l)$ of all pixel points is the same P value, and then the clustering center was determined through [13]. The AP algorithm pseudocode is expressed as follow:

Algorithm 1: AP Algorithm

Input: data set *i*
Output: *P*

$r(i,l) \leftarrow s(i,l) - \max_{i \neq l}\{a(i,l) + s(i,l)\}$

if *i≠l*

$a(i,i) \leftarrow \min\{0, r(l,l)\} + \sum_{i \ s.t. \ i \in \{i,l\}} \max\{0, r(i,l)\}$

IF *i=l*

$a(l,l) \leftarrow \sum_{i \ s.t. \ i \neq l} \max\{0, r(i,l)\}$

where, the representative matrix $R = [r(i,l)]_{n \times m}$ indicates that X_i is suitable as the class representative point of X_i. The appropriate matrix $A = [a(i,l)]_{n \times m}$ represents the appropriate degree for X_i to select X_l as the point of class representation. Meanwhile, since AP clustering method is still loosely presented in the data structure and the clustering accuracy will also decrease accordingly, the semi-supervised AP clustering algorithm will be adopted in this study [14]. In other words, in the iterative process of AP clustering processing, supervisory information will be introduced to guide the clustering algorithm to obtain the optimal results.

3 Experimental Results and Analysis

3.1 Citrus Anthracnose Image Retrieval Experiment

This study used self-developed citrus anthracnose image library, which was used to collect more than 8000 images of citrus anthracnose conditions of citrus production areas in 9 provinces (cities), such as Hunan, Jiangxi, Fujian, and Zhejiang. Such citrus anthracnose diseases have many types, and the diseases of individuals have certain differences. Therefore, the image features in the database were extracted, then randomly-obtained example images were input for preliminary retrieval based on the similarity between images. Afterward, the positive and negative example images after the preliminary retrieval were marked; and SVM optimization training was conducted according to them to improve the accuracy of the system. Then the trained classifier was used to carry out distribution processing of database according to the classification of relevance and irrelevance. Finally, relevant images were clustered, and the most representative class center point was selected from each class and displaced according to similarity.

In this paper, setting the leaf citrus anthracnose diseases as the study case, the experiment focused on the diversity retrieval results of different clustering algorithms. The specific classification was as follows:

a. Traditional retrieval based on similarity image retrieval.

b. Relevance feedback retrieval, retrieval results obtained through SVM feedback.

c. AP clustering diversity retrieval, the duplicate images are deleted by AP clustering algorithm based on relevance feedback retrieval.

d. Semi-supervised AP clustering diversity retrieval, using the semi-supervised AP clustering algorithm based on the optimization of AP deduplication for diversity retrieval.

3.2 Results Analysis

The first 20 images screened out by the four retrieval methods are as follows:

(a) Preliminary retrieval results

(b) SVM relevance feedback results

(c) AP clustering retrieval results

(d) Semi-supervised AP clustering retrieval results

Figure 5: Results of diversity retrieval with maximum divergence

Fig. 5 showed the results of the first 20 retrieval images. The images of citrus leaf diseases were taken in this study as examples.

Fig. 5(a) is the preliminary retrieval results. It can be seen that although many retrieved images were consistent with the example images of citrus anthracnose, some images are normal leaves without disease, indicating that there were many errors in the retrieval results. Especially the first image in the third row, there are some drops on the leaf that are similar in color to the anthrax lesions. It actually is an identification error. Besides, some images retrieved were fruit lesion images.

Whereas after SVM relevance feedback, the resulting images are shown in Fig. 5(b). Compared with the preliminary retrieval results, the number of wrong examples has been reduced. However, there are still a large number of problems; and some normal leaves are still covered.

Obviously, the above two methods are not suitable for further extending the function of the image retrieval system. According to Fig. 5(c), after AP clustering retrieval, the retrieval results showed good relevance. In particular, the main retrieval results are focused on the leaf. Nevertheless, there are still a few false examples.

After semi-supervised AP clustering retrieval, the overall effect was greatly improved. As shown in Fig. 5(d), some wrong examples have disappeared, but these wrong examples may still be displayed in subsequent sequence. The overall retrieval results were well.

To sum up, in the results of diversity retrieval with maximum divergence, semi-supervised AP clustering retrieval > AP clustering retrieval > SVM relevance feedback results > preliminary retrieval.

4 Conclusion

The self-collected citrus anthracnose database in this study covers 8 types of conditions such as leaf, branch and fruit citrus anthracnose and adopts three feature extraction methods: Gist feature extraction, cumulative color histogram, and Gabor texture feature. According to the result, the effect of Gist feature extract is the best. Preliminary retrieval was conducted based on image retrieval system. Tests based on SVM relevance feedback optimization and AP class and AP multi-clustering algorithm were carried out.

In conclusion, based on traditional detection methods, this image detection system is optimized successively by SVM relevance feedback optimization, AP class and AP multi-clustering algorithm in turn. This study can partly meet the requirement of intelligent and efficient control of citrus anthrax in modern production areas and provide new ideas and technical support for the control of citrus anthracnose.

In the Internet application, some scholars proposed diversity similarity ordering methods, combined with the visual information and semantic information of the image, and put forwarded some criteria to optimize the diversity retrieval, so that the retrieval results have a certain typicality. The methods based on clustering can select the samples close to the pre-clustering center to realize retrieval diversity and improve the accuracy of retrieval [15–16]. To adapt to the complex morphology and color characteristic information of citrus, the system needs to improve the learning performance to improve the generalization ability of the algorithm by kernel transformation in the study. Meanwhile, to realize the more intelligent feedback of the system, automatic intelligent robots will be introduced for image acquisition and image region feedback to realize intelligent and automatic control of citrus anthracnose and provide program reference and technical support for relevant agricultural information management.

Funding Statement: This work was supported in part by the National Natural Science Foundation of China under Grant 61772561; in part by the Key Research and Development Plan of Hunan Province under Grant 2018NK2012; in part by the Science Research Projects of Hunan Provincial Education Department under Grant 18A174; in part by the Degree & Postgraduate Education Reform Project of Hunan Province under Grant 209; and in part by the Postgraduate Education and Teaching Reform Project of Central South University of Forestry & Technology under Grant 2019JG013.

Conflicts of Interest: We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

References

[1] R. Pydipati, T. F. Burks and W. S. Lee, "Identification of citrus disease using color texture features and discriminant analysis," *Computers and Electronics in Agriculture*, vol. 52, no. 1–2, pp. 49–59, 2006.

[2] F. Liu, "Study and realization of citrus canker detection system," Chongqing University, Chongqing, 2009.

[3] X. X. Zhao, G. Liu, W. X. Li, J. M. Hao, X. P. Hou *et al.,* "Infrared spectroscopy combined with LDA and BPNN based on wavelet transform to detect citrus osbeck anthracnose," *Chinese Journal of Lasers*, vol. 41, no. 1, 2014.

[4] Z. Yan and Z. Lin, "The augmented homogeneous coordinates matrix-based projective mismatch removal for partial-duplicate image search," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 181–193, 2018.

[5] V. Y. Belozyorov, "A novel search method of chaotic autonomous quadratic dynamical systems without equilibrium points," *Nonlinear Dynamics*, vol. 86, no. 2, pp. 835–860, 2016.

[6] S. Paul, M. Magdon-Ismail and P. Drineas, "Feature selection for linear SVM with provable guarantees," *Pattern Recognition*, vol. 60, pp. 205–214, 2016.

[7] J. Zhang, L. S. Shen and D. D. Feng, "A Survey of Image Retrieval Based on Visual Perception," *Acta Electronica Sinica*, vol. 36, no. 3, pp. 494–499, 2008.

[8] A. P. Ramos, P. Talhinhas, S. Sreenivasaprasad and H. Oliveira, "Characterization of colletotrichum gloeosporioides, as the main causal agent of citrus anthracnose, and c. karstiias species preferentially associated with lemon twig dieback in Portugal," *Phytoparasitica*, vol. 44, no. 4, pp. 549–561, 2016.

[9] H. Z. Yin, Z. Y. Fu, C. J. Peng, C. Y. Zhao, T. K. Bai *et al.,* "Integrated test and demonstration effect of whole-process green control technology for citrus diseases and pests," *China Plant Protection*, vol. 53, no. 2, pp. 51–57, 2018.

[10] L. Zhang, J. J. Bi and Z. Y. Lin, “Research on image extraction algorithms based on color features,” *Science and Technology Innovation Herald*, vol. 33, pp. 189, 2008.

[11] M. Romanos-Diogenes, “Spark deficient gabor frames,” *Pacific Journal of Mathematics*, vol. 294, no. 1, pp. 159–180, 2018.

[12] W. Jiang, Y. X. Wang, W. Zhang, J. T. Sun, J. Yang *et al.,* “NDT Technology for cantaloupe maturity based on LIBSVM,” *Jiangsu Agricultural Sciences*, vol. 44, no. 11, pp. 396–398, 2016.

[13] X. F. Yang and L. F. Liu, “SVM bus station short-term passenger flow forecast,” *Journal of Wuhan University of Technology* (*Transportation Science and Engineering Edition*), vol. 40, no. 1, pp. 36–40, 2016.

[14] Y. B. Han and P. Han, “An iterative weighted updating semi-supervised AP clustering algorithm with accelerated operator,” *Application Research of Computers*, no. 2, pp. 376–378, 2015.

[15] H. Wang, X. Q. Lv and Y. Huang, “AP clustering image labeling based on distance measure learning,” *Computer Engineering and Applications*, vol. 53, no. 23, pp. 159–164, 2017.

[16] X. Z. Ge, W. L. Xiong and B. G. Xu, “A multi-model soft-measurement method based on adaptive AP clustering,” *Control Engineering*, vol. 24, no. 10, pp. 2070–2076, 2017.