

The Application of Image Retrieval Technology in the Prevention of Diseases and Pests in Fruit Trees

Wang Zhi-jun^{1,2,*}, Liu Xin², Jiang Meng², and Cheng Shu-han^{2,**}

¹ College of Horticulture Science and Engineering, Shandong Agricultural University,
Taian 271018 China

² College of Information Science and Engineering, Shandong Agricultural University,
Taian 271018 China

{wzj,shcheng}@sdau.edu.cn, liuxinmails@sina.com,

jiangmengbz@163.com

Abstract. Diseases and pests in fruit trees is a key factor influencing the quality and quantity of fruit. Image retrieval technology into diseases and pests detection is an important way to improve the fruit tree management level and the quality and quantity of the fruit. This paper reviews the content-based image retrieval technology and the application of technology in the aspect of plant diseases and insect pests and proposed pest detection methods, steps and future research directions.

Keywords: image retrieval technology, pests and diseases of fruit trees, Preventive and treatment.

1 Introduction

China is a huge producer of fruit. Total harvested area and yield as well as a variety of fruit production are the highest in the world. In the year of 2010, the fruit production of our country is 214.014 million tons[1]. Fruit industry becomes the third largest industry in planting after grain and vegetable in our country. It has already been a pillar industry of China's rural economies and an important way of promoting the employment and income for farmers. China's fruit industry has the advantage of scale but the problems of low total yield and poor quality are becoming more pronounced in which diseases and pests in fruit trees is a key factor influencing the quality and quantity of fruit. In practical production, most orchard workers make diagnosis of fruit diseases and pests by experience and feeling. Although they have obtained some results, they caused harm to the growth of fruit trees. Unable to accurately detect diseases and peats in fruit trees, fruit producers blindly apply lots of fertilizer and pesticide in order to prevent the occurrence of diseases and pests in fruit trees which is not only a waste of financial resources, material resources and manpower but also of no use. And it also affects the fruit quality and yield and destroys the ecological environment. So building an effective

* The main interesting research topics are computer networks and agriculture informationization.

** Corresponding author.

disease and pests monitoring and system, introducing image retrieval technology into diseases and pests detection and realizing digitized, intellectualized, informative and networked fruit production management is an important way to improve the fruit tree management level and the quality and quantity of the fruit.

2 Review of Image Retrieval Technology

Image Retrieval Technology originated in the 1970s. Text-based Image Retrieval (TBIR) describes image content via text messages. Text information is often called keyword comments and annotation statements and this method is actually transforming the image retrieval into corresponding text retrieval. But TBIR has shortcomings such as subjectivity in understanding images, hefty workload and inefficiency [2]. Therefore TBIR already can't satisfy the people's needs. In the 1990s, there appears Content-Based Image Retrieval, a technology that combines color, texture, shape, layout, and other characteristics of the picture and then analyses them.

2.1 Original Data-Based Query

According to Neal Krawetz[3], the key technology of Original Data-based Query is called 'Perceptual hash algorithm'. It is used to generate a 64 - bit integer for each image, called 'fingerprint' character string and then compare the different fingerprints of these images.

Determine how many bits are not the same in this 64-bit integer. In theory this is equivalent to computing the 'hamming distance'. If the number of different bits is less than 5, then we can say that these two pictures are the same but if the number is greater than 10, they are different. This comparison is very specific and it is only helpful when using relatively exact matches. It demands high levels of image and is always used to search thumbnails and similar pictures. But it is unable to distinguish transmutative images. It has obvious advantages: speed and high accuracy.

2.2 Color-Based Retrieval

Color Histogram is the most common methods of color specification [4]. The thought is color space histogram matching. It will first divide color space into several fixed subspaces and then add up the number of pixels in these subspaces of each image and in the end use the histogram to measure the similarity between images. This algorithm has advantage of simple calculation and insensitivity of translation and rotation and so on [5]. Commonly used methods of color specification also include Color Moments, Color Sets color aggregation vector and so on.

Now the method based on global color feature which is brought up by Swain and Ballard[6-7] mainly use color histogram. This method counts the statistical probability of the appearance of each color based on the color histogram. Then it uses the

intersection of the color histograms to judge the similarity of the two images. Later Pass and Zabih[8] improved the algorithm and proposed to use color aggregation vector as the index of the image. To some extent, aggregate information in the aggregation vector still retains the color space information of the image [9]. Sangoh Jeong[10] et al. came up with the idea of extracting color histogram by GMVQ. K.Konstantinidis et al [11]. Put forward a kind of connected fuzzy color histogram retrieval algorithm based on $L^*a^*b^*$ space. [12]

Image retrieval is mostly based on color histogram which has the advantages of high operation speed, low storage requirement and insensitivity of size scale and rotation changes of the image. The drawback of color histogram is that query results based on different color space retrieval systems may vary. In addition, it contains the frequency of only one color and loses the pixel's location information. Also the accuracy is not very high [2]. Therefore we usually combine other retrieval methods to retrieve plant diseases and pests of fruit trees. For instance on the basis of detecting the disease spot area, we can use color-based retrieval technology [13] to identify diseases while shape and other characteristics are almost the same.

2.3 Shape-Based Retrieval

Such methods to describe shape feature currently include two major classes: edge-based and region-based shape methods. By the former, the description mainly includes straight line description, spline curve fitting, Fourier descriptors, Gaussian parameter curve and invariant moment description. As for the later one, it mainly includes irrelevant torque of the shape, area of the region, aspect ratio of the shape, eccentricity and so on[14]. In the shape- based retrieval, edge detection research is extensively used. At present there exist many image edge extraction algorithms such as Canny operator, Robert operator, Prewitt operator, Sobel operator and wavelet operator [15]. The edge detection algorithm has the advantages of simpleness and speed. But it is a bit more demanding. The algorithm directly processes on the original image, so the detection effect is poorer when the original image is disturbed by noise. If the color mutation of the image is not particularly evident, the original edge may also be smoothed while filtering out the noise. Fourier descriptor is also too complex. Invariant moment can't guarantee the invariance while the object is greatly rotated. In real production images may be greatly influenced by the weather conditions, so the effect is not stable.

Wavelet analysis with excellent time-frequency localization characteristics is widely used in the field of image processing and pattern recognition. Wavelet-based edge detection algorithm [16] can perfectly guarantee the accuracy of the edge and overcome the quality troubles came from traditional algorithm.

Shape-based image retrieval can pre-match diseases and pests of plants in order to narrow the matching region. It can narrow the scope of diseases according to the shape of diseased area and it also can reduce the disease types though retrieving the shape of the pests. Combine with other retrieval methods, it will be more accurate.

2.4 Texture-Based Image Retrieval

Texture feature which can quantitatively describe spatial information of image to some extent refers to varying patterns composed of gray scale and color in two-dimensional space. It is image features related with the surface material. The texture analysis methods can be basically divided into four classes: statistics method, structure method, model method and space method/frequency domain analysis method. There are mainly two kinds of texture image feature descriptions. One is concurrence matrix, raised by Haralick and Shanmugam. It highlights the spatial dependencies of the texture. The other is visual characteristics method, raised by Tamura. It analyses the texture from six visual features: thickness, contrast, orientation, degree of linearity, rules and roughness. And this method mainly emphasizes visual effect [17]. Owing to the relationship between texture characteristics and surface material we can use this to retrieve the diseases. The material of the diseased area will change, which is more convenient for retrieving diseases. Likewise, it can also retrieve diseases according to the different physical features of pests.

This image retrieval has rotation invariance and stronger resistance to noise. Its disadvantages is that when the resolution of the image changes the calculated texture may be greatly biased. Besides, due to the possible affection of illumination and reflection, texture reflected from two-dimensional image is not always the real texture form 3D object surface. In diseases and pests retrieval, images may be greatly influenced by environmental factors, so the stability and accuracy is poor, but it can be used as auxiliary retrieval combined with other retrieval methods.

2.5 Content-Based Multi-feature Image Retrieval

Texture, color and shape are the most commonly used and easily extracted features in image retrieval. Extraction is not only based on one character, but based on multiple characteristics. The effect of the image retrieval system is assessed from three aspects such as precision, recall ratio and query speed. Experiments have shown that the image retrieval system based on multi-feature is superior to the image retrieval system based on single vision feature in every aspect [2]. Different representation reflects the characteristics of the media from different angle. Integrate multiple features in accordance with users' actual demand in order to improve the efficiency of image retrieval [5]. One important problem needs to be solved in multi-feature retrieval is the normalization in algorithm matches. Their expectation and variance may be quite different, so they must be normalized so that different algorithm is equal in weight [18].

The most commonly used algorithm in content-based multi-feature image retrieval is SIFT algorithm. This algorithm was put forward in 1999 by David Lowe and completed in 2004[19-20]. Multi-feature retrieval is mainly used in diseases and pests retrieval, combined with other retrieval methods, in order to improve the accuracy and speed. SIFT algorithm mainly includes five steps: scale-space maximum detection, key point positioning, direction and size determining, key point description and feature vector matching. The algorithm has the advantage of basing on multi-characteristic content and

it can accurately extract the image features and improve the image recognition and matching rate. But the process is too complex time-consuming and lack of a general optimal algorithm and the computational cost is huge.

3 The Application Status of Image Retrieval Techniques in the Prevention of Diseases and Pests

Machine vision and image processing technology have been applied in the field of agriculture and plant protection in the US and other developed countries. The application of image processing has made a certain process and development in fruit tree industry. But at home, the application of image processing technology in industry and agriculture is just beginning and still in an infancy. However, many colleges and universities began to study with image detection technique applied to agroforestry and achieved some results. But there is still a long way to go to practice. We should speed up the research process of the application of image detection technology in fruits industry in order to achieve the aim of improving production efficiency so that we can finally realize intelligentizing and automation on production and management.

In foreign countries, YutakaSasaki et al.(1999) contraposed the impact the different spectral reflectance characteristics and optical filtering have on disease, studied automatic diagnosis technology of cucumber anthracnose using genetic algorithm to identify diseases[21].

Insect identification system, studied by Dr. Jeffrey and T. Drake et al. is used for rapid identification and classification in massive samples. Image segmentation algorithm can extract single segmentation in a large number of samples and it can segment most insect samples in order to obtain better effect [22].

Huang (2007) proposed a method of detection and classification of three kinds of diseases in phalaenopsis though image analysis technology and neural network and used adaptive index transformation method to segment diseased area. Extract color and texture features and use BP neural network to classify normal and diseased plants with an accuracy of 97.2% [21].

In recent years, domestic scholars also made certain research in this field. The expert system of plant diseases and pests forecast and control technology raised by Wang Qinglei et al.(2005) has forecast and prevention function such as automatic identification of pests, assistant identification of pests, diseases and pests prevention and control strategy of plant diseases and pests[22].

Song Kai et al. (2007) applied multiple classifier of SVM to identify various diseases and a pest of corn leaves [21].

Huang Xiaoyan et al. discussed the color digital image segmentation algorithm based on mathematical morphology and proposed the method of filtration colorized digital image with morphology template which can be used in the segmentation of colorized digital images of stored-grain pests[22].

Zhang Gongtao et al. completed image matching of stored-grain pests in 2009[23].

4 Steps of Diseases and Pests Image Testing in Fruit Trees

Presently, No image retrieval method is applicable to any situation. Different use needs different analysis. Aiming at diseases and pests image testing in fruit trees, we can use the following steps to complete.

First of all, all the tree pest species should be collected as much as possible. Inviting relevant experts to classify the pests according to the structure characteristics and color characteristics of pests, such as wings, head size, shard shape, and the number of feet, color, etc. Pest characteristics are extracted and it's convenient for matching testing. Diseases also should be classified comprehensively and systematically and the feature set is planned out. Then according to the different characteristics of diseases and pests in fruit trees, Designing image processing algorithm of feature extraction which is appropriate, efficient and accurate. On this basis:

(1) Preprocessing the image. Because the weather, light and other environment have deep impact on image acquisition, using edge detection, filtering and other algorithms for image to processing in advance. Cutting the edge of the useless background part and improving image recognition rate, the detection accuracy is improved. For disease, we can use edge detection and detection method based on color to extract the shape of the disease parts in advance; for pests, we need to remove background and extract the pest images. This narrows the matching space, reduces the amount of calculation, improves the matching speed and reduces the storage space.

(2) Normalization of characteristic value. The dimensions of the various features are not the same. There is also a great difference in magnitude. Subsequent processing will have a bad influence on the performance of the system. Before using a particular algorithm, we need to normalize the raw data for eliminating the influence of magnitude and making the indicators comparable. It is convenient for subsequent processing and it can improve the calculation speed and efficiency.

(3) Feature extraction. According to the pest's classification and characteristics of the analysis, we select the appropriate algorithm and extract characteristic values. Extractions based on color, shape and combining feature can be used to find the correct method of high speed. For disease, retrieval method based on color and shape can be used. According to the characters of disease, it basically can be achieved. Compared with the pest, Disease characteristics is much simpler. Considering lookup speed, accuracy and recall ratio three factors, appropriate algorithms can be chose by Comprehensive comparison.

(4) Similarity computation, match the similar images, so match the types of diseases and pests. Appropriate calculation method can be chose according to different algorithms. Combined with database data, we can provide professional prevention measures and right guidance to the farmers.

Based on the above steps, research directions that image processing can be applied to fruit industry can be analyzed:

(1) Study new algorithm that quicker and more efficient. Not only to study the image feature extraction algorithm, must we study the image similarity matching algorithm. Dealing with the matching image fast and accurately and reducing the requirements and dependencies of the image quality.

(2) Reduce costs. Based on the function realization, we can reduce equipment costs and costs of achieving functional operation.

(3) Real time and networking. Realize perfect combination of orchard and the Internet and achieve the real-time detection of diseases and pests.

(4) Intelligence and digital. Combined with the plant diseases and pests' image detection and expert answers and diagnosis online, we can accurately distinguish diseases and pests and provide solution, prevention and control measures.

5 Conclusion

Image retrieval, of course, can't be only limited to the above methods. It is still an imperfect field now, waiting to explore in various aspects and make progress every day. Image retrieval method is also springing out endlessly, varying from situation to situation. Yet there also have lots of improvements on the existing image retrieval algorithms. The future development in fruit tree production is informatization, unification and digitization. For now, the image detection part of fruit tree management platform based on digital and information is still in an embryonic stage, the potential development and application prospect is huge. In consideration of the rapid development of the Internet and the Internet of Things, combined with the existing image detection technology match detection of diseases and pests in fruit trees can be basically realized. The present situation of low yield and poor quality can be gradually improved in our country. It is widely believed that with the continuous development of technology and the unremitting efforts of researchers' image detection on diseases and pests in fruit trees will become more mature. This will promote further development of China's fruit industry and motivate the process of digitalization, informatization and intelligentization.

References

1. Regional Economic Yearbook of China (2011)
2. Feng, G.-G., Qi, Y.-H., Xiao, Y.-B.: Review on Technology of Text-based Image Retrieval. China Academic Journal Electronic Publishing House (2007)
3. Neal Krawetz, Looks Like It
4. Jain, A.K., Vailaya, A.: Image retrieval using color and shape. *Pattern Recognition* 29(8), 1233–1244 (1997)
5. Lou, B.-F., Shao, L.-Q.: Content-Based static Image Retrieval. *Study of Library Science* 3, 84–86 (2006)
6. Swain, M.J., Ballard, D.H.: Indexing via color histograms. In: *Proc. ICCV*, pp. 390–393 (1990)
7. Swain, M.J., Ballard, D.H.: Color indexing. *IJCV* 7(1), 11–32 (1991)
8. Zabih, G.R., Miller, J.: Comparing images using color coherence vectors. In: *Proc. of ACM Intern. Conf. Multimedia*, Boston, MA (1996)
9. Wei, L.-M.: Review on Technology of Text-based Image Retrieval (2012)

10. Jeong, S., Won, C.S., Gray, R.M.: Image retrieval using color histograms generated by Gauss mixture vector quantization. *Computer Vision and Image Understanding* 94(1/3), 44–66 (2004)
11. Konstantinidis, K., Gasteratos, A., Andreadis, I.: Image retrieval based on
12. Zhang, S.-L., Ren, S.-P., Wang, X.-F.: Research status and development trend of Content-based image retrieval technology (2007)
13. Shi, B.-X., Zhang, M.-X., Qiao, X.-N., Hao, R.-Z.: Retrieval Technique of Image Based on Color Features. *Microelectronics & Computer* (4), 164–167 (2010)
14. Liu, Q.-Y.: Retrieval Technique of Image Based on shape features. *Journal of Information* 4 (2004)
15. Gu, J., Xie, J.: A Fast Image Matching Retrieval Method Using Sobel. Operator. *Computer Programming Skills and Maintenance* (2010)
16. Dong, W.-J., Zhou, M.-Q., Li, X., Geng, G.-H.: Research of the Edge Detection Technique Based on Wavelet Analysis. *Computer Engineering and Applications* (2004)
17. Sun, J.-D., Cui, J.-T., Wu, X.-S., Zhou, L.-H.: Color Image Retrieval Based on Color and Shape Features. *Journal of Image and Graphics* (2004)
18. Deng, C.-Q., Feng, G.: Content-based Image Retrieval Using Combination Features. *Computer Applications* 7 (2003)
19. Lowe, D.G.: Object Recognition from Local Scale-Invariant Features (1999)
20. Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints (2004)
21. Li, Z.-R.: Reserach On Identify Technologies Of Apple's Disease Based on Image Analysis. Northwest A & F University (2010)
22. Cai, Q.: Reserach of Vegetable Leaf-Eating Pests Based on Image Analysis. Northwest A & F University (2010)
23. Zhang, H., Mao, H., Qiu, D.: Feature extraction for the stored-grain insect detection system based on imagerecognition technology. *Transactions of the CSAE* 2, 134–138 (2009)