Regular paper

Journal of the KIECS. pp. 959-968, vol. 14, no. 5, Oct. 31. 2019, t. 97, pISSN 1975-8170 | eISSN 2288-2189
http://dx.doi.org/10.13067/JKIECS.2019.14.5.959

토마토 잎사귀 질병 감지를 위한 이미지 처리 메커니즘

박정현* · 이성근**

An Image Processing Mechanism for Disease Detection in Tomato Leaf

Jeong-Hyeon Park* · Sung-Keun Lee**

요 약

농업 분야에서 여러 가지 센서들과 임베디드 시스템을 활용하여 한 무선 센서 네트워크 기술이 적용되고 있는 추세이다. 특히, 센서 네트워크를 활용하여 작물의 질병을 조기에 진단할 수 있는 많은 연구가 진행되고 있다. 기존 병충해 진단 연구들은 실제 농가에 적용하기 어려운 부분이 존재한다. 본 논문은 이를 개선하고자 하였으며, 화상카메라를 통해 받아온 작물의 잎사귀 이미지를 분석하여 병충해를 초기에 감지 가능한 알고리즘을 제안한다. 실제 시설원예 및 노지 환경 농가의 캡쳐한 이미지 내에서 감염 의심 영역을 개선된 K 평균 클러스터링 기법을 통해 분류하였다. 그 후 엣지 검출, 엣지 추적 기법을 활용하여 해당 영역의 잎사귀 내부 존재 여부를 확인하였다. 인근 농가에서 촬영한 토마토 잎사귀 이미지를 이용하여 성능 평가를 수행하였다. 기존 논문의 방법 보다 제안 알고리즘의 감영 영역 분류 능력이 우수한 것으로 나타났다.

ABSTRACT

In the agricultural industry, wireless sensor network technology has being applied by utilizing various sensors and embedded systems. In particular, a lot of researches are being conducted to diagnose diseases of crops early by using sensor network. There are some difficulties on traditional research how to diagnose crop diseases is not practical for agriculture. This paper proposes the algorithm which enables to investigate and analyze the crop leaf image taken by image camera and detect the infected area within the image. We applied the enhanced k-means clustering method to the images captured at horticulture facility and categorized the areas in the image. Then we used the edge detection and edge tracking scheme to decide whether the extracted areas are located in inside of leaf or not. The performance was evaluated using the images capturing tomato leaves. The results of performance evaluation shows that the proposed algorithm outperforms the traditional algorithms in terms of classification capability.

키워드

Disease Detection Of Tomato, Leaf Disease, K Means Clustering, Image Processing
토마토 질병 감지, 잎사귀 질병, K 평균 클러스터링, 이미지 처리

* ㈜브이엠소프트 전임연구원(pjh6364@nate.com)
** 교신저자 : 순천대학교 멀티미디어공학과
· 접 수 일 : 2019. 08. 16
· 수정완료일 : 2019. 09. 15
· 게재확정일 : 2019. 10. 15
· Received : Aug. 16, 2019, Revised : Sep. 15, 2019, Accepted : Oct. 15, 2019
· Corresponding Author : Sung-Keun Lee
Dept. Multimedia Eng. Sunchon National University,
Email : sklee@scnu.ac.kr

Ⅰ. Introduction

This paper aims at providing more convenient and efficient agricultural environment by combining IT to agriculture. One of the biggest challenges in cropping is disease and insect pests. Because damage caused by disease and insect pests directly affect earnings of agricultural household, it is critically important to prevent this happening in advance and detect it at an early stage. Currently, disease detection is done manually. In this case, labor costs may be high and it takes much time. Furthermore, there is another problem that human sensory system has limitation in terms of detection capability. Owing to an evolution of sensor technology and a development of low cost and low altitude automatic aerial system recently, crops can be monitored continuously and sequentially[1][2]. Under this circumstance, there are numerous studies on a variety of sensor network for growing crops based on IT such as sensor network and image processing technology actively being conducted across agriculturally advanced countries such as the United States, Netherlands, and Spain[3][4][5].

This paper introduces an image processing-based algorithm which enables to investigate leaf images taken by camera sensor and detect disease and insect pests in advance. We utilizes an enhanced k-means clustering algorithm for detecting the infected areas in order to determine the occurrence of disease at leaf. The rest of this paper is organized as follows. Chapter 2 describes related works and chapter 3 proposes disease detection algorithm for crop leaf. Chapter 4 describes the procedure to receive the image and apply the proposed algorithm to the image. Finally chapter 5 concludes this study.

Ⅱ. Related Work

Diseases occurring at crops are normally broke out at leaf and accompanied with a variety of symptoms[6]. Leaf diseases of crops are determined by ratio of the infected areas of the leaf[7]. Therefore, infection of crop leaf can be determined by analyzing a leaf image. Crop leaf disease detection process is carried out in this order: obtaining image, pre-processing, extracting feature points, and identifying symptom of bacterial disease. Types of crop are divided into fruit crop, vegetable crop, commercial crop and major grain crop. Since the symptoms appear in different forms at various parts depending on type of infected diseases, the mechanism to be applied can be vary[5].

To get rid of noise from the obtained image, median filter is applied to the image or the image is converted to one with color-space suitable for processing the image. In addition to this, this procedure includes the process to apply log function to gray value or make image binary with use of visual characteristics such that the degree of recognizing light by human eye is similar to log function. The second step is to extract the interested areas by investigating the features of the parts to be detected.

There are several feature extraction methods: block-wise, Gray-Level Cooccurrence Matrix (GLCM), Discrete wavelet transform(DWT) and Principal Component Analysis(PCA). In the block-wise feature extraction method[8], noise is removed from the image through pre-processing. Then binary image is generated. The image is divided into a number of blocks with specific size. Each block is investigated and the number of pixels used for presenting the image is stored in vector sequentially. When vector is ready, feature extraction can be done by examining the vector. DWT and PCA are methods used for data compression, face recognition and fruit detection as well as feature extraction[9]. PCA is a method that can be used for data compression, face recognition and noise removal since it enables to identify property on how numerous data sets are distributed. If pixel data consisting of image are applied to vector, n-dimensional vector is generated where the size of image is n. We apply PCA to 100 images in n dimension and then sort the images in an increasing order of variance values. K-means clustering is a scheme to bind a set of similar data into k clusters[10-11]. As initial values, the number of clusters and the location of center are set. Then each data is categorized to the close cluster. Once all data are assigned to clusters, center of each cluster will be recalculated. The center part needs to be repeatedly recalculated until the center value is unchanged. If the algorithm is applied to the color image, color sets with similar colors are grouped together into one cluster. Thus this is good way to extract specific color.

Ⅲ. Image processing method to detect crop leaf disease

Image processing algorithm for detecting crop leaf disease is conducted as following sequence. It is necessary to prepare the target images collected around horticulture facility and outdoor environment taken by camera sensors. Various parts including leaves, stems, soils and fruits would appear in the images. The first step is to extract the parts which are suspicious of infection. All parts except for infection-suspicious part will be eliminated.

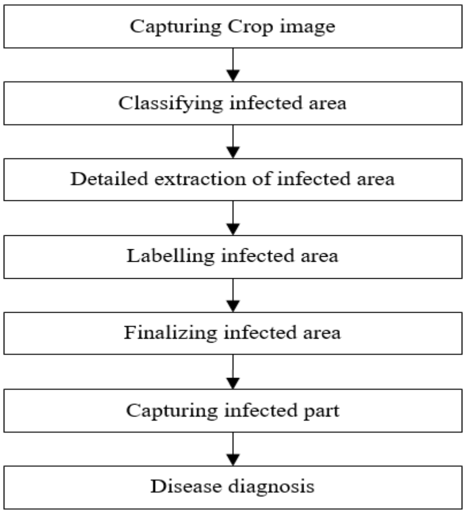

Fig. 1 Disease detection algorithm for crop leaf

Within the infection-suspicious part, we also need to remove the part which seems not inside of leaf and the part which is hardly identified what it is because it is too small. The parts which are suspicious to infection will be labelled, extended to the bigger rectangular area and combined with other labelled part which is considered to be adjacent. Finally, the original image and the rectangle for the suspicious part are synthesized. Cropping out the rectangle from the synthesized image and using the self-diagnosis service, we determine the type of disease appearing at the part. Figure 1 shows the detailed operation procedure of the algorithm proposed in this paper.

3.1 Enhanced K-means clustering algorithm

In this paper, we study an algorithm that can apply to both outdoor environment and horticulture facility and complement weak point that clustering is negatively affected by light. Once the scene is exposed to light, R, G, B values of the scene image get increased all together. Consequently, the image looks white. Instead of using R, G and B values individually, we carry out clustering process using R-B, R-G and G-B value for overcome this whiting problem. Even though the light increases all the values of R, G and B, the difference between two values is still unchanged. Except for the part that appears white or the part appears black due to shadow, all can be recognized as one cluster. At this time, the values of R-B, R-G, and G-B can be unique value representing the corresponding object, which will be used to determine whether it is the inside part of leaf or not. Prior to starting k-means clustering process, k value and the center value of K need to be set initially. Without specific background in an image, it is hard to decide the number of clusters which is enough to categorize the image. To do this, it is necessary to know information on the distribution of colors in the image. Each pixel existing in the image is categorized into one of seven color areas. Seven color areas consists of one achromatic color area and six color areas which are areas depending on R,G,B value orders such as R>G>B, R>B>G, G>R>B, G>B>R, B>G>R, and B>R>G. If the highest and the lowest value of R, G, B are not much different, the pixel is mapped into the achromatic color area. Once all pixels are completed to be categorized, the color areas are sorted in an ascending order of the total number of pixels belonging to the corresponding area. Only the area with its percentage higher than a specific threshold can be determined to become a cluster. Then K value is set for those areas with higher percentage than threshold. The center value of K is designated to be a random pixel existing at the corresponding color area. K-value setting through the way of categorizing pixels into color areas is feasible because outdoor environment or horticulture facility has characteristics having limited color area.

3.2 Classifying an infected area

After K-means clustering process is completed, the image is divided into several clusters. In this study, we focus on detecting tomato leaf diseases such as tomato leaf mold disease and tomato yellow leaf curl virus(TYLCV) disease so we set the detection color to be brown. Detection process consists of the step to select the cluster which contains the infected area and the step to extract the more specific detection area within the selected cluster. Within the selected cluster, there exist three different sub-areas: (a) infected area, (b) area with similar colors to the infected area, and (c) unclassified area due to ambiguous colors. Among those areas, the unclassified area with ambiguous colors contains noise values which implies that the part lose its nature color by being exposed to light rather than showing specific object. Thus, the ambiguous colored area needs to be erased by applying median-filter which is generally used for filtering out the noise area. Next, Areas containing similar colors to the infected area needs to be eliminated. Since our aim is to detect leaf disease, it is necessary to remove the parts which are determined to be outside of the leaf. First step is to prepare several crop leaf images and then calculate average value for R-G, G-B, and R-B values for every pixel. This value can be called unique color value for leaf. Pixels surrounding the anticipated infected area and unique color value of the crop leaf are compared. Margin of error with unique color value needs to be set properly through experiments. If they are 100% identical, this area is determined to exist in the leaf. Sometimes, the infected area could spread around the edge of the leaf not right inside of leaf. In this case, more complicated detection process is additionally required. If only some parts of the surrounding pixels are determined to be inside of leaf and the other neighboring pixels are clearly divided by an edge, we need to track the dividing edge. Though several edge detection methods exist, Canny edge detection algorithm is suitable for this case since leaf edge should be detected in a linear form with simple and clear line.

Fig. 2 Extraction of infection area at leaf edge

Once an edge is extracted, the direction going further of the edge will be explored using the edge tracking method. In this direction, pixels in an area which is suspicious to the infection are compared to the unique color of leaf. If it is the unique color, it means that the area can be considered to be an infection area at edge. In this case, therefore, it should not be cut off. All other cases than this are considered to be non-infected area, so that it should be removed. Figure 2 illustrates the classification of infection area at leaf edge.

3.3 Infected area extraction through labelling

Upon completion of extracting infected area, only areas which are expected to be infected remain in the cluster. At this stage, other than these areas will be colored by black. Next is labelling process which is for splitting the infected area separately. Labelling is done by making label at each group of pixels which is encountered by scanning the whole image sequentially. Labelling allows us to distinguish the order between areas and identify orientations of the edge such as up, down, left and right for the area. Area around edge needs small margin rather than cutting off sharply so that the infected area can be correctly detected. While searching location of the infected areas sequentially, two infected areas which seem to be located adjacently are combined by means of their labels. Then the edge locations are reconfigured and researched. Once the searching process finishes, the adjacent infected areas will be all combined into bigger one so that we can easily look and determine if it is infection or not.

3.4 Disease self-diagnosis service

Once we get the extracted image for the infected area, it is necessary to decide what disease it is. There are numerous ways to do this. One way is to compare the extracted image to the standard images that can be obtained by calling disease self-diagnosis service API which is offered by Rural Development Administration (RDA). Another way is to use Lucene-based disease diagnosis service. If we pass the extracted image to the service, it will automatically list up the disease and insect pest images with similar shape or color structure.

Ⅳ. Performance evaluation of the proposed mechanism

In this chapter, we validate the performance of the proposed crop leaf disease detection image processing scheme by evaluating it using real images. Scripts for performance evaluation are implemented in C++ and Java on Visual Studio 2010 and Eclipse using OpenCV. The target images were taken at tomato farm near Suncheon Bay with focus on the leaves outstandingly infected. Figure 3 is the original image.

Fig. 3 Tomato leaf original image

4.1 Implementation and execution of K-means clustering

Though k-means clustering function is provided in OpenCV, we implemented it in Java language on Eclipse environment without using the function in OpenCV in order to set pixel value in the clustering process. Firstly, we loaded the raw image file which was saved in Photoshop. Then we move R, G, B values for each pixel to array. While the values are stored to array, R-G, G-B and R-B values are also computed and stored at the same time as a preparation for applying the proposed mechanism. Once the image is ready, K value and center value are set. Then every pixel is compared to K and assigned to the closet K cluster. In this paper, it is done by comparing the distance between two points. Secondly, the center value of K is updated with the center value of the pixels which are assigned to the cluster. The first and the second steps are repeated until the value is unchanged.

Figure 4 is the image as a result of k-means clustering. K-value is set to 3 according to the area analysis result. Upon completion of clustering process, the number of clusters classified is identical to K. Among the clusters, only the cluster images corresponding to the infection area are selected. To select the cluster image corresponding to the infection area, color groups of the target leaf disease to be found are necessary beforehand. Based on these color groups, the third cluster was selected in this experiment.

Fig. 4 K-means clustering result image(K=3)

4.2 Elimination of non-infected area

The infected area cluster in figure 4 indicates that unnecessary parts also exist aside from the infected area. When using k-means clustering, it is inevitable that pixels that have relative color difference are included in the same cluster. Therefore, it is necessary to remove this uninfected area from the cluster in several steps. First, it needs to be done by calculating R,G,B ratio of leaf crop disease color group. Second, median filter is applied to the cluster in order to filter out small noise.

Third, too tiny area needs to be eliminated since it is too small to identify whether it is infection or not. These processes are carried out by running Java-coded program. For doing this, mask was created and images were retrieved sequentially. The size of mask is 6x6 so as to get rid of the area with 20 pixels or less. In this process, the area where the infected pixels reside in inside of the mask and outside of the pixels are black is to be eliminated. To investigate the inside of the leaf, adjacent pixels are compared to unique value of leaf. In case that adjacent pixels are all different from the unique value of leaf, leaf edge investigation process is proceeded. Edge extraction is carried out using Canny edge method. Then the path following the edge is searched.

Fig. 5 Image after eliminating non-infected area.

Based on the edge, pixels at the direction of infected area are compared to those at opposite direction. If pixels at the direction of infected area are identical to the unique value of leaf and pixels at opposite direction are not the same, the pixels are considered to be inside of leaf. Figure 5 shows the result image.

4.3 Extraction of infected area

Next step is to pick the infected areas only and to combine those areas into one bigger object through labelling. Labelling is done by searching the pixel in the image sequentially from the first pixel. If the infected area is encountered while searching the pixel, adjacent pixels of the pixel are retrieved one by one at up, down, left and right direction. If they are not black, the pixel is labelled. Once the labelling finishes at one area, searching is resumed. This labelling process is repeated until all the infected areas get labelled with individual number. These numbers allow us to find out where the boundary pixels facing at four directions, up, down, left and right are within the infected area. Using the information on location, adjacent areas among the infected areas can be combined into bigger areas.

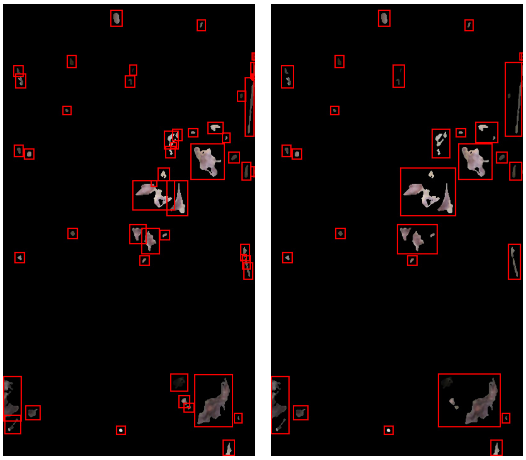

Fig. 6 Image resulting from labelling and combination of areas.

Since the adjacent infected areas are highly likely existing at the same leaf, combination of those adjacent areas enables easily to detect leaf disease and insect pests. Figure 6 indicates that the infected area becomes more simplified and turns to have easier look.

4.4 Applying original image

Figure 7 is the original image on which rectangle is drawn. This image indicates that the proposed method could find out significantly small infected area which seems hard to spot with person's eyes. However, it turns out that the proposed mechanism made wrong decision that some other objects such as soil and crop supports are also the infected area. Even when the other objects rather than crop leaf are seen in between leaves, the neighboring pixels of those objects are considered to be leaf in the proposed algorithm so the objects are decided to be infected part. This problem is required to be resolved with additional algorithm in further study.

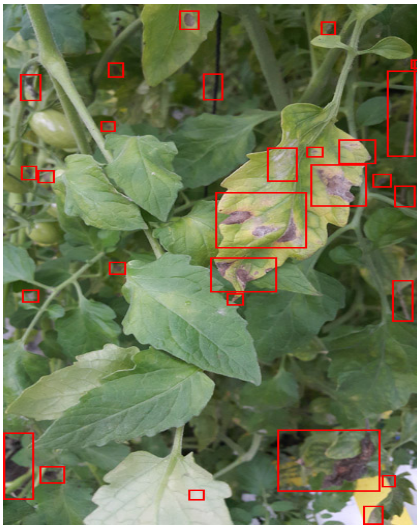

Fig. 7 Image resulting from completion of extraction of infected area

Ⅴ. Conclusion

In growing crop, it is critically important to diagnose the disease and insect pests at an early stage and treat it in timely manner. Highly contagious disease is particularly critical since it can directly damage to agricultural household if it fails to be discovered at an early stage. In spite of this, manual observation and investigation to the crops across the huge size of area is hardly possible because it requires massive time and efforts. In this paper, we conducted an early-stage study on crop leaf disease diagnosis system using leaf image sent by camera sensor widely deployed in sensor network environment. We used an enhanced k-means clustering method to detect the infected area, and an edge detecting and tracking method to determine if the pixel is residing in inside of leaf or not. Additionally, labelling, removal of unnecessary area and combination are carried out to detect the infected area in more detail. The proposed algorithm was implemented using C++ and Java. Performance of the proposed method was evaluated using the tomato leaf images taken at farms in neighboring area. The results from the evaluation indicate that the proposed algorithm outperforms the traditional k-means clustering algorithm in detecting the infected area. In addition, infected area detection function was successfully verified by means of determining whether it is inside of leaf and labelling. However, the proposed method also has limitation that it cannot be generally used because it requires specific color representing infection area which will vary at each case. For general use of this algorithm, this should be complemented in future. Another part necessary to improve is implementation for enabling more accurate classification of infected and uninfected area. Future study will be to improve this part and make it more general so as to use it in any crop not for just specific crops.

Acknowledgements

This research was supported by the MSIP(Ministry of Science, ICT and Future Planning), Korea, under the ITRC(Information Technology Research Center) support program (IITP-2018-2013-1-00877) supervised by the IITP.

References

[1] J. Park, S. Lee, and W. Oh, "Congestion control mechanism for efficient network environment in WMSN," *J. of the Korea Institute of Electronic Communication Sciences,*

vol. 10, no. 2, Feb. 2015, pp. 289 - 295.

[2] Y. Choi and S. Joung, "A Design and Implementation of Mobile based Smart Green House System," *J. of the Korea Institute of Electronic Communication Sciences*, vol. 9, no. 4, Apr. 2014, pp. 475 - 482.

[3] S. Sankaran, A. Mishra, and C. Davis, "A review of advanced techniques for detecting plant diseases," *Computers and Electronics in Agriculture*, vol. 72, no. 1, Feb. 2010, pp. 1 - 13.

[4] A. Camargo and J. S. Smith, "An image processing based algorithm to automatically identify plant disease visual symptoms," *Biosystems Engineering*, vol. 102, no. 1, Jan. 2009, pp. 9 - 21.

[5] J. D. Pujari, R. Yakkundimath, and A. S. Byadgi, "Image Processing Based Detection of Fungal Diseases in Plants," *Procedia Computer Science*, vol. 46, no. 4, Aug. 2015, pp. 1802 - 1808.

[6] A. N. Rathod, B. Tanawal, and V. Shah, "Image Processing Techniques for Detection of Leaf Disease," *International J. of Advanced Research in Computer Science and Software Engineering*, vol. 3, no. 11, Nov. 2013, pp. 397 - 399.

[7] S. Weizheng, W. Yachun, C. Zhanliang, and W. Hongda, "Grading method of leaf spot disease based on image," In *Proc. Computer Science and Software Engineering*, Wuhan, Hubei, China, 2008, pp. 491 - 494.

[8] J. D. Pujari, R. S. Yakkundimath, and A. S. Byadgi, "Statistical methods for quantitatively detecting fungal disease from fruits' images," *International J. of Intelligent Systems and Applications in Engineering*, vol. 1, no. 4, Aug. 2013, pp. 60 - 67.

[9] B. J. Woodford, N. K. Kasabov, and C. H. Wearing, "Fruit image analysis using wavelets," In *Proc. the Iconip/Anziis/Annes*, Perth, Austrakia, 1999. pp. 88 - 91.

[10] S. A. Ali, N. Sulaiman, A. Mustapha, A and N. Mustapha, "K-Means Clustering to Improve the Accuracy of Decision Tree Response Classification," *Information technology journal*, vol. 8, no, 8, Aug. 2009, pp. 182-191.

[11] S. Lee, "Comparison of Initial Seeds Methods for K-Means Clustering," *J. of Internet Computing and Services*, vol. 13, no. 6, June 2012, pp. 1 - 8.

저자 소개

박정현(Jeong-Hyeon Park)

2014년 순천대학교 멀티미디어공학과 졸업(공학사)
2017년 순천대학교 대학원 멀티미디어공학과 졸업(공학석사)
1997년 ~ 현재 ㈜브이엠소프트 전임연구원
※ 관심분야 : WSN, 영상처리, 멀티미디어 통신, 인터넷 QoS

이성근(Sung-Keun Lee)
(교신저자)

1985년 고려대학교 전자공학과 졸업(공학사)
1987년 고려대학교 대학원 전자공학과 졸업(공학석사)
1995년 고려대학교 대학원 전자공학과 졸업(공학박사)
2017년 ~ 2018년 : Georgia Institute of Technology ECE 방문교수
1997년 ~ 현재 순천대학교 멀티미디어공학과 교수
※ 관심분야 : WSN, 스마트농업, 멀티미디어통신, 강화학습 기반 QoS 보장 기술