BIOSYSTEMS ENGINEERING 144 (2016) 52–60

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/issn/15375110

Review

A review on the main challenges in automatic plant disease identification based on visible range images

Jayme Garcia Arnal Barbedo *

Embrapa Agricultural Informatics, Av. André Tosello, 209 – C.P. 6041, Campinas, SP, 13083-886, Brazil

ARTICLE INFO

Article history:
Received 11 September 2015
Received in revised form
26 January 2016
Accepted 29 January 2016
Published online 21 February 2016

Keywords:
Plant diseases
Automatic identification
Visible symptoms
Digital image processing

The problem associated with automatic plant disease identification using visible range images has received considerable attention in the last two decades, however the techniques proposed so far are usually limited in their scope and dependent on ideal capture conditions in order to work properly. This apparent lack of significant advancements may be partially explained by some difficult challenges posed by the subject: presence of complex backgrounds that cannot be easily separated from the region of interest (usually leaf and stem), boundaries of the symptoms often are not well defined, uncontrolled capture conditions may present characteristics that make the image analysis more difficult, certain diseases produce symptoms with a wide range of characteristics, the symptoms produced by different diseases may be very similar, and they may be present simultaneously. This paper provides an analysis of each one of those challenges, emphasizing both the problems that they may cause and how they may have potentially affected the techniques proposed in the past. Some possible solutions capable of overcoming at least some of those challenges are proposed.

1. Introduction

Plant disease identification is one of the most basic and important activities in agriculture. In most cases, identification is performed manually, either visually or by microscopy. The problem with visual assessment is that, being a subjective task, it is prone to psychological and cognitive phenomena that may lead to bias, optical illusions and, ultimately, to error. On the other hand, laboratorial analyses such as molecular, immunological or pathogen culturing-based approaches are often time consuming, failing to provide answers in a timely manner. In this context, it is compelling to develop automatic methods capable of identifying diseases in a rapid and reliable way. The vast majority of automatic methods proposed so far rely on digital images, which allows the use of very fast techniques. However, intrinsic and extrinsic factors mean these methods remain too error prone, which was the motivation for the current review.

Most of the methods described in the literature are based on digital images of symptoms in the visible and near-infrared bands (Barbedo, 2013), with those bands being considered in isolation or represented in multi and hyperspectral images. Although multi and hyperspectral images can potentially

* Tel.: +55 19 3211 5880.
E-mail address: jayme.barbedo@embrapa.br.
http://dx.doi.org/10.1016/j.biosystemseng.2016.01.017

Nomenclature

IR	Infrared
LED	Light Emitting Diode
VR	Visible Range

carry more information than normal photographs, they are usually captured by expensive and bulky sensors, while conventional cameras are ubiquitous and present in many consumer-level electronics stores. This has resulted in developing systems based on the visible range, which also leads to a more focused discussion. More information on multi and hyperspectral imaging applied to plant pathology can be found in Sankaran, Mishra, Ehsani, and Davis (2010) and Bock, Poole, Parker, and Gottwald (2010).

Some of the methods exploring visible range images focus on detecting a single disease of interest amidst other diseases, healthy tissue, nutritional problems and pests (Barbedo, Tibola, & Fernandes, 2015; Oberti et al., 2014; Polder, van der Heijden, van Doorn, & Baltissen, 2014; Pourreza, Lee, Ehsani, Schueller, & Raveh, 2015, Pourreza, Lee, Etxeberria, & Banerjee, 2015; Zhang, Yuan, Pu, Loraamm, Yang, & Wang, 2014; Zhou, Kaneko, Tanaka, Kayamori, & Shimizu, 2014), while others try to detect and discriminate different diseases. Although progress has been made regarding the disease classification problem, the vast majority of the methods are only capable of discriminating among a small number of diseases (Phadikar, Sil, & Das, 2013; Pydipati, Burks, & Lee, 2006; Sanyal & Patel, 2008). In general, this is too limited for real-world applications, because the number of pathogens that can simultaneously infect a plant and cause disease symptoms is usually higher. Also, nutritional deficiencies (Pagola et al., 2009; Romualdo et al., 2014; Wiwart, Fordonski, Zuk-Golaszewska, & Suchowilska, 2009) and pests (Clément, Verfaille, Lormel, & Jaloux, 2015; Koumpouros et al., 2004; Škaloudová, Křivan, & Zemek, 2006) may produce symptoms that mimic very closely the characteristics of some diseases. To make matters even more complicated, there are some challenges that affect virtually all studies devoted to the automation of the disease diagnosis process and that have not yet been properly investigated. The objective of this review was to identify some of the most important of those challenges, to explore in depth their causes and their impact on the performance of the techniques proposed so far, and to propose some possible avenues to be explored in order to mitigate or eliminate their adverse effects. The challenges selected as the most impactful were the following:

- The background often contains elements that can make it very difficult to correctly segment the region of interest where the symptoms are manifest.
- Capture conditions are difficult to control, which may cause the images to present characteristics that are difficult to predict and make the disease identification more challenging.
- Most symptoms do not have well defined boundaries, rather gradually fading into normal tissue, making it difficult to clearly define which are the healthy and diseased regions.
- A given disease may possess very different characteristics depending on its stage of development, and sometimes on where it is located on the plant.
- Symptoms produced by different diseases may be present simultaneously, manifesting either physically separated or combined into a "hybrid" symptom that may be difficult to identify.
- Symptoms produced by different diseases may be visually similar, which forces the methods to rely on very tenuous differences to discriminate among them.

The first two challenges can be viewed as extrinsic factors, while the remaining four are intrinsic to the problem.

It is important to highlight that, although the focus is on the identification of plant diseases, the challenges discussed here are also relevant for disease severity measurement, with some references on the matter also being included. In fact, the only major difference is that accurately outlining the symptoms is not necessarily critical for disease identification, while it is paramount for severity measurement. All other challenges cited above have roughly the same importance for both issues, especially considering that the disease identification may be a necessary intermediate step for the severity measurement, particularly when multiple diseases are expected to coexist.

Because of space limitations, technical details about the methods, including the software used, were omitted, but a comprehensive discussion on the matter can be found in Barbedo (2013). The discussions presented were based on research reported primarily for leaf symptoms (by far the most explored), however they are, for the most part, valid for other plant parts, including stem, fruit and flowers.

2. Extrinsic factors

2.1. *Image background*

Leaf segmentation is the first step of most image-based tools for leaf analysis. If some kind of panel (preferably white or blue) is placed behind the leaf, this task can usually be performed automatically without much problem. On the other hand, if the background contains plants, leaves, soil and other elements, the segmentation may be a challenge.

Segmenting the leaf is particularly difficult when the background has a significant amount of green elements, for example as in Fig. 1. The segmentation of leaves from busy backgrounds is a problem that has received some attention: Zhang and Meng (2011) directly separated the lesions from leaf and background using a two-step hierarchical matching procedure; Alenyà, Dellen, Foix, and Torras (2013) used depth information for localizing the leaves and extracting them from the rest of the image; while Wang, He, Han, Ouyang, and Li (2013) used the so-called marker controlled watershed segmentation, which is based on the selection of certain local minima from the image's gradient as control markers, for separating leaves from the rest of the image.