Computers and Electronics in Agriculture 110 (2015) 70–77

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

In-field crop row phenotyping from 3D modeling performed using Structure from Motion

Sylvain Jay *, Gilles Rabatel, Xavier Hadoux, Daniel Moura, Nathalie Gorretta

UMR ITAP, Irstea, 361, rue J.F. Breton, B.P. 5095, 34196 Montpellier Cedex 5, France

ARTICLE INFO

Article history:
Received 5 March 2014
Received in revised form 27 July 2014
Accepted 16 September 2014

Keywords:
Leaf area estimation
Phenotyping
Plant 3D modeling
Plant height estimation
Plant/background discrimination
Structure from Motion

ABSTRACT

This article presents a method for crop row structure characterization that is adapted to phenotyping-related issues. In the proposed method, a crop row 3D model is built and serves as a basis for retrieving plant structural parameters. This model is computed using Structure from Motion with RGB images acquired by translating a single camera along the row. Then, to estimate plant height and leaf area, plant and background are discriminated by a robust method that uses both color and height information in order to handle low-contrasted regions. The 3D model is scaled and the plant surface is finally approximated using a triangular mesh.

The efficacy of our method was assessed with two data sets collected under outdoor conditions. We also evaluated its robustness against various plant structures, sensors, acquisition techniques and lighting conditions. The crop row 3D models were accurate and led to satisfactory height estimation results, since both the average error and reference measurement error were similar. Strong correlations and low errors were also obtained for leaf area estimation. Thanks to its ease of use, estimation accuracy and robustness under outdoor conditions, our method provides an operational tool for phenotyping applications.

1. Introduction

In a context of a greener and more competitive agriculture, the creation and selection of varieties consuming less water, nitrogen or even pesticides, have become a topic of major interest. Phenotyping methods are therefore needed in order to compare varieties and to relate genotypes to phenotypes. In particular, these methods aim at retrieving numerous parameters that characterize the crop row structure, such as plant height or leaf area. They must be non-destructive since these measurements have to be carried out all along the plant life. They must also be fast and automatic so that a maximum of data may be processed with a minimum of user interactions.

Plant structural parameters can be accurately retrieved from a three-dimensional (3D) model that can be computed in several ways. Active systems involving a light source have been used, e.g. with LiDAR (Rosell et al., 2009; Weiss and Biber, 2011) or the depth-imaging system Kinect Microsoft© (Chéné et al., 2012). On the other hand, stereovision-based methods are passive and provide both the 3D structure estimate and color if an appropriate RGB camera is used. Because they can be implemented with commercial cameras, such methods have been widely used for agricultural applications (Ivanov et al., 1995; Andersen et al., 2005; Rovira-Más et al., 2008).

Many authors have implemented stereovision to build plant 3D models using two view angles. These are often provided by two cameras that are separated by a fixed baseline distance and that need prior laboratory calibration. However, these methods do not always fulfill phenotyping-related constraints. For example, in their original implementation, some of them cannot be easily implemented in the field (Shrestha et al., 2003) or are dedicated to a single variety and thus not applicable for varietal selection (He et al., 2002). Only a few methods can be used for such applications. For example, Kise and Zhang (2008) have developed a system using one tractor-mounted stereocamera for 3D crop row structure mapping. Their system estimates the height and position of crop rows in order to reliably guide the tractor. Leemans et al. (2013) have also used images acquired by a stereocamera for leaf area index retrieval in wheat crops. Recently, Lati et al. (2013) have developed a novel approach for the estimation of plant height, leaf cover area and biomass volume, which can even handle occluded areas to some extent (i.e., overlapping leaves).

On the other hand, using Structure from Motion (SfM) to retrieve the crop row 3D structure is easier than using stereovision

* Corresponding author. Tel.: +33 4 67 16 64 59.
E-mail address: sylvain.jay@irstea.fr (S. Jay).

http://dx.doi.org/10.1016/j.compag.2014.09.021
0168-1699/

with several cameras. Indeed, as explained in Section 2.1, SfM is implemented with only a single camera acquiring images while moving along the row. Furthermore, since camera intrinsic parameters are automatically estimated during the reconstruction process, SfM does not need any prior calibration and is therefore a convenient tool for in-field phenotyping applications. It has been used for plant 3D modeling (Santos and de Oliveira, 2012), but its potential for plant structural parameter retrieval has not been explored much yet.

Retrieving plant structural parameters from crop row 3D models often necessitates to discriminate plant from background. Usual discrimination methods are color-based and involve thresholding a vegetation index map computed from RGB bands. Many of these vegetation indices have been reviewed by Meyer and Neto (2008). One of the most widely used is the Excess Green Index (ExG) first introduced by Woebbecke et al. (1995). However, even though this index is normalized and thus insensitive to the light source intensity and leaf inclination (Gée et al., 2008), it is affected by shadows. Indeed, in shaded areas, the light mainly originates from multiple scattering caused by the nearby environment (Vigneau, 2010). Therefore, the resulting light variations cannot be accurately modeled with a single multiplicative factor. Recently, Lati et al. (2013) have discriminated plant from background in a more appropriate two-dimensional color space also based on RGB bands. They have used a hue-invariant transformation that takes into account spectral properties of natural light. However, this method is not automatic because for each situation, several parameters that depend on the sensor and contrast between plant and background have to be set beforehand. In addition, like other color-based discrimination methods, this approach fails if this contrast is low.

In this study, we propose a 3D modeling based method to characterize crop rows under outdoor conditions, i.e., to retrieve their structure, plant height and leaf area. Phenotyping-related needs (non-destructive, automatic, fast) are considered, and special attention is given to the robustness against heterogeneous lighting and low-contrasted regions. First, we perform 3D reconstruction using SfM and second, we retrieve the plant structural parameters from the crop row 3D model.

This paper is organized as follows. Section 2 presents the proposed method. The data sets and implementation details are described in Section 3. Section 4 shows the results of 3D reconstruction, discrimination and retrieval of plant structural parameters. Lastly, Section 5 draws conclusions and provides some guidelines for further work.

2. Proposed method

The framework of the proposed method is illustrated in Fig. 1. The first part deals with the 3D reconstruction performed using SfM, while in the second part, various processing steps are applied to the 3D model in order to retrieve the plant structural parameters. Both parts are described in the two following sections.

2.1. 3D reconstruction using Structure from Motion

Stereovision enables an object 3D structure to be retrieved from 2D images acquired from different view angles. Using SfM, the view angles are obtained by moving a single camera around the object of interest. The motion is either free or forced by a translation stage as illustrated in Fig. 2. In this latter case, the camera is translated at a given speed and acquires an image every time interval Δt. If Δt is small enough, images are overlapping so every point is seen from various view angles. Techniques based on ray intersection then allows us to retrieve the 3D position of such a point as well as the associated camera positions and orientations (Kraus, 2007).

In this method, we implemented SfM with Micmac (Ign, 2013), a digital surface model freeware developed by the National Institute of Geographic and Forest Information. In the following, we briefly describe the three usual main steps of SfM and their Micmac implementation. For more information, the reader can refer to the associated documentation (Ign, 2013). These steps are also illustrated on a real example in Fig. 3.

As shown in Fig. 3a, the first step consists in identifying points that are seen in several images. For this purpose, Micmac implements the Scale-invariant feature transform (SIFT) (Lowe, 1999), which has been designed to provide feature descriptors that are invariant to scale, rotation, translation and exposure. Then, basically, for every couple of images, a feature point in the first image is matched with the most similar feature point in the second image (in terms of Euclidean distance). This step thus allows many points of the scene to be detected in different images.

In the second step, the 3D position of every matched feature point is estimated together with the camera internal calibration and its position and orientation for every image. To do so, Micmac implements an iterative process in which, at the iteration N, ray intersection and matched feature points are used to orientate the Nth image with respect to the $(N-1)$ previous ones (the first image being arbitrarily orientated). A distortion model (usually radial) with several degrees of freedom (one for the focal length, two for the position of the principal point, and others for distortion coefficients) is used to estimate the camera internal calibration from the image set. At the end of each iteration, a bundle adjustment (Hartley and Zisserman, 2004) is also performed in order to limit error accumulation. A sparse point cloud describing the positions of feature points is therefore obtained as illustrated in Fig. 3b.

Lastly, in the third step, the estimated camera orientations and positions are used to retrieve the 3D positions of non feature point pixels, thus generating a dense point cloud. In Micmac, a matching process using normalized cross-correlation is implemented. For a given pair of overlapping images, a pixel in the first image is associated with the pixel that is on the epipolar line in the second

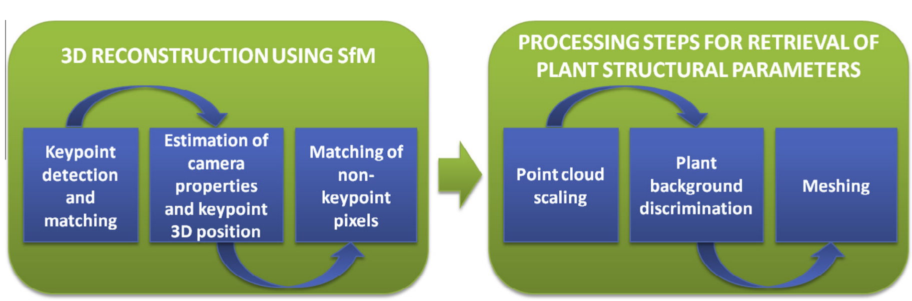

Fig. 1. Framework of the proposed method.