Computers and Electronics in Agriculture 148 (2018) 37–44

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Original papers

Improved image processing-based crop detection using Kalman filtering and the Hungarian algorithm

Esmael Hamuda*, Brian Mc Ginley, Martin Glavin, Edward Jones

National University of Ireland, University Road, Galway, Ireland

ARTICLE INFO

Keywords:
Image processing
Smart weeding
Tracking algorithm
Kalman filter
Hungarian algorithm
Crop detection

ABSTRACT

There is increasing interest in the use of image processing techniques for crop detection in intelligent weeding applications. An effective system for crop detection requires a high degree of adaptability to challenging circumstances such as different weather conditions and image capture conditions (vibration, variations in speed, etc.). To achieve the goal of a robust crop detection system, we have extended a previously-developed detection algorithm that is based on a combination of color-space and shape analysis, through the addition of object tracking. While the previous algorithm performed well in general, performance in sunny conditions was not as robust, opening up the possibility of improvement. The tracking algorithm consists of two steps. Firstly, we apply Kalman filtering to predict the new position of an object (a cauliflower plant in this case) in video sequences. Secondly, we use a data association algorithm (the Hungarian algorithm) to assign each detected crop that appears in each image to the correct crop trajectory. The recall matrix was used to evaluate the detection and tracking performance. With the help of tracking algorithm, detection failures were reduced, especially in sunny conditions, such that overall detection performance was raised from 97.28 to 99.3404%.

1. Introduction

Computer vision technology is starting to play a crucial role in several agricultural applications such as weed control, crop fertilization, plant species recognition and detection, growing phase determination, plant disease detection, and harvesting fruits. In general, computer vision is used to detect specific plants in a sequence of video, based on a combination of feature extraction and classification. Effective feature extraction and region detection in image can help to reduce the superfluous keypoints and improve the efficiency of the system (computational time) (Jauregi et al., 2009). Moreover, Soliman (2013) reported that the real challenge for feature detection and posterior image matching is to achieve robust feature detection with the following characteristics: (i) consistency: positions of detected objects should be insensitive to noise, scale, orientation, clutter, illumination; (ii) accuracy: objects should be detected as close as possible to the correct positions and features; (iii) speed: should be sufficiently fast.

Furthermore, challenging real-world conditions such as weather variability, presence of shadows in sunny conditions, natural similarities between the target object (weed or crop) and the background, affect the performance of machine vision techniques (Slaughter et al., 2008; Hamuda et al., 2016).

One of the challenges that most affects the plant based segmentation accuracy is sunny conditions because this leads to similarities in color between crops and weeds. For example, when the sunshine is strong, the surface of some leaf types (such as cauliflower or corn leaf), acts as a mirror (specular reflection); as a result, it may be segmented into the wrong category (Hamuda et al., 2016). Dealing with a moving camera may generate issues such as vibration, speed variations, and crops inadvertently covered by soil caused by moving machinery. These problems translate into loss of detection of moving objects between frames, and may result in e.g. inaccurate operation of automatic weeding or spraying machinery that relies on accurate detection of a plant across multiple frames. This paper addresses the problem of reducing detection errors in sunny conditions in particular, through the application of multiple-object tracking and data association.

The organization of the rest of this paper is as follows. The related work is given in Section II. The proposed method is detailed in Section 3. Section 4 describes the testing framework and performance metrics used. Section 5 discusses the results and finally, Conclusions are given in Section 6.

2. Related work

Generally, object tracking is a difficult task to achieve, particularly in uncontrolled conditions including: object motion, changing

* Corresponding author.
E-mail address: e.hamuda2@nuigalway.ie (E. Hamuda).

https://doi.org/10.1016/j.compag.2018.02.027
Received 18 June 2017; Received in revised form 21 February 2018; Accepted 25 February 2018

appearance patterns of both the object and the background, non-ridged object structures, occlusions, and camera motion (Yilmaz et al., 2006). Tracking multiple objects in video sequences has received a lot of attention in many applications such as surveillance, transportations, military, navigation, games, etc. Some tracking applications such as surveillance cameras track a person moving against a static background; this approach has also been used for tracking of fruit in sequential images of mango trees (Stein et al., 2016).

Stein et al. used a multi-view approach using LiDAR to detect, track, count and locate mango fruit. The detection process was done based on a state-of-the-art R-CNN detector, and pair-wise correspondences are established between images using trajectory data provided by a navigation system. The Hungarian algorithm and epipolar geometry were applied to perform the tracking of detected objects between frames. Results showed that the system was able to count fruit with high precision and accuracy (1.36% error compared to ground-truth). However, using a multi-view approach may increase complexity of the system, whereas it is desirable to achieve good performance with a single view system and hence achieve computational simplicity.

Tracking multiple objects in a dynamic scene adds a further level of complexity in computer vision applications. This problem increases in complexity when there are similarities between the different target objects (Pathan et al., 2009; Fan et al., 2016), as might be expected in e.g. crop detection applications. As with single object tracking, numerous approaches have been proposed to address these problems and have demonstrated good performance. According to the comprehensive survey presented in (Yilmaz et al., 2006), there are three common approaches that are used for tracking multiple objects: point tracking, kernel tracking and silhouette tracking.

The Kernel-based tracking approach is robust to occlusion, clutter, and distraction, however, some spatial information in the target is lost, and the method cannot perform well where the target object and its background have similar color. The Silhouette-based Tracking Approach is less sensitive to appearance variations, however, it requires training, particularly in shape matching. With respect to point-based tracking, this approach is simple and very useful to track small objects, which can be applied to track objects like a cauliflower that occupies a relatively small region in video sequences. The primary limitations of using point tracking occur in two scenarios: the occurrence of occluded plants where the growth stage is quite late, which is not considered here and where the false detection rate is high. The previously developed detection algorithm used here has relatively good performance in sunny conditions (98.04%), so false detections are not a significant issue here. Based on this, in this paper, we use the point tracking approach. Within the class of point tracking methods, there are three common approaches in the literature: Kalman filtering, particle filtering, and Multiple Hypothesis Tracking (MHT) (Parekh et al., 2014). According to Marron et al. (2007) the Kalman filter is a good option for environments where the number of objects is reasonably small, e.g. on the order of 5 objects. The Kalman filter is widely used because it has excellent properties including low computation, and the use of a sophisticated recursive and optimal estimator for one-dimensional linear systems with Gaussian error statistics (Anderson and Moore, 2005). Other point based tracking algorithms (Particle filter and Multiple Hypothesis Tracking) are computationally more expensive than Kalman filter (Orlande et al., 2011; Kim et al., 2015), which makes the Kaman Filter more attractive for real time processing.

In general, Kalman filtering used for multiple object tracking consists of two stages: the first using the Kalman filter itself and the second using a data association method to create robust object tracks.

Chang and Gong (2001) used a Bayesian network with Kalman filtering for people tracking in an indoor environment to address the matching issue between multiple people being tracked. With the same conditions that Chang used, Nguyen et al. (2003) applied a Kalman filter in a distributed monitoring system to track moving people. Vasuhi and Vaidehi (2014) proposed a Combined Gaussian Hidden Markov Model and Kalman Filter (CGHMM-KF) to detect and track multiple people in different environments.

Other researchers have used the Kalman filter with a single camera and the Hungarian algorithm (Kuhn, 1955) as an assignment method to track multiple objects. For example, Lütteke et al. (2012) track vehicles using this approach. Nandashri and Smitha (2015) proposed an algorithm to track multiple people by using the Kalman filter, and using the Hungarian algorithm to address the occlusion problem that may occur between individuals. Yussiff et al. (2014) proposed a human tracking algorithm with a Kalman filter and the Hungarian algorithm to link data in the previous frame to the current frame. Vasuhi et al. (2015) used a similar approach to track multiple people in an outdoor environment.

In this study, we also apply the Kalman filter and the Hungarian algorithm to track multiple crop plants in video sequences. To achieve this, the centroids of detected objects need to be determined and fed to the proposed tracking algorithm in order to improve any misdetections that may occur in certain frames due to different weather conditions and image capture conditions (vibration, variations in speed etc.). The proposed method is described in detail in the following section.

3. Proposed method

While detection allows us to locate object positions in each frame, this in itself does not provide information about the movement of individual objects between frames so objects cannot be tracked over time. Explicit tracking is necessary to follow object instances as they move through the field of view, according to an assumed movement model. The use of a tracking algorithm such as Kalman filtering that includes a model of motion allows refinement of the detection co-ordinates to produce a smoother track across multiple frames.

Fig. 1 shows a block diagram of a general scheme for the tracking algorithm proposed in this paper. This includes: (i) detection stage (described in Section 3.1); (ii) extraction of feature centroid (the x and y coordinates of the centre of each detected object); these coordinates will be used to predict the current location of the track; (iii) tracking stage, which tracks detected objects and includes association of detected objects to plant trajectories.

3.1. Crop detection

The proposed tracking algorithm relies on an initial detection algorithm which has been previously described by Hamuda et al. (2017). This detection method was proposed to identify cauliflower plants from video sequences acquired outdoors in the west of Ireland, where weather conditions can be quite variable and unpredictable. The method was tested against different environmental conditions (cloudy, partially cloudy, and sunny) for different stages of growth (from June until the end of September 2015). Various circumstances, such as partial occlusion between crop and weeds, partial crop disappearance from the scene, leaves which were partially eaten by insects, light changes, motion caused by the wind, different type of shadows, and various backgrounds (soil, nylon, stones, and other residues) were also included. A top-view camera position (a camera moving along the rows of cauliflower seedlings) was adopted to capture the image sequences with a frame resolution of 2704×1520 pixels. The method involves several steps: (i) the original frame sequences are reduced to a quarter of their original size for computational efficiency; (ii) blur filtering is applied to enhance and reduce the noise from the captured image; (iii) the blurred image (RGB image) is converted to the HSV color space; (iv) the HSV image is filtered between the minimum and maximum values and thresholding applied to identify regions of interest (ROI); (v) the ROIs are further refined by using a morphological erosion and dilation process; (vi) moments are calculated to determine the position and mass distribution of objects in video sequences; (vii) regions that satisfy the detection rule (area ≥ 100 and the perimeter ≥ 30) were deemed to be detected plants. The detected positions (centroid points) are the