Computers and Electronics in Agriculture 142 (2017) 139–154

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Original papers

Multiple camera fruit localization using a particle filter ☆

S.S. Mehta [a,*], C. Ton [b], S. Asundi [c], T.F. Burks [d]

[a] Department of Industrial and Systems Engineering, University of Florida, Shalimar, FL 32579, United States
[b] Research and Engineering Education Facility, University of Florida, Shalimar, FL 32579, United States
[c] Aerospace Science Engineering Department, Tuskegee University, Tuskegee, AL 36088, United States
[d] Department of Agricultural and Biological Engineering, University of Florida, Gainesville, FL 32611, United States

ARTICLE INFO

Article history:
Received 7 November 2016
Received in revised form 7 August 2017
Accepted 7 August 2017
Available online 5 September 2017

Keywords:
Fruit motion
Wind load
Non-Gaussian noise
Layered vision system

ABSTRACT

Apart from socioeconomic factors, success of robotics in agriculture lies in developing economically attractive solutions with efficiency comparable to that of the humans. Fruit localization is one of the building blocks in many robotic agricultural operations (e.g., yield mapping and robotic harvesting) that determines 3D Euclidean positions of the fruits using one or several sensors. It is crucial to guarantee the performance of the localization methods in the presence of fruit detection errors and unknown fruit motion (e.g., due to wind gust), so that the desired efficiency of the subsequent systems can be achieved. For instance, inaccurate localization may severely affect fruit picking efficiency in robotic harvesting. The presented estimation-based localization approach provides estimates of the fruit positions in the presence of fruit detection errors and unknown fruit motion, and it is based on a new sensing procedure that uses multiple ($\geqslant 2$) inexpensive monocular cameras. A nonlinear estimator called particle filter is developed to estimate the unknown position of the fruits using image measurements obtained from multiple cameras. The particle filter is partitioned into clusters to independently localize individual fruits, while the behavior of the clusters is manipulated at global level to maintain a single filter structure. Since the accuracy of localization is affected by errors in fruit detection, the presented sensor model includes non-Gaussian fruit detection errors along with image noise. Fruit motion can significantly reduce harvesting efficiency due to errors in locating moving fruits. In contrast to existing methods, the dynamics of fruit motion are derived and included in the localization framework to obtain time-varying position estimates of the moving fruits. A detailed theoretical foundation is provided for the new estimation-based fruit localization approach, and it is validated through extensive Monte Carlo simulations. The performance of the estimator is evaluated by varying the design parameters, measurement noise, number of fruits, amount of overlap in clustered fruit scenarios, and fruit velocity. Correlation of these parameters with the performance of the estimator is derived, and guidelines are presented for selecting the design parameters and predicting performance bounds under given operating conditions.

1. Introduction

The US fresh fruit industry is facing growing global market pressures that threaten its long-term viability. The combination of low commodity prices both domestically and abroad, high labor prices and low labor productivity present significant challenges for the fresh fruit market. In addition, the labor intensive and injury-prone working conditions experienced in specialty crop harvesting are leading to decline in skilled labor availability and increase in harvesting costs (Gongal et al., 2015). For example, Florida citrus, which is approximately 480,121 acres, had harvesting costs about 2–4 times Brazilian harvesting cost in the period from 1979 to 2009. In 2008–09, the delivered-in cost of Florida orange was \$1.070 per pound solids while that of the Brazilian orange was \$0.725 per pound solids. The harvesting cost alone for Florida orange in 2015–2016 was \$1208 per acre compared to the production cost of \$2328 per acre. According to economic studies, harvesting cost must be reduced by 50% to maintain global competitiveness (Brown, 2002). Robotic harvesting is being investigated as an alternate solution to manual picking to improve

☆ This research is supported in part by a grant from the United States Department of Agriculture Small Business Innovation Research (USDA-SBIR) award #2016-33610-25473 and the USDA NIFA AFRI National Robotics Initiative #2013-67021-21074. The USDA-SBIR awardee is GeoSpider, Inc, an agricultural machinery startup company of which Dr. Tom Burks is the CEO. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funding agency.

* Corresponding author.
E-mail addresses: siddhart@ufl.edu (S.S. Mehta), asundi@mytu.tuskegee.edu (S. Asundi), tburks@ufl.edu (T.F. Burks).

http://dx.doi.org/10.1016/j.compag.2017.08.007
0168-1699/

productivity, reduce harvesting costs, and, in general, create a sustainable agriculture industry.

Vision systems are ubiquitous in agricultural robotics due to their ability to provide information rich image feedback of the environment at reasonably low cost. The main drawback of vision systems is the absence of depth information since an image is formed by projecting the three-dimensional (3D) Euclidean space onto a two-dimensional image plane. Various operations in robotic harvesting, e.g., path planning and servo control, can benefit from knowing of the depth of the objects or features viewed by camera. Consider an example of robotic harvesting of specialty crops, such as citrus, where it is necessary to determine the order in which the fruits should be picked to minimize the robot travel time. To design such optimal trajectories for robots, it is necessary to know the Euclidean position (or depth) of the fruits. The process of obtaining Euclidean position by recovering or measuring the depth information is referred to as localization.

Vision-based localization approaches can be broadly classified into systems that use vision (or camera) as the only sensor and systems that use range measurement sensors in conjunction with vision. The later class of systems employ sensors such as lidar, sonar, and infrared range finders to determine depth map (e.g., see Harrell et al., 1990; Ceres et al., 1998; Jiménez et al., 2000; Tanigaki et al., 2008; Bulanon and Kataoka, 2010; Feng et al., 2012). Lidars and sonars typically provide range estimates by measuring the time-of-flight, i.e., the time required for the light or sound to travel to an object and back, while infrared sensors obtain depth using the principle of triangulation. The features or objects of interest seen in a camera image are subsequently localized by fusing the image with the obtained depth map. However, range measuring sensors are not widely used in outdoor fruit mapping applications, which can be due to high equipment cost of lidars ($3000–$5000) or inaccuracies of infrared sensors and sonars in cluttered outdoor conditions. Additionally, accurate calibration can be an issue as the response of infrared and sonar varies with ambient conditions (e.g., humidity, wind, dust) and the object of interest (e.g., color, flatness, surface area). The time-of-flight cameras are an emerging technology that uses similar principle as laser range finders to generate 3D maps by capturing an entire scene with light pulse. Due to high speeds and less complexity, the time-of-flight cameras can be promising in agricultural applications (see Karkee et al., 2014; Gongal et al., 2016). However, at present the applicability of the time-of-flight cameras is limited due to high cost (up to $12,000) and low image resolutions (less than 320×240 pixels). Also, the performance of infrared and time-of-flight cameras operating in the near-infrared range is affected by solar infrared noise making them best suited at night with supplemental lighting. In contrast, vision-only localization systems determine the unknown depth of an object using machine vision principles. Structure-from-motion (Muscato et al., 2005; Baeten et al., 2008) based approaches, using a single monocular camera, can be used to identify the unknown depth from the knowledge of the camera motion (displacement or velocity). Our previous work in Mehta and Burks (2014) relied on a known object model for depth recovery using a single monocular camera. Stereo-vision or triangulation based approaches use the notion of disparity (i.e., the difference in image location of the same 3D point viewed by two cameras) to determine the depth. Despite added complexity, stereo-vision is widely used in fruit localization (e.g., see Buemi et al., 1996; Kondo et al., 1996; Recce et al., 1996; Plebe and Grasso, 2001; Van Henten et al., 2002, 2003; Wang et al., 2013; Font et al., 2014). A comprehensive review of fruit localization methods can be found in Gongal et al. (2015).

At the University of Florida, we are investigating economically viable and straightforward sensing solutions for robotic harvesting. With increase in available processing power and advent of parallel architectures that use FPGAs and GPUs, stereo-vision-like multiple camera approaches hold significant potential. One of the approaches being considered, called layered vision system, consists of multiple sensing layers wherein each layer in turn is comprised of multiple ($\geqslant 2$) inexpensive ($30–$50) monocular cameras. The layers can be assigned specific tasks in robotic harvesting based on the location of the cameras. For example, a layer with cameras farthest from the tree canopy can perform fruit mapping while a layer comprised of robots' hand-held cameras can be responsible for visual servo control. Additionally, the layers can be connected to share information to improve detection efficiency and accuracy. Motivated by layered sensing, we propose a multiple camera fruit localization approach where the number of cameras viewing fruits can be more than two, i.e., beyond standard stereo-vision. The fruit detection problem is assumed to have been solved (e.g., Plebe and Grasso, 2001; Hannan et al., 2009; Bulanon et al., 2009; Krizhevsky et al., 2012; Payne et al., 2014; Zheng et al., 2015; Hung et al., 2015; Sa et al., 2016; Shelhamer et al., 2017; Bargoti and Underwood, 2017) to segment fruits from the background and yield the coordinates of the fruit centroids in the image space, i.e., 2D image coordinates. Fruit centroid is a point in the 3D Euclidean space, e.g., the center of sphere for spherical fruits. The image coordinates of the fruit centroid will then correspond to the center of the circle projected on an image plane. Spherical fruits (e.g., apples, oranges, peaches, and tomatoes) that have a well-defined fruit centroid are an obvious choice for the presented localization method. However, non-spherical fruits and vegetables, such as cucumber and zucchini, can also be localized provided the definition of their centroid is established *a priori* – recognizing that the image projection of asymmetrical fruits will vary based on the camera's viewpoint – and the image coordinates of the centroid can be obtained reliably during fruit detection. Apart from noise inherently present in digital cameras, unstructured lighting conditions in outdoor environments may result in poor image quality, inconsistent object segmentation, and consequently errors in obtaining the image coordinates of the true fruit centroids (Tian and Slaughter, 1998). Noise filtering (Jiménez et al., 2000; Bulanon et al., 2004; Hannan et al., 2009) and robust segmentation (Tian and Slaughter, 1998; Tang et al., 2000) approaches may alleviate harmful effects of noise, however noise cannot be completely eliminated.

In this paper, a Bayesian probabilistic approach is taken to obtain belief of the true fruit positions instead of attempting to identify unique or exact solution from the noisy image measurements. Specifically, a nonlinear estimator called particle filter is presented to obtain the posterior distribution of the fruit positions using noisy and erroneous image measurements of the fruit centroids. In particle filtering, the posterior distribution is represented using a set of particles (hypotheses). The paper uses a single particle filter to estimate the location of multiple fruits. Although a single filter ensures computational tractability and offers robustness to false positives (i.e., false fruit detections) via particle resampling, it has a limitation to multiple target (or fruit) tracking problems. A particle filter after resampling may lose particle resolution over some targets, i.e., it may not be able to localize all targets. To address this issue while maintaining robustness to false detections, the particles are partitioned into multiple subfilters using a simple clustering procedure such that each subfilter is assigned to exactly one fruit. The subfilters estimate the position of an individual fruit, and they are processed independently from each other. A single filter structure enables the global behavior of the estimator to be affected by dissolving and then reforming the subfilters based on measurement-particle association following each new measurement. It enables one to eliminate defunct filters arising from false detections and to accommodate newly detected fruit. Some parallels can be drawn between the presented approach and the track-oriented and Bayesian multiple hypothesis tracking (MHT)