Computers and Electronics in Agriculture 139 (2017) 231–242

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Original papers

Flexible system of multiple RGB-D sensors for measuring and classifying fruits in agri-food Industry

Rodrigo Méndez Perez [a], Fernando Auat Cheein [a,*], Joan R. Rosell-Polo [b]

[a] Department of Electronic Engineering, Universidad Técnica Federico Santa María, Valparaíso, Chile
[b] Research Group in AgroICT & Precision Agriculture, Department of Agricultural and Forest Engineering ETSEA, University of Lleida-Agrotecnio Center, Lleida, Spain

ARTICLE INFO

Article history:
Received 6 August 2016
Received in revised form 14 May 2017
Accepted 16 May 2017

Keywords:
Fruit detection
Depth sensor
Fruit classification
Phenotyping

ABSTRACT

The productivity of the agri-food sector experiences continuous and growing challenges that make the use of innovative technologies to maintain and even improve their competitiveness a priority. In this context, this paper presents the foundations and validation of a flexible and portable system capable of obtaining 3D measurements and classifying objects based on color and depth images taken from multiple Kinect v1 sensors. The developed system is applied to the selection and classification of fruits, a common activity in the agri-food industry. Being able to obtain complete and accurate information of the environment, as it integrates the depth information obtained from multiple sensors, this system is capable of self-location and self-calibration of the sensors to then start detecting, classifying and measuring fruits in real time. Unlike other systems that use specific set-up or need a previous calibration, it does not require a predetermined positioning of the sensors, so that it can be adapted to different scenarios. The characterization process considers: classification of fruits, estimation of its volume and the number of assets per each kind of fruit. A requirement for the system is that each sensor must partially share its field of view with at least another sensor. The sensors localize themselves by estimating the rotation and translation matrices that allow to transform the coordinate system of one sensor to the other. To achieve this, Iterative Closest Point (ICP) algorithm is used and subsequently validated with a 6 degree of freedom KUKA robotic arm. Also, a method is implemented to estimate the movement of objects based on the Kalman Filter. A relevant contribution of this work is the detailed analysis and propagation of the errors that affect both the proposed methods and hardware. To determine the performance of the proposed system the passage of different types of fruits on a conveyor belt is emulated by a mobile robot carrying a surface where the fruits were placed. Both the perimeter and volume are measured and classified according to the type of fruit. The system was able to distinguish and classify the 95% of fruits and to estimate their volume with a 85% of accuracy in worst cases (fruits whose shape is not symmetrical) and 94% of accuracy in best cases (fruits whose shape is more symmetrical), showing that the proposed approach can become a useful tool in the agri-food industry.

1. Introduction

Although LiDARs (Light Detection and Ranging) have been widely used in new agricultural technology and applications, the information they provide is associated with distance only. Thus geometric characterization is possible (Andújar et al., 2013) but the processing of valuable vegetative information is difficult to be further enchanced. In this context, artificial vision systems (such as monocular or stereo cameras, as well as NIR –near infra-red cameras–) are used for monitoring groves (Chéné et al., 2012; Nissimov et al., 2015; Gongal et al., 2015) and its growing (Gongal et al., 2015; Mehta and Burks, 2014). In particular, we can find nowdays vision systems for unharvested fruit recognition (Schöler et al., 2015; Jay et al., 2015; Xu and Payandeh, 2015), leaf density estimation (Erdal et al., 2015) and flowers detection and classification (Rosell-Polo et al., 2015), among other tasks. It is to be noted that within the artificial vision field, light structured sensors are the current research focus in many academic groups, such as (Rosell-Polo et al., 2015). Light structured sensors (LSS, such as the commercial *Kinect* made by Microsoft Corporation, Redmond, WA, USA) are low cost sensors whose usability in the agronomic context is still under study, as can be seen in a previous work of the authors (Rosell-Polo et al., 2015). LSS sensors can be used to estimate foliage density, flower density, geometric characteristics

* Corresponding author.
E-mail addresses: rodrigo.mendez.11@alumnos.usm.cl (R. Méndez Perez), fernando.auat@usm.cl (F.A. Cheein), jr.rosell@eagrof.udl.cat (J.R. Rosell-Polo).

http://dx.doi.org/10.1016/j.compag.2017.05.014
0168-1699/

of the orchard or the stems and even terrain parameters (Andújar et al., 2013), using the depth readings and RGB images provided by the sensor (for this reason they are also called RGB-D or depth sensors). However, the use of LSS or RGB-D sensors in the agri-food industry still is an open issue to be addressed. In this work, we explore the possibility of implementing multiple RGB-D sensors in order to take advantage of the color and depth information.

Multiple vision based solutions have been proposed to solve growing characterization problems. In Jay et al., 2015 the authors developed a system capable of generating a 3D model of plants and classify the different types of plants by analysing their leaves. To achieve this, a mechanical architecture is used, where a color camera is mounted on a metal girder and takes multiple captures from different angles. A similar solution is presented in Yeh et al., 2013, where two color cameras are mounted in a robotic arm, which moves around a leafy vegetable taking multiple pictures. These solutions and others which implement multiple RGB-D sensors proposed to solve similar problems in other areas (Susanto et al., 2012; Satta et al., 2013; Caon et al., 2011, among many others) use specific set-up or need a previous calibration in order to operate.

To overcome the above mentioned issues, in this work a flexible and portable system of multiple RGB-D sensors capable of self location and self calibration of the sensors to then start detecting, classifying and measuring fruits applicable to the agri-food industry is proposed and validated. It is considered the case where different types of fruits are transported in a conveyor belt at the same time. To obtain an accurate classification and characterization, we use computer vision and advanced soft computing methods, which are explained in detail in the following sections. The characterization process considers: classification of fruits, estimation of its volume and the number of assets per each kind of fruit. The entire system works in real-time, with a sampling time of 0.1 s, and does not need an expert operator to install it. Processing times is a key issue to face when working in conveyor belts, in order to avoid missing transported fruits or misclassifying them.

This paper is organized as follows. Section 2 presents the proposed system. In Section 3 the implementation and validation of the system and its methods are described. Lastly, Section 4 draws conclusions and provides the guidelines for further work.

2. Materials and methods

According to the requirements stated in Section 1, the proposed system must be capable of integrating the information acquired from multiple RGB-D sensors without knowing their exact position and orientation, and once the position and orientation of the sensors have been estimated, the RGB-D system should be able to detect, classify and measure the characteristics of the fruits that pass through. The following sections describe in detail the system developed in this work.

2.1. System architecture

The framework of the proposed system is illustrated in Fig. 1. The first stage deals with the self localization of the sensors, whereas the second stage faces the processing steps to detect, classify and measure the fruits. Briefly,

- We use multiple RGB-D sensors randomly placed over a fruit table. Such sensors correspond to the Kinect v1, manufactured by Microsoft.
- Next, we solve the localization problem: the main goal is to be able to place the Kinect sensors in the work place without increasing the costs of the system, i.e., avoiding further calibration. Then, we study the sensor errors and the error propagation associated with our goal.
- Finally, we analyze the processing stage: RGB and depth information are used to detect and classify fruits.

Following, each part of the system architecture shown in Fig. 1 is presented in detail.

2.2. RGB-D sensor

The RGB-D sensor used is the Microsoft Kinect v1. It can obtain dense depth estimates using a structured light pattern. The device contains a color camera, an active infra-red camera and a laser projector. The RGB-D sensor uses an infra-red structured random light pattern and interferences will occur if two or more sensors point to the same area. In order to avoid such interference it is possible to alternate the laser projectors (by switching them on and off) and obtaining depth images alternately.

As any other RGB-D sensor, the dephts obtained by the Kinect from Microsoft are affected by measurement errors, which have been widely studied (Andersen et al., 2012; Khoshelham, 2012; Langmann et al., 2012). The minimum distance that it is able to measure is about ~800 mm and the maximum distance is about ~4000 mm.

2.3. Localization

As it was mentioned earlier, a requirement for the system to be able to localize the sensors is that each sensor must share (partially) its field of view with at least another sensor. The sensors localize themselves by estimating the rotation and translation matrices that allow to transform the coordinate system of one sensor into the other. To achieve this the *Iterative Closest Point* (ICP) algorithm (Besl and McKay, 1992), which is capable of estimating the rotation and translation between two point clouds, is used. Since we are able to obtain a depth image from every sensor, it is possible to transform them to point clouds and compute the rotation (R) and translation (T) between two sensors that share part of their fields of view as:

$$
\begin{aligned}
&[R, T] = ICP(X_i, X_j) \\
&X_{j(k)} \approx RX_{i(k)} + T \quad \forall\ k \in [1, M]
\end{aligned}
\tag{1}
$$

where $X_i \in \Re^{3\times M}$ corresponds to the point cloud captured by the sensor $i, X_j \in \Re^{3\times M}$ is the point cloud captured by the sensor j and k is a point that belongs to the point cloud, which is composed of

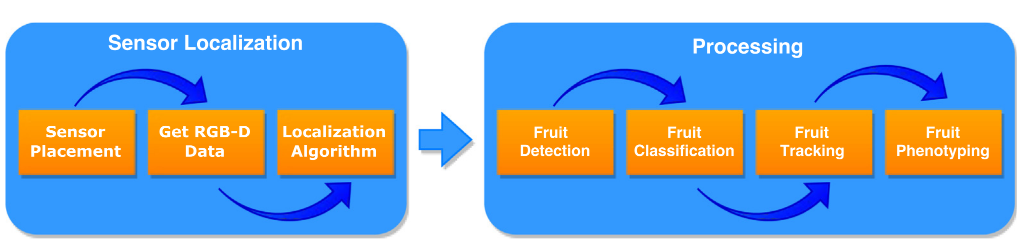

Fig. 1. Framework of the proposed work.