BIOSYSTEMS ENGINEERING 171 (2018) 272–289

Available online at www.sciencedirect.com

ScienceDirect

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

Research Paper

Statistical models for fruit detectability: spatial and temporal analyses of sweet peppers

Polina Kurtser[*], *Yael Edan*

Department of Industrial Engineering and Management, Ben-Gurion University of the Negev, Beer Sheva, Israel

ARTICLE INFO

Article history:
Received 31 March 2017
Received in revised form
17 January 2018
Accepted 20 April 2018

Keywords:
Detectability modelling
Spatial viewpoints
Sweet pepper
Fruit detectability
Harvesting
Robotics

Statistical models for fruit detectability were developed to provide insights into preferable variable configurations for better robotic harvesting performance.

The methodology includes several steps: definition of controllable and measurable variables, data acquisition protocol design, data processing, definition of performance measures and statistical modelling procedures. Given the controllable and measurable variables, a data acquisition protocol is defined to allow adequate variation in the variables, and determine the dataset size to ensure significant statistical analyses. Performance measures are defined for each combination of controllable and measurable variables identified in the protocol. Descriptive statistics of the measures allow insights into preferable configurations of controllable variables given the measurable variables values. The statistical model is performed by back-elimination Poisson regression with a loglink function process. Spatial and temporal analyses are performed.

The methodology was applied to develop statistical models for sweet pepper (*Capsicum annuum*) detectability and revealed best viewpoints. 1312 images acquired from 10 to 14 viewpoints for 56 scenes were collected in commercial greenhouses, using an eye-in-hand configuration of a 6 DOF manipulator equipped with a RGB sensor and an illumination rig. Three databases from different sweet-pepper varieties were collected along different growing seasons.

Target detectability highly depends on the imaging acquisition distance and the sensing system tilt. A minimum of 12 training scenes are necessary to discover the statistically significant spatial variables. Better prediction was achieved at the beginning of the season with slightly better prediction achieved in a temporal split of training and testing sets.

1. Introduction

Despite intensive R&D on harvesting robots, to date no commercial harvesting robot exists (Bac, van Henten, Hemming, & Edan, 2014). One major limitation of current developments is the low detection rates of around 85% (Bac, van Henten, et al., 2014) caused by the complex and highly variable agricultural environment.

Most robotic harvesters are equipped with a vision sensor in an eye-in-hand configuration (Bac, van Henten, et al., 2014). Current agricultural robotics detection algorithms usually use

* *Corresponding author*. Ben-Gurion University of the Negev, P.O.B. 653, Beer-Sheva 8410501, Israel.
E-mail address: kurtser@post.bgu.ac.il (P. Kurtser).

https://doi.org/10.1016/j.biosystemseng.2018.04.017

Nomenclature

P_i	Proportion of visible targets from viewpoint i
P_{ij}	Proportion of visible targets from a combination of viewpoints i and j
T_i	Number of visible targets from viewpoint i
T_{ij}	Number of visible targets from a combination of viewpoints i and j
$\widehat{T}_i$	Predicted number of visible targets from viewpoint i
T_{ik}	Number of targets detection from viewpoint *i* of type *k*
T	Total number of targets in a scene
$\overline{T}$	Average number of targets visible from a viewpoint
S_i	Weighted score of viewpoint i
w	Relative weight given to each target type
A_i	Bounding box area of object *i*
RA_i	Relative bounding box area

a single preset viewpoint (Bac, van Henten, et al., 2014; McCool et al., 2016; Sa et al., 2016), however since single viewpoint visibility is limited (Bulanon, Burks, & Alchanatis, 2009; Hemming, Ruizendaal, Hofstee, & van Henten, 2014) multiple viewpoints are necessary to improved detection rates. Since cycle times are critical (Bac, van Henten, et al., 2014) it is important to direct the robot to a minimum number of viewpoints that can provide maximal detectability.

The search for an optimal viewpoint has been extensively investigated in many fields (Fleishman, Cohen-Or, & Lischinski, 2000; Foix, Alenyà, & Torras, 2011; Maver & Bajcsy, 1993; Reed & Allen, 2000). However, the high scene variability in a harvesting application, which is inherent to the biological nature of the scene, reduces the ability to calculate a-priori the best viewpoints due to the very limited geometric information about the scene. The complexity of the fruit detection task is due to the unstructured and dynamic nature of both the objects and the environment (Gongal, Amatya, Karkee, Zhang, & Lewis, 2015; Kapach, Barnea, Mairon, Edan, & Ben-Shahar, 2012; McCool et al., 2016; Sa et al., 2016): fruits have a high inherent variability in size, shape, texture, and location; in addition, occlusion and variable illumination conditions significantly influence the detection performance. This research aims to determine the dependency of detectability on the chosen viewpoint, and to analyse the temporal and spatial relations between consecutive scenes. The proposed methodology for developing statistical prediction models of fruit detectability for vision based robotic harvesters enables to prove correlations between controllable and measurable variables and the detectability performance measures, providing insights into preferable configurations for better robotic harvesting performance.

The paper starts with a literature survey of common practices in detectability and visibility research. Section 3 outlines a methodology for viewpoint detectability modelling and definition of performance measures. It aims to present the minimum size of training sets for which the suggested controllable variables are still found to be significant and therefore will lead to a correct sensing plan that will increase detectability. Section 4 presents the results of application of the proposed methods on a case study database of *Capsicum annuum* (sweet peppers).

2. Background

2.1. *Detectability and visibility*

Current agricultural robotic detection algorithms aim to maximise the true positive rate and minimise the true negative rate (Vitzrabin & Edan, 2016b). The standard procedure applied in the computer vision community for defining true positive and true negative rates is labelling images (Russell, Torralba, Murphy, & Freeman, 2008). The labelling process includes either a bounding box or pixel-wise labelling resulting in segmented image into areas representing the targets and the background. This is performed by annotators reviewing the images for unlimited amount of time (Deng et al., 2009; Russell et al., 2008). In some cases the annotators are requested to classify the targets annotations into fully revealed targets, partially occluded targets and truncated targets (Geiger, Lenz, & Urtasun, 2012). The labels are often used both as a training set of supervised learning algorithms as well as a benchmark for the detection algorithms.

However, the annotation of a single image is not sufficient for object detectability within a scene. The number of targets within the scene is defined as the ground truth. The common procedure for obtaining *detectability ground truth* is placing several sensors that simultaneously sense the same scene and then combining and annotating the joint number of targets in the scene (Dollar, Wojek, Schiele, & Perona, 2012; Russell et al., 2008). By doing so a benchmark of joint detectability using vision sensors is generated. The placement of the sensors in a way that covers the whole operational area is critical for proper ground truth acquisition, providing detection of all targets relevant to the robotic task (e.g., reachable by the robotic manipulator, to be avoided by an autonomous vehicle).

An alternative type of ground truth, that has been used for agricultural applications (Bac, van Henten, et al., 2014; Hemming, Ruizendaal, et al., 2014), and some SLAM applications (Blanco, Moreno, & Gonzalez, 2009) is the *visibility ground truth*. Visibility in robotic harvesting is defined as "*The visible part of a fruit in an image expressed as a percentage of total fruit area which would be seen in an image without occlusion*" (Hemming, Ruizendaal, et al., 2014). To obtain the visibility ground truth an in-field human observer manually counts the actual number of targets in the scene. The visibility analysis compares between targets manually labelled from the vision system to the actual number of targets present within the scene. This allows a visibility benchmark given sensors capable of detecting targets occluded by other objects in the scene. While it is an important analysis for evaluating the final performance of a robotic harvester, it does not provide a true benchmark for vision-based robots that cannot "see-through" obstacles. Furthermore, it is limited in the number of data that can be analysed since it requires accurate counting of all fruit in the analysed scenes. Reported numbers should be