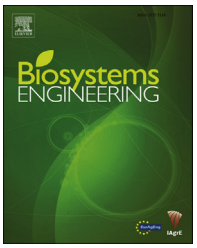
BIOSYSTEMS ENGINEERING XXX (2016) 1–14

Available online at www.sciencedirect.com

ScienceDirect

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



Research Paper

Special Issue: Robotic Agriculture

Colour-agnostic shape-based 3D fruit detection for crop harvesting robots

Ehud Barnea, Rotem Mairon, Ohad Ben-Shahar *

The Interdisciplinary Computational Vision Lab, Computer Science Department, Ben Gurion University of the Negev, Israel

ARTICLE INFO

Article history:
Published online xxx

Keywords:
Agrobotics
Shape
Highlights
Symmetry
RGB-D
Green sweet pepper

Most agricultural robots, fruit harvesting systems in particular, use computer vision to detect their fruit targets. Exploiting the uniqueness of fruit colour amidst the foliage, almost all of these computer vision systems rely on colour features to identify the fruit in the image. However, often the colour of fruit cannot be discriminated from its background, especially under unstable illumination conditions, thus rendering the detection and segmentation of the target highly sensitive or unfeasible in colour space. While multispectral signals, especially those outside the visible spectrum, may alleviate this difficulty, simpler, cheaper, and more accessible solutions are desired. Here exploiting both RGB and range data to analyse shape-related features of objects both in the image plane and 3D space is proposed. In particular, 3D surface normal features, 3D plane-reflective symmetry, and image plane highlights from elliptic surface points are combined to provide shape-based detection of fruits in 3D space regardless of their colour. Results are shown using a particularly challenging sweet pepper dataset with a significant degree of occlusions.

1. Introduction

A major and challenging task of agricultural robotic systems is that of detecting and localising fruit. Systems designed to count or harvest fruit require an accurate detection scheme that is able to overcome challenges such as naturally occurring changes in illumination, shape, pose, colour, and viewpoint. All these factors are prevalent in natural environments, making target detection in agricultural settings particularly challenging.

Colour is often a distinctive and indicative cue for the present of fruit. Indeed, fruits are frequently red (sweet peppers, strawberries, etc...), orange (oranges, persimmons, etc...), or yellow (bananas, sweet peppers, etc...), and thus they stand out from the green foliage. In such conditions, their presence is easily identified in images by simple colour processing, as indeed practiced in an overwhelming majority of agricultural vision (*agrovision*) methods (Kapach, Barnea, Mairon, Edan, & Ben-Shahar, 2012). Obviously, in certain cases the use of colour introduces high degree of uncertainty. Some fruits are simply green (certain apples, sweet peppers, cucumbers, etc...)

* *Corresponding author*. Tel.: +972 8 6477868.
E-mail address: ben-shahar@cs.bgu.ac.il (O. Ben-Shahar).
URL: http://www.cs.bgu.ac.il/~icvl

http://dx.doi.org/10.1016/j.biosystemseng.2016.01.013
1537-5110/

Nomenclature

S	Saturation
V	Luminance
I	Input image
I_x,I_y	Partial derivatives of I
C_i	the i-th circle in a group of concentric circles
r_i	Radius of circle C_i
$C_i(p)$	The set of all pixel locations at radius r_i from p
o_j	The j-th orientation value at a location among $C_i(p)$
e_j	The j-th expected orientation value at a location among $C_i(p)$.
W	Spatial neighbourhood
$S_W(p)$	The second moment matrix approximation of neighbourhood W of p
$w(r)$	Weight associated with r
G_σ	Gaussian kernel with variance σ
λ_1, λ_2	Eigenvector of matrix $S_W(p)$, where $\lambda_1 > \lambda_2$
$S(C_i)$	Percentage of computed and expected orientations at $C_i(p)$ that agree
T_h	Accumulated support threshold
T_a	Similarity threshold
T_s	Agreement threshold
n_x	Surface normal at point x
θ, ϕ	Angles representing a vector
p_v	Viewpoint
n_p	Normal of symmetry plane

Abbreviation

HSV	Hue Saturation Value
HoG	Histogram of Orientation Gradients
SVM	Support Vector Machine

and thus they are hardly distinguishable from the foliage on the basis of colour alone (see Fig. 1a). Moreover, illumination may not be good enough to render colour differences between fruit and background or too unreliable to allow robust detection even for non-green fruits (see Fig. 1b). Following these observations, the focus of this work is a fruit detection framework that is completely agnostic of colour and thus able to cope with such challenges. The two cues that we propose to exploit are depth (or shape) and highlights.

The recent advances in the research and development of range sensors and depth cameras have made depth information of scenes (both indoors and outdoors) readily available in the form of RGB-D images or 3D point clouds. Depth modality, which is inherently different than colour and intensity, has lately been employed to solve many kinds of general computer vision problems, such as object recognition (e.g., Lai, Bo, Ren, and Fox (2011)), object detection (e.g., Hinterstoisser et al., 2011; Spinello & Arras, 2011, pose estimation (e.g., Shotton et al., 2013); Aldoma et al., 2011) and segmentation (e.g., Silberman & Fergus, 2011).

Seeking to improve agricultural robots, RGB-D images were also used to detect several types of fruit (Chi & Ling, 2004; Edan, Rogozin, Flash, & Miles, 2000; Hannan & Burks, 2004). Indeed, the depth modality becomes particularly useful in agricultural applications as it makes the shape and the geometry of visible objects explicit with relatively little sensitivity to illumination conditions (Harrell, Slaughter, & Adsit, 1989). Furthermore, when fruit colour is similar to the surrounding foliage, shape becomes a more prominent cue toward separating the two.

Another visual cue that is independent of colour is highlights. Highlights tend to appear more often on the smoother, more specular, and typically elliptical (in the sense of differential geometry, see Do Carmo & Do Carmo, 1976) fruit regions (compared to the foliage) in places where the surface normal bisects the angle between illumination and viewing directions. While not categorical, the presence of highlights thus increases the probability that the image region from which they reflect belongs to a fruit. The challenge, of course, is to properly find such highlights and discriminate them from other bright image regions or highlights that originate from other non-fruit entities within the image.

Combining both highlight detection and 3D shape/range data, and inspired by recent advancements in general computer vision and previously suggested fruit detection systems (Kapach et al., 2012), here a colour-agnostic fruit detection framework composing of two steps is proposed: a rapid highlight-based candidate generation step, followed by a costlier 3D shape-based detection step.

In the first step, regions that are likely to contain fruit are found by detecting intensity highlight signatures in the image. In the second step, the likely regions are processed with a depth-based object detector that is partially invariant to changes in pose (Barnea & Ben-Shahar, 2014). In order to

(a) (b)

Fig. 1 – Example of agricultural images where colour cannot assist in detection. (a) Green fruit amidst green foliage. (b) A distinctive coloured fruit (left) becomes more similar to the surrounding foliage under different illumination conditions (right).

Please cite this article in press as: Barnea, E., et al., Colour-agnostic shape-based 3D fruit detection for crop harvesting robots, *Biosystems Engineering* (2016), http://dx.doi.org/10.1016/j.biosystemseng.2016.01.013

account for the different poses in which a fruit may appear, the detector processes each image region by first finding the best 3D symmetry reflection plane (assuming that the region contains a fruit), and then accumulating local shape features (i.e., surface normals) relative to a 3D frame defined using the detected symmetry.

Accumulated features are then classified using existing machine learning methods. The generalisation power of the learning scheme then permits not only certain robustness over shape variations, but also for occlusions.

In what follows, we provide a survey of relevant literature (Section 2), an elaboration on the detection framework (Section 3), an evaluation of our method on a challenging bell-pepper dataset (Section 4), and finally, we discuss our conclusions and future work (Section 5).

2. Background

Object detection in colour images has been a subject of research for many years. Instead of covering it here again we refer the reader to a recent review of the literature by Kapach et al. (2012), or the earlier review by Jimenez, Ceres, and Pons (2000a). The different subsections that follow discuss background material related to the use of range data, visual highlights, and symmetry, to provide the background relevant to the approach presented later.

2.1. *Detection of fruit and general object categories in depth images*

With the introduction of depth data to agrovision, new opportunities as well as challenges emerge, particularly how to properly use depth data, or how it may be used in conjunction with RGB data. Jimenez, Ceres, and Pons (2000b), for example, exploited the spherical shape of oranges by looking for and aggregating a set of local primitives that are likely to belong to spherical objects. However, the more common approach of exploiting RGB-D data in agrovision is the case of RGB followed by depth analysis. Specifically, the colour of fruits is used to segment the region of interest (usually of a cluster of fruits) from the RGB image, and the registered depth map is then used on the segmented parts to localise fruits in space. For example, Monta and Namba (2003) used this cascade for the detection of tomatoes, where depth data was also used to distinguish individual fruits that are part of a single colour segment. Fruit candidate regions were generated by thresholding the colour channels, and separating single fruits by examining and thresholding the spatial distance between adjacent pixels in the candidate regions. Needless to say that in all systems of that type, the colour of the fruit is highly discriminative and easily detected and that all processing, both in the colour and the depth domains, is highly fruit-dependent.

Clearly, ways of using range or RGB-D data also arise outside agrovision in the general computer vision literature, and ideas that presently develop in the latter can serve the former as well. Recently, a baseline study by Janoch et al. (2011) employed the popular part-based detector by Felzenszwalb, McAllester, and Ramanan (2008) for object detection based on a variant of the HoG algorithm (Dalal & Triggs, 2005). These algorithms employ the sliding window approach, in which every window in the scene is represented as a point in high dimensional space by computing a set of features for the data inside the window. Machine learning classifiers then learn to classify a window as containing an object or not by finding a function that separates the high dimensional feature space into a part containing mostly windows with objects and another part containing mostly windows without objects. In the study by Janoch et al. (2011), the sliding window approach was combined with a representation based on histograms of edge orientations, but applied it directly on depth images as if they were intensity/colour images. Perhaps expectedly, this yielded inferior performance, suggesting that depth should not be treated as if it was intensity or colour. Borrowing these insights, Tang et al. (2012) also used the HoG formulation but with histograms of surface *normals* that are characterised by two spherical angles.

These features have been shown to produce better results than those obtained by the HoG algorithm over intensity/colour images and better than HoG over depth images, indicating again that depth should not be treated as intensity. A different approach to the analysis of depth information attempts to facilitate prior segmentation (Bo, Lai, Ren, & Fox, 2011; Redondo-Cabrera, López-Sastre, Acevedo-Rodriguez, & Maldonado-Bascón, 2012). Kim, Xu, and Savarese (2013), for example, proposed employing such information by generating a small set of segmentation hypotheses, and then use both HoG features and depth features from these segmented regions in a part-based model generalised to 3D. Their scheme resulted in a feature vector containing both appearance and 3D shape features, which gave better results in most categories. Other ways to use depth information include estimating object size (Janoch et al., 2011; Saenko et al., 2011) and combining detector responses from different views (Lai, Bo, Ren, & Fox, 2012).

Regardless of the colour or the depth features employed, an important issue of object detection is the treatment of object pose. As is clearly needed for agrovision, a robust object detector would be general enough to capture its sought-after target (fruit, in the agrovision case) at different poses, and there are different ways of doing so. The naïve way, as shared by most object detectors, is to rely on machine-learning classifiers to be able to generalise diverse training data. However, machine-learning methods have limitations (like any other method) and cannot always be expected to generalise well. In order to achieve better results, some researchers try and provide the learning algorithm with simpler examples to learn from. This is done by estimating the object's pose prior to the classification phase (Lin & Davis, 2008) and using it to align the object to a canonical pose or to calculate features in relation to the estimation.

Recapping on the above, current RGB-D algorithms in agrovision are too specific and fruit-dependent, while general computer vision algorithms are too general and perform inferiorly on agricultural data. Here we seek to bridge the two approaches and seek a general detector that can be successfully applied for the detection of fruit. Using both highlights and symmetry of objects, these aspects are reviewed before moving on to discuss our main contribution.

Please cite this article in press as: Barnea, E., et al., Colour-agnostic shape-based 3D fruit detection for crop harvesting robots, *Biosystems Engineering* (2016), http://dx.doi.org/10.1016/j.biosystemseng.2016.01.013

2.2. Detection of highlights

Theoretically, when a light source illuminates an object in a visual scene, part of the light is immediately reflected back, while the remainder infiltrates the object. Of the infiltrating light, some would pass deeper through the object and some would reflect back onto its surface and into the air. The light reflected immediately is called *specular*, while the light reflected after penetrating the object is called *diffuse*. The former is typically reflected according to the law of specular reflection (and more generally, on Snell's law) while the latter may reflect in many different directions depending on the interactions exhibited inside the material.

The physical properties of the illuminated object determine the specular and diffuse components of the light reflected from it. Many common materials exhibit a mixture of both components while those that have more specular reflection are known as glossy or shiny objects. In a two dimensional projection of a visual scene, at viewing angles in which the reflection dominates, these reflections often appear as bright spots of light called *specular highlights* (Beckmann & Spizzichino, 1963).

Specular highlights are often regarded as a nuisance for practical computer vision applications. A main reason lies in their characteristic high intensity and low saturation values, which appear as intense white regions in the image of the scene. Often, these regions hinder image processing algorithms that are based on colour information and decision thresholds (e.g., segmentation and edge detection). Additional difficulties rise from the viewpoint dependent appearance of specular highlights, which interferes with image registration and subsequent image processing tasks (e.g., stereo matching and object recognition). Naturally, these properties of specular highlights have been used in order to remove them from acquired visual data and improve the performance of a wide range of computational tasks. Several methods used the viewpoint dependent appearance of highlights in order to detect them either by acquiring images of the same scene from multiple views (Lee & Bajcsy, 1992; Lin, Li, Kang, Tong, & Shum, 2002; Nayar, Fang, & Boult, 1997) or by changing the light source direction (Lin & Shum, 2001; Park & Tou, 1990; Sato & Ikeuchi, 1994). These approaches, however, are not always applicable since they require modifying the general setting of the scene.

Removal of specular highlights without using visual cues related to their view dependent appearance (i.e., using a single image) is more challenging. Several methods rely on an image of the diffuse component, generated according to a reflection model and the parameters of the acquisition device (Mallick, Zickler, Belhumeur, & Kriegman, 2006; Mallick, Zickler, Kriegman, & Belhumeur, 2005; Shen & Cai, 2009; Tan & Ikeuchi, 2005b). Other approaches for highlights removal analyse the distributions of image colours within a colour space (Tan & Ikeuchi, 2005a; Tan, Quan, & Lin, 2006). These methods, however, are not capable for real-time applications and focus on the removal of specular highlights without their explicit detection.

While their removal may be beneficial to some applications, specular highlights may also be regarded as informative visual cues. For example, specular highlights were used in order to detect shiny and transparent objects by Osadchy, Jacobs, and Ramamoorthi (2003) who exploited them as unique signals to estimate the pose of objects (Netz & Osadchy, 2011), and in the extreme, when the object was entirely specular, for the full reconstruction of its geometry from several (Adato & Ben-Shahar, 2011, Adato, Vasilyev, Zickler, & Ben-Shahar, 2010) or even one image (Vasilyev, Zickler, Gortler, & Ben-Shahar, 2011). Indeed, under challenging viewing conditions, where conventional (e.g., colour, brightness) features are often unreliable or insufficient for the detection of visual objects, the unique appearance of specular reflections may be exploited to detect, characterise and model the objects on which they form. This is relevant for all objects of some specular characteristic, and clearly true for many types of fruits. Here the focus is on sweet peppers, but one could consider many other fruits as well.

Indeed, due to their physical properties, sweet pepper fruits almost always produce specular reflections of the light source illuminating them. Considering the abundance of visual information in the natural cluttered scenes of sweet pepper fruits, specular highlights can be significant signals that can be exploited for their detection.

Here a novel method to detect specular highlights with a specific application to localise sweet pepper fruits (or more generally any fruit of sufficiently glossy and smooth shape) is presented. Our method is based on a model of specular highlights that exploits a particular relationship between highlights and image gradients. The model uses no prior knowledge of lighting direction and requires no calibration. As will be shown, it not only can be computed in real time but it also greatly enhances the reliability of detecting specular highlights compared with simply relying on their luminance and saturation.

2.3. Detection of symmetry

Symmetry is a phenomenon occurring abundantly in nature. Extensive research has been carried out trying to detect all kinds of symmetry in both 2D (e.g., Park et al. 2008) and 3D (e.g., Mitra, Pauly, Wand, & Ceylan, 2012). Applications for detecting symmetry are numerous, with special interest in object detection. Indeed, when searching for objects in a clutter, symmetry is not only naturally organised perceptually (Wertheimer, 1923), but is also indicative of natural or man-made objects (Rosen, 2011). In this work fruit detection is included in range data with symmetry detection in order to better discriminate fruit from clutter. The challenge then becomes one of detecting symmetry.

Complete symmetry detection (including the detection of multiple types of symmetry and across multiple objects and scales in an image) is a difficult problem due to the different types of symmetry found in nature. For this reason, research is usually focused on specific symmetries, ranging from rigid transla tion (Zhao & Quan, 2011) and rigid reflection (Loy & Eklundh, 2006; Podolak, Shilane, Golovinskiy, Rusinkiewicz, & Funkhouser, 2006), to non-rigid symmetry (Raviv, Bronstein, Bronstein, & Kimmel, 2007), reflection relative to general curves or curved glide-reflection (Lee & Liu, 2012), or a hierarchy of different symmetries (Thrun & Wegbreit, 2005).

Proposed symmetry detection methods may also be classified as solving for either partial or global symmetries. While

Please cite this article in press as: Barnea, E., et al., Colour-agnostic shape-based 3D fruit detection for crop harvesting robots, *Biosystems Engineering* (2016), http://dx.doi.org/10.1016/j.biosystemseng.2016.01.013

an image containing various symmetric objects is likely to contain a great deal of local symmetries, the image itself may not necessarily be globally symmetric. Partial symmetry detection entails finding the symmetric parts of the image, in contrast to global symmetry detection in which all of the image pixels are expected to participate. More formally, global symmetry, being a special case of partial symmetry (Mitra et al., 2012), is characterised by a transformation that maps the entire data to itself, while for partial symmetry a sought-after transformation maps only a subset of the data to itself. Occasionally, local symmetries are treated as global symmetries after confining the region of interest to local region of the data.

As a final note, it should be mentioned that as a tool facilitating object detection and pose localisation symmetry should be discriminative enough. This is not the case for completely spherical objects that have infinite number of symmetry axes and planes. As long as the target deviates from spherical (as is indeed the case with many fruits, sweet peppers included) the contribution of symmetry increases.

3. Material and methods

With the goal of developing a colour-agnostic detection scheme, we turn to visual cues other than colour. For this reason, non-colour information in the image plane is employed, as well as shape information given directly by a depth sensor. In the main (and second) part of our algorithm the entire space is searched for fruit instances, by first detecting the symmetry plane that best describes the data around each location, followed by accumulation of shape features, and the classification of that data with a pre-trained classifier. While accumulating local features and running a classifier are considered relatively fast, the detection of symmetry is somewhat slower. For this reason, a rapid pruning stage is incorporated that processes the 2D visual data and finds locations that are likely to contain fruit based on specular highlights. The areas in the image that are filtered out by this first stage are then ignored by the 3D detector, allowing for faster detection. Both the pruning and the 3D detection make no assumptions about visibility and thus incorporate intrinsic capacity to cope with occlusions. In the next three sections the different parts of our algorithm are described. The pruning algorithm is based on highlight detection, a shape-based detector (that assumes an estimation of symmetry is available), and a symmetry detection process.

3.1. *Highlight detection for data pruning*

Specular highlights tend to exhibit high luminance (V) and low saturation (S) values in digital images. Our highlight detection begins by generating an initial list of candidate regions based on the values of these two channels in the hue saturation value (HSV) representation (Levkowitz, 1997) of the input image I. More specifically, each pixel $x \in I$ is filtered according to

$$f(x) = \begin{cases} 1 & \text{if } V_x/S_x > t \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where V_x and S_x are its corresponding luminance and saturation values. This equation classifies a pixel as a possible candidate (indicated by the value 1) when its luminance to saturation ratio is larger than a certain threshold value t. Otherwise, the pixel is classified as a non-candidate (indicated by the value 0). The parameter t can be determined in a supervised manner and in our experiments it was selected empirically to be 0.8. Once the input was so binarised, the algorithm aggregated pixels to connected components which formed candidate specular regions.

Figure 2 shows the resulting binary representation of candidate specular regions for a sample scene of sweet pepper fruit. Already in this example one can notice that high luminance and low saturation are not exclusive properties of specular highlights. Therefore, the list of initial candidates did not indicate highlights by itself. However, regions filtered out by Eq. (1) were very unlikely to be highlights and thus this preliminary step served mainly to reduce the computational resources required for the next computational step, in which specular highlights were distinguished from among the candidate regions. This computation relied on the approximately isotropic structure that characterises specular highlights. In particular, highlights appear to possess a particular distribution of image gradients that surround a singularity. As can be seen in Fig. 3, these image gradients tended to organise radially around the highlight centre, in what may be considered a local pinwheel (when coded by colour). To find those local pinwheels as more evidence of the existence of a highlight their expected structure was modelled as a group of concentric circles of increasing radii, C_1, C_2, …, C_n. Each circle encoded the expected local orientation values along its perimeter and thus provides a multi-scale signature of the orientation at all discrete angles around the highlight (see Fig. 3). The resultant pattern was the one that was sought in the gradient map of the input image.

In order to detect local orientation (gradient) patterns a reliable orientation map was firstly required from the image. While there are numerous ways of doing so, here the process of Carson, Thomas, Belongie, Hellerstein, and Malik (1999) was followed that uses the eigenvectors of the second moment matrix to produce a two dimensional map encoding the estimated orientation in the range $[0, 2\pi]$ at each location. More formally, for a specific pixel, $p = (x,y)$ the second moment matrix is approximated by

$$S_W(p) = \begin{bmatrix} \sum_r w[r](I_x[p-r])^2 & \sum_r w[r]I_x[p-r]i_y[p-r] \\ \sum_r w[r]I_y[p-r]I_x[p-r] & \sum_r w[r](I_y[p-r])^2 \end{bmatrix} \quad (2)$$

where I_x and I_y are the partial derivatives of the intensity image, r varies over a set of image positions in the spatial neighbourhood W of p, and $w(r)$ is a weight associated with each r. In practice, the matrix components for all pixels were obtained by convolving a Gaussian kernel G_σ with each combination of the partial derivatives. The variance σ of the kernel determines the size of the environment W used to integrate gradients at a specific pixel.

The orientation at p is computed based on the eigenvectors of the matrix $S_W(p)$. These eigenvectors were λ_1 and λ_2, where $\lambda_1 > \lambda_2$. The orientation of gradients on the window used to build the matrix was given by the orientation of λ_1.

Please cite this article in press as: Barnea, E., et al., Colour-agnostic shape-based 3D fruit detection for crop harvesting robots, *Biosystems Engineering* (2016), http://dx.doi.org/10.1016/j.biosystemseng.2016.01.013

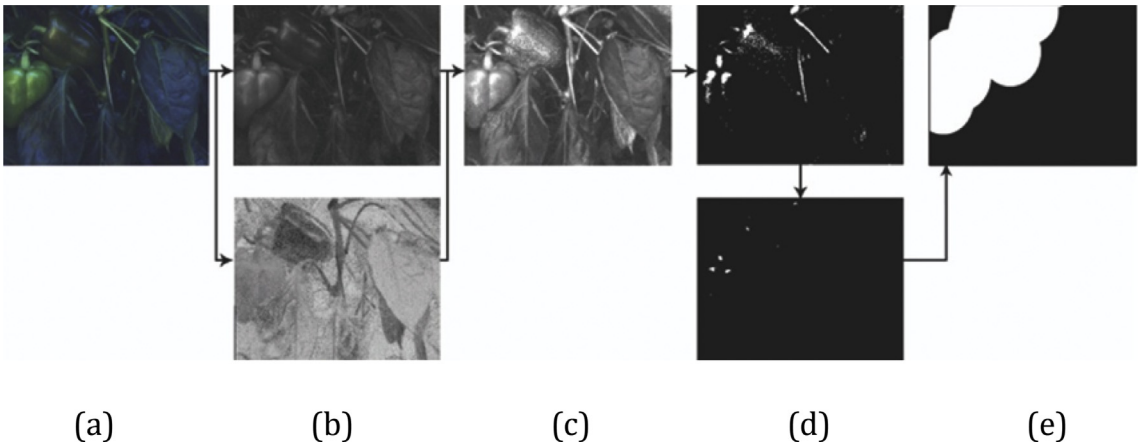


Fig. 2 – Initial selection of candidate regions based on luminance and saturation. (a) An RGB input image with highlight appearing on pepper fruit. (b) The luminance (top) and saturation (bottom) channels are extracted from the HSV representation of the input image. (c) Information from both channels is combined according to Eq. (1) into a map that encodes the luminance saturation ratio at each location. (d) The ratio map is thresholded to obtain a binary representation of candidate regions with high luminance-saturation ratio (top). As the binary image indicates, high luminance and low saturation are not exclusive properties of highlights and non-highlight regions may become candidates as well. From among the candidates, highlights are distinguished based on their structure and represented in a binary indicator function (bottom). (e) The centres of mass of the binary indicators are considered as highlight markers and are dilated by a predefined radius to produce the final pruning mask.

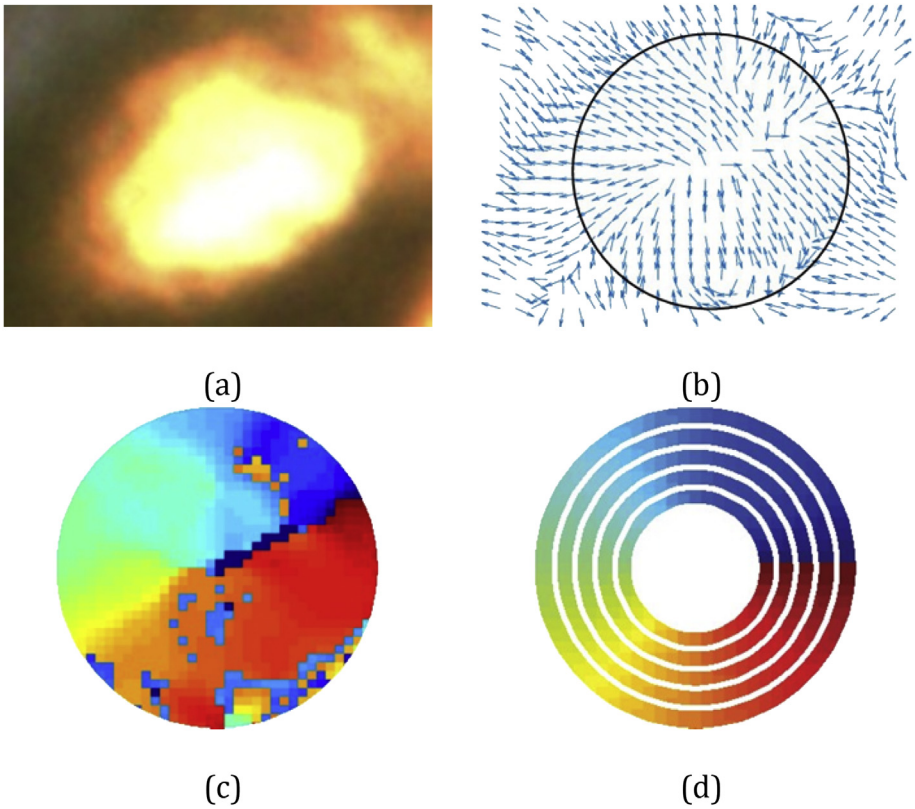


Fig. 3 – The signature distribution of orientation gradients around specular highlights. (a) A close-up of a typical highlight. (b) The typical circular distribution of image gradient in the proximity of the highlight. (c) The same structure, now coded by colour, is reminiscent of a pinwheel. (d) A multi scale model of the pinwheel used for highlight detection.

Please cite this article in press as: Barnea, E., et al., Colour-agnostic shape-based 3D fruit detection for crop harvesting robots, *Biosystems Engineering* (2016), http://dx.doi.org/10.1016/j.biosystemseng.2016.01.013

With both the orientation map and the desired pattern defined, each candidate pixel p was classified by the degree to which the orientation pattern around it matched that of the pinwheel. C_i was a circle of radius r_i from the set of concentric circles, C_1, C_2, …, C_n, that formed our model. For a specific candidate pixel p, C_i induces the set of all pixel locations $C_i(p)$ at radius r_i from p.

The estimated orientation values computed at locations in $C_i(p)$ were o_1, o_2, …, o_m . Each value o_j corresponded with an expected orientation value e_j that is encoded by C_i. In order to measure the agreement between the two corresponding orientations, the periodicity of orientation values must be considered: a shorter distance was expected for orientations that were proximate *radially*, rather than by orientation value (e.g., the distance between 0 and 360 should be 0).

To this end, the distance between o_j and e_j was measured based on their complex number representations: $(cos(o_j) + i \cdot sin(o_j))$ and $(cos(ej) + i \cdot sin(ej))$ respectively. Thus, the distance between o_j, and e_j was computed as follows

$$d(o_j, e_j) = \sqrt{(\cos o_j - \cos e_j)^2 + (\sin o_j - \sin e_j)^2} \quad (3)$$

With this distance measure, the agreement between the measured and expected orientations is determined by similarity up to a given threshold

$$agreement(o_j, e_j) = \begin{cases} 1 & d(o_j, e_j) < T_a \\ 0 & otherwise \end{cases}. \quad (4)$$

Now, each circle C_i was labelled as supporting the hypothesis that p is a highlight by the percentage of corresponding orientations that agree. If that percentage, denoted $S(C_i)$, was larger than a threshold T_s, then C_i voted positively (indicated by the value 1). Otherwise, C_i did not support p (indicated by 0). Formally, this was computed by

$$support(C_i) = \begin{cases} 1 & S(C_i) > T_s \\ 0 & otherwise \end{cases}. \quad (5)$$

Finally, if the accumulated support from all circles C_1, C_2, …, C_n, exceeded a certain threshold T_h, p was considered as a highlight pixel. Otherwise, it was classified as a non-highlight pixel:

$$highlight(p) = \begin{cases} 1 & \sum_i support(C_i) > T_h \\ 0 & otherwise \end{cases}. \quad (6)$$

The result of all these steps was a binary map representation of detected highlight "markers" on fruits. To conclude, these markers were transformed into regions where entire fruits were located. Thus, to produce the final pruning map the centre of mass of each marker was computed and dilated by a fixed radius r pixels according to the maximum observable size of fruits in the environment. Note that this can be easily calibrated for any particular environment using the minimal viewing distance and maximal expect fruit.

3.2. Fruit detection in 3D

To conveniently take advantage of the depth data mentioned in Section 2, the space with a fixed-size 3D box was scanned by sliding over the cloud of points supplied by a SwissRanger 4000 depth camera by Heptagon (acquired from Mesa Imaging), Rüschlikon, Switzerland. This is done in an efficient manner, without considering empty parts of space, places that are too far away from the camera or those containing just a small batch of nearby points. Needless to say this is carried out for those parts of the point cloud that correspond to the image regions that survived the 2D pruning step discussed above. For each box, the best symmetry plane passing through the centre of the box was found, and the features were calculated using the data points inside the box. This was followed by a classification of the resultant feature vector using a support vector machine (SVM) classifier that constructs a classification hyperplane that maximised the distance between the hyperplane and closest example data-points in the high dimensional feature space (Cristianini & Shawe-Taylor, 2000; Chang & Lin, 2011). Since several boxes were usually classified as containing the same object, nearby detections were removed using a non-maximum suppression process. Figure 4 shows this process graphically and the rest of this section discusses its details.

It is assumed that there is a template plane-symmetric object that one seeks to find in the point cloud (in our case, a sweet pepper, but the description is indifferent to the type of object of interest). The heart of the computation was the calculations performed for every 3D box as it slides and scans the point cloud. Firstly, the symmetry plane of the data inside the box was estimated (assuming it indeed contains an object of interest). This inference is described in the next subsection. With this plane computed, the shape inside each 3D box was modelled as a long feature vector relative to the symmetry plane, thus obtaining a partial pose-invariant representation.[1] More specifically, the feature vector was based on histograms of surface normals that are computed and represented relative to the estimated symmetry plane.

It was assumed that there was a reference frame whose origin is the object's centre, and its three orthonormal basis vectors $\mathbf{r}$, $\mathbf{i}$, $\boldsymbol{n_p}$ are such that $\boldsymbol{n_p}$ is the normal of the object's symmetry plane, and vectors $\mathbf{r}$ and $\mathbf{i}$ span the symmetry plane and selected in a particular and consistent way as described later (see Fig. 5a). Seeking a histogram-based representation of the shape inside a 3D box, the surface normal for each point was calculated by fitting a plane to the points in its vicinity and representing these vectors not in the camera coordinate system but in the symmetry-based reference frame just described.

To further leverage the partial pose-invariant power of the approach these normal vectors were chosen using a somewhat unconventional version of two spherical angles as discussed below.

Where $\boldsymbol{p}$ is the centre of the box and $\boldsymbol{x}$ be a point in the point cloud within this box, $\boldsymbol{n_x}$ be the surface normal associated with $\boldsymbol{x}$ and let $\overline{\boldsymbol{n}}_{\boldsymbol{x}}$ and $\overline{\boldsymbol{x}}$ is their projections on the symmetry plane ($\boldsymbol{p}$, $\boldsymbol{n_p}$), Fig. 5b depicts the first angle in our

[1] Note that a representation based on a symmetry plane is invariant to pose only partially, since transformations of the objects that keeps its symmetry plane in tact will no change this transformation. Thus, ambiguity remains for all changes of pose induced by rotation of the object about an axis perpendicular to the symmetry plane.

Please cite this article in press as: Barnea, E., et al., Colour-agnostic shape-based 3D fruit detection for crop harvesting robots, *Biosystems Engineering* (2016), http://dx.doi.org/10.1016/j.biosystemseng.2016.01.013

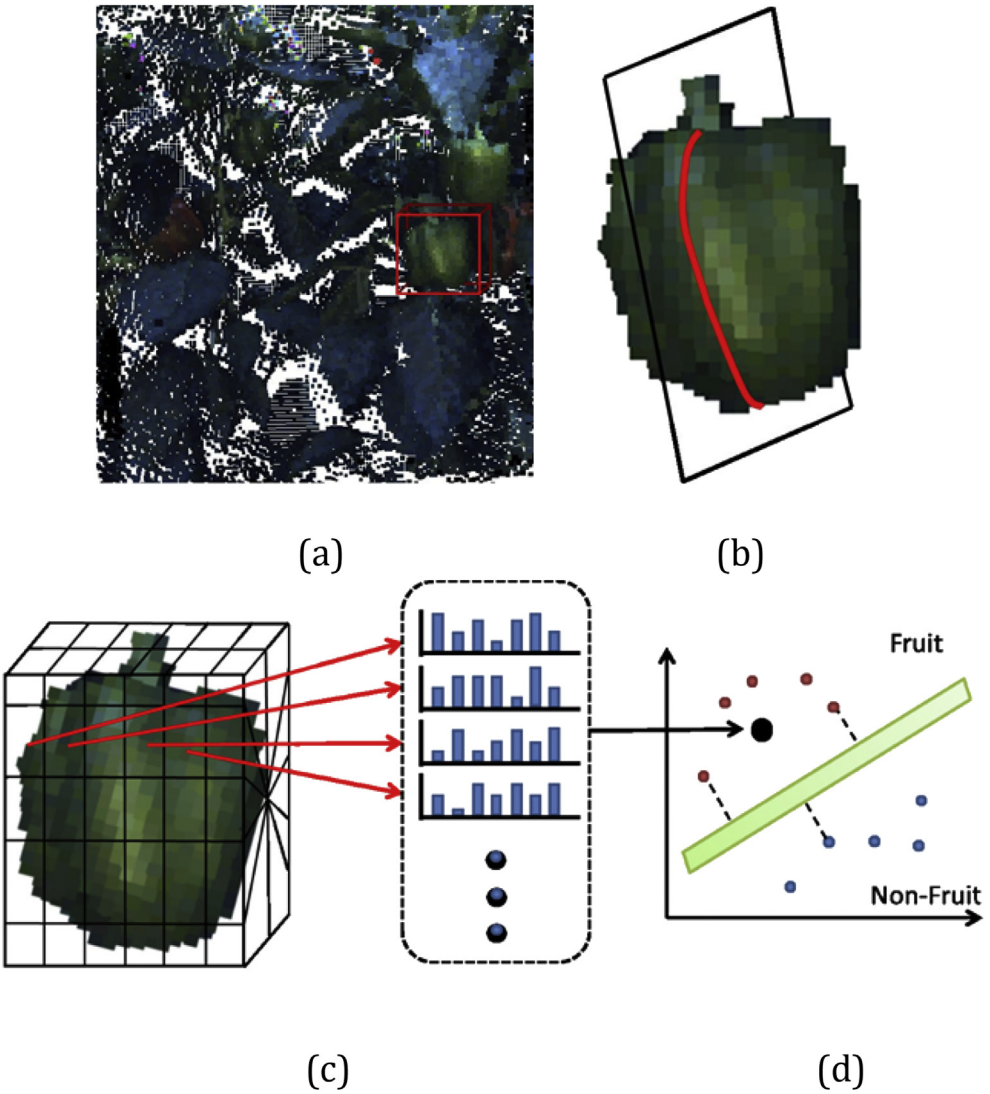


Fig. 4 – An overview of the 3D detection process of fruit in the point cloud provided by the range camera. (a) A typical point cloud from the RGB-D sensor and the sliding box positioned around one measurement (in this case, one that contains a fruit). (b) A close up of the content of the box with the best symmetry plane detected. (c) Features are calculated to represent the content of the box in the feature space. (d) The training set (red for fruit and blue for non-fruit) is used to construct a Support Vector Machine (SVM) classifier in feature space, with which the query feature vector is classified, in this case, as a fruit.

representation which is denote as $\theta \in [0, \pi]$ and defined as the angle between the surface normal $\boldsymbol{n}_x$ and plane normal $\boldsymbol{n}_p$:

$$\theta = \cos^{-1}(\boldsymbol{n}_x \cdot \boldsymbol{n}_p). \tag{7}$$

Figure 5c depicts the second angle $\varphi \in [0, 2\pi)$ in the representation, defined as the signed angle between the projected normal $\overline{\boldsymbol{n}}_x$ and the vector connecting the box centre $\boldsymbol{p}$ with the projected point $\overline{\boldsymbol{x}}$:

$$\phi = \cos^{-1}\left(\frac{\overline{n}_x}{\|\overline{n}_x\|} \cdot \frac{p - \overline{x}}{\|p - \overline{x}\|}\right). \tag{8}$$

This was followed by an addition of π depending on the direction of $\overline{\boldsymbol{n}}_x$ relative to direction vector:

$$direction = \boldsymbol{n}_p \cdot \frac{\boldsymbol{p} - \overline{\boldsymbol{x}}}{\|\boldsymbol{p} - \overline{\boldsymbol{x}}\|}. \tag{9}$$

These two angles provided a representation that depended only on the estimated symmetry plane. Indeed, both Fig. 5b, c and the equations above indicate that both θ and φ of each point can be determined solely by from $\boldsymbol{n}_p$, $\mathbf{x}$, and the centre of the box $\boldsymbol{p}$.

With a scheme to represent the normal vector of each point in the cloud in a symmetry plane-dependent way, the following step is the accumulation of surface normals using histograms. To do so, the space of a 3D box was divided into several bins, each of which were represented by a histogram of their own. A

Please cite this article in press as: Barnea, E., et al., Colour-agnostic shape-based 3D fruit detection for crop harvesting robots, *Biosystems Engineering* (2016), http://dx.doi.org/10.1016/j.biosystemseng.2016.01.013

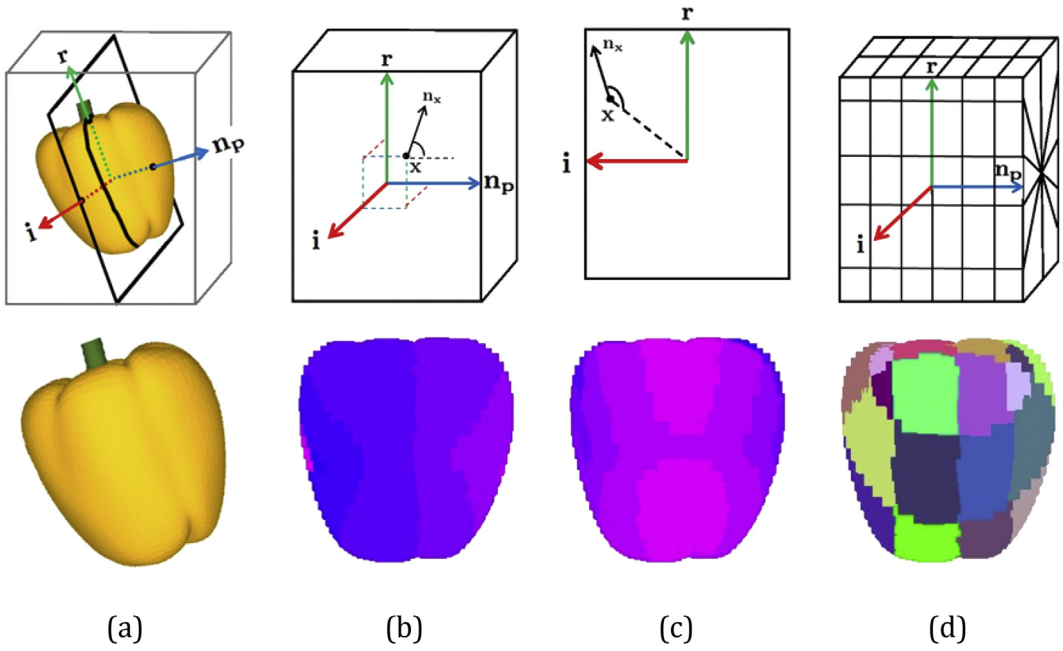


Fig. 5 – The representation of point clouds inside the sliding box is done relative to the symmetry plane and the corresponding reference frame of the points inside it. The first row illustrates the different vectors, angles, and bins used. The second row provides a colour coded visualisation of these quantities calculated for the synthetic pepper and reference frame in the left panel. (a) The input for the representation is a set of points inside a box (depicted here as the sweet pepper object), together with a reference frame induced by the (estimated) object's symmetry plane. (b) The basic representation unit is the surface normal vector associated with each point in the cloud and computed by fitting a plane to the points in its vicinity. These vectors are then represented relative to the reference frames, an operation depicted here by rectifying the frame to an upright position. In particular, all normal vectors are represented by two spherical angles, where θ shown here is the horizontal angle with colour ranging from purple to blue as the angle increases (refer to the text for details). (c) A depiction of the vertical angle φ with colour ranging from blue to purple as the angle increases. Note that x and $\bar{x}$ coincide from this viewpoint. (d) The spatio-angular bins that divide the box. The colour map depicts how different subsets of points on the pepper are associated with different bin.

box was divided into a 2D set of angular bins (Fig. 5d), where each bin was defined according to two parameters - its Euclidean distance from the symmetry plane, and angular distance from the (arbitrarily chosen) basis vector $\boldsymbol{r}$.

The surface normals of all the points that fall in the same bin were accumulated in two 1D histograms according to the two angles described above. All the histograms were normalised and concatenated to form a single feature vector that was then classified using the C-SVC formulation of SVM with an RBF-kernel (Chang & Lin, 2011).

To train the SVM classifier the used dataset was split into a training and a testing set Bishop (2007). In order to generate false and positive training examples, feature vectors were computed for boxes containing fruit and boxes that did not contain fruit (randomly chosen from locations without fruit). These two classes of feature vectors were then used to train the classifier.

3.3. *3D plane-reflection symmetry detection*

The computational procedure above assumed a symmetry plane was at hand in order to construct the representation, the training, and the classification. In practice, this plane should be estimated from the measured data in each box (Fig. 6).

In fact, taking into account every relevant box in space greatly simplifies the symmetry plane estimation task. If it

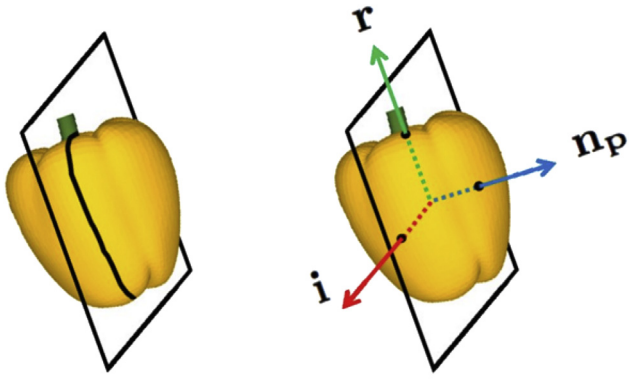


Fig. 6 – The symmetry plane and reference frame of a synthetic pepper. The vector n_p is the normal of the plane (blue), r points up (green), and i is the cross product of the two vectors (red).

Please cite this article in press as: Barnea, E., et al., Colour-agnostic shape-based 3D fruit detection for crop harvesting robots, *Biosystems Engineering* (2016), http://dx.doi.org/10.1016/j.biosystemseng.2016.01.013

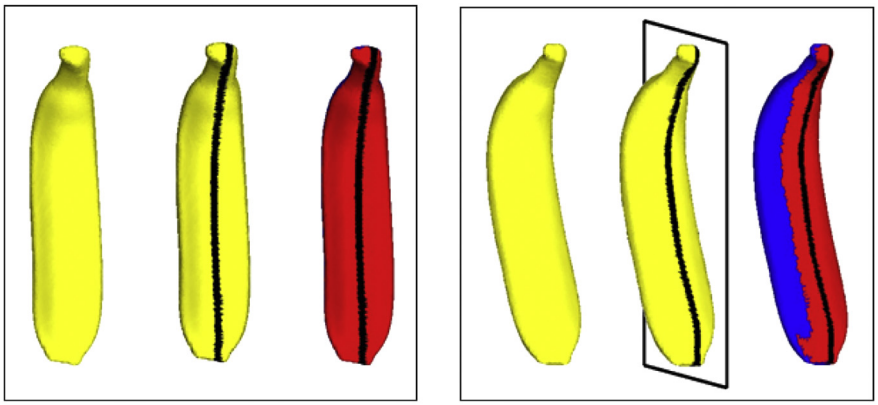


Fig. 7 – Symmetry planes and visible symmetric partners. For each of the two cases we show a banana imaged from a particular point of view (left), estimated symmetry plane (middle, depicted by the intersection of the plane with the banana), and a colour map of points with (red) and without (blue) symmetric partners. (a) A banana imaged frontally has virtually all of its points in the range data possess symmetric partners. Note how all points are marked red. (b) A banana imaged obliquely has the same symmetry plane (now shown rotated, of course) but an important subset of its points has no visible symmetric partners in the cloud (shown in blue).

contains a fruit, most points inside the box are likely to belong to that object, while the number of outliers is usually not too great; a property that distinguishes range data from intensity/colour data. More importantly, scanning the entire space, provided assurance that some box will have a centre point that coincides with the sought-after symmetry plane. Since a plane can be represented with a point and a normal, only the normal remains to be solved for. Be that as it may, assuming that points are generated from a perspective imaging device from a single viewpoint, the corresponding symmetric counterpart of many inlier points (or even all of them) is simply not visible. These points were first identified following a scoring strategy that ranks every possible reflection plane normal. For that purpose, the two angles comprising the normal's spherical representation[2] were quantised and the best pair was chosen using a score penalising point pairs that are spatially symmetric (relative to the candidate plane) while having non-symmetric surface normals (in contrast to Thrun and Wegbreit (2005)). The formal details follow below.

Prior to calculating a score for a candidate symmetry plane, the inlier points without symmetric partners were dealt with. Self-occlusion dictates that when observing an object from one side of the symmetry plane, most of the visible inlier points that are observed will be the ones that share the same side with the camera. For the same reason, inlier points that are observed on the other side should all have visible symmetric points on the camera's side (as shown by the banana in Fig. 7). Knowing this, these points were found on the closer side of the candidate plane with no partners on the farther side and excluded from the score calculation. To do so, surface normals were used, observing that a point $\boldsymbol{x}$ on a surface with an estimated surface normal $\boldsymbol{n}_x$ is visible from viewpoint $\boldsymbol{p}_v$ if

$$\boldsymbol{n}_x \cdot (\boldsymbol{p}_v - \boldsymbol{x}) > 0. \tag{10}$$

Therefore, points whose symmetric partner had a surface normal that points away from the camera were sought. A point $\boldsymbol{x}$ with estimated normal $\boldsymbol{n}_x$ was reflected over a candidate symmetry plane with centre point $\boldsymbol{p}$ and normal $\boldsymbol{n}_p$ by:

$$\tilde{\boldsymbol{x}} = \boldsymbol{x} - 2 \cdot \boldsymbol{n}_p \cdot d_x, \tag{11}$$

where d_x is the signed distance between the point $\boldsymbol{x}$ and the plane. Correspondingly, $\boldsymbol{x}$'s normal was reflected as well by:

$$\tilde{\boldsymbol{n}}_x = \boldsymbol{n}_x - 2 \cdot \boldsymbol{n}_p \cdot d_n, \tag{12}$$

where $\boldsymbol{n}_x$ is the normal we wish to reflect and d_n is the signed distance between the normal's head and the candidate plane, centred at the camera's axes origin with normal $\boldsymbol{n}_p$. Thus, $\boldsymbol{x}$ has no symmetric partner if:

$$\tilde{\boldsymbol{n}}_x \cdot (\boldsymbol{p}_v - \tilde{\boldsymbol{x}}) \leq 0. \tag{13}$$

Following that, each point was assigned a *point reflection score* that measured the "wellness" of its reflection. An observed point $\mathbf{y}$ with normal $\boldsymbol{n}_y$ closest to $\tilde{\boldsymbol{x}}$ (and in the same side) was found using a kd-tree data structure and the score of $\boldsymbol{x}$ was determined by:

$$x_{score} = d + w \cdot \alpha, \tag{14}$$

where d is the distance between $\tilde{\boldsymbol{x}}$ and $\mathbf{y}$, α is the angle between $\tilde{\boldsymbol{n}}_y$ and $\boldsymbol{n}_y$ (see Fig. 8), and w is a weighting factor. A lower value implied good symmetry and the best plane was chosen as the one that minimised the mean score of all the contributing points. In order to have a complete reference frame for the symmetry plane $\boldsymbol{n}_p$ was endowed with another unit length

[2] A normal's magnitude is always 1 and thus only its direction should be represented.

Please cite this article in press as: Barnea, E., et al., Colour-agnostic shape-based 3D fruit detection for crop harvesting robots, *Biosystems Engineering* (2016), http://dx.doi.org/10.1016/j.biosystemseng.2016.01.013

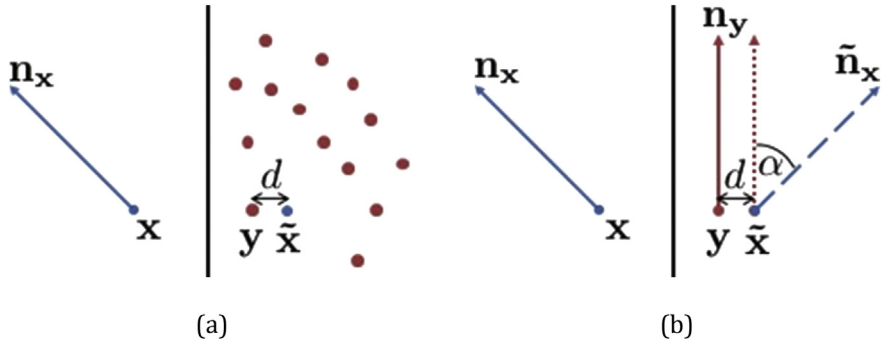


Fig. 8 – Determining a reflection score for point *x* using its reflection $\tilde{x}$, the closest point *y* is found (a), then the point is scored according to the distance *d* and normal difference α (b).

reference vector $\boldsymbol{r}$ that lay on the plane. It was chosen to be on the verge of visibility according to Eq. (10), and to be directed upwards. Summarising these constraints, we get:

1. $\|\boldsymbol{r}\| = 1$ ($\boldsymbol{r}$ is of unit length)
2. $\boldsymbol{r} \cdot (\boldsymbol{p}_v - \boldsymbol{p}) = 0$ ($\boldsymbol{r}$ is on the verge of visibility)
3. $\boldsymbol{r} \cdot \boldsymbol{n}_p = 0$($\boldsymbol{r}$ is on the symmetry plane)
4. $\boldsymbol{r} \cdot [0,1,0] \geq 0$ ($\boldsymbol{r}$ points up)

Therefore, $\boldsymbol{r}$ can be calculated with:

$$\boldsymbol{r} = \boldsymbol{n}_p \times \frac{\boldsymbol{p}_v - \boldsymbol{p}}{\|\boldsymbol{p}_v - \boldsymbol{p}\|}, \tag{15}$$

and the reference frame can be completed by calculating the third orthonormal vector:

$$\boldsymbol{i} = \boldsymbol{n}_p \times \boldsymbol{r}. \tag{16}$$

An illustration of the estimated symmetry together with a complete reference frame is shown in Fig. 6.

4. Results and discussion

To evaluate the performance of our developed fruit detection system it was tested on an RGB-D dataset of green and red bell peppers taken in a greenhouse. The RGB-D data was generated by a SwissRanger depth camera registered with a simple RGB camera. The dataset consists of 88 realistic images taken in a conventional greenhouse in the Netherlands during the harvesting season (see Fig. 1) with the cameras placed about 500–600 mm from the foliage, at different heights from the ground, and facing in a direction perpendicular to it. No changes were made to the peppers or foliage apart from the normal treatment given by the grower.

For the test, 581 of the peppers with visibility 50% or more (as judged by a human observer) were considered. The sweet peppers selected were labelled for position in both the RGB and the range data by a human expert in order to facilitate both training and quantitative testing. The work was done in the context of the cRops harvesting robot, as shown in Fig. 9.

Before discussing the results, it should be emphasised that fundamentally the proposed approach is a computational process whose components may be plugged-and-played in various ways, replaced by other algorithms, or even removed completely. The pruning stage, for example, was designed for both computational efficiency (by quickly discarding image locations if they are unlikely to contain fruits) and for reducing the probability of false positives by the 3D detector (by eliminating data where the 3D detector may wrongfully detect targets). It is proposed to use highlights, but if other visual cues (including colour) are readily available they may be used too (or instead). Similarly, the symmetry estimation method may also be replaced by another method that quickly estimates a consistent reference frame for the object, or it may even be replaced by a blind method that ignores it completely and always returns a fixed reference frame.

The latter approach may be particularly appealing if the input range data is so noisy that symmetry estimation becomes chaotic. Indeed, in order to understand the significance of each such phase, this detector was evaluated with and without both of these modules.

In order to empirically evaluate the detector, a common method was followed that is used in general computer vision challenges such as the PASCAL VOC (Everingham, Van Gool,

Fig. 9 – The cRops gripper and manipulator, shown here with the sensory rig that includes both RGB and range cameras. Registration was done in software and provided RGB-D point cloud used for the analyses and experimental evaluations discussed in this paper.

Please cite this article in press as: Barnea, E., et al., Colour-agnostic shape-based 3D fruit detection for crop harvesting robots, *Biosystems Engineering* (2016), http://dx.doi.org/10.1016/j.biosystemseng.2016.01.013

Williams, Winn, & Zisserman, 2010). The dataset images were randomly divided into two sets, where the first set is used for training (including cross validation) and the second for testing. This process was repeated several times (in our case, four) and performance was averaged to report mean performance. For each random split of the data, the detector was evaluated using the following procedure. Firstly, the detector was executed on all the test images and a list of all detections was saved. These results were then validated against the manually labelled "ground truth" data mentioned above in order to mark each detection as true or false positive according to its distance from a "ground truth" target. In our case a detection was defined as a true positive if its centre was no further than 50 mm from the centre of a ground truth sweet pepper. Clearly, this tolerance should depend on the application and can be relaxed or tightened accordingly. Finally, a precision-recall curve was generated, from which the average precision (AP) measure was calculated, serving as the final score. Precision was defined as the fraction of true positives from the returned detection hypotheses, Recall was defined as the fraction of true positives from all the positives, and Average Precision (AP) was defined as the mean of maximal precision values at a set of eleven equally spaced intervals of the recall axis (Everingham et al., 2010). The entire process was slightly contaminated by correct detections of highly occluded fruits that were not labelled as "ground truth" sweet peppers because their visibility was <50%.

Table 1 lists the mean average precision for each variation of our algorithm, while Fig. 10 shows the corresponding average precision-recall curves. As can be seen, the best combination was the one including symmetry detection and pruning based on highlights. While the differences may appear small, they represent average performance for all recall values, including those for unrealistic recall values where precision essentially vanishes and the detector practically considers everything as a positive. Compared to state-of-the-art in object recognition from range data (Janoch et al., 2011) this performance was superior, especially when considering the significant level of complexity of agricultural data compared to typical indoor scenes used in existing computer vision studies.

As mentioned above, to assess the contribution of each component in our pipeline we processed the results both with and without it. The results with pruning are always better than those without it, and when pruning is included the contribution of symmetry is very significant. Somewhat unexpectedly, when pruning was not included (and this the 3D detector processed all range data) the inclusion of symmetry detection in the 3D detector provided no improvement.

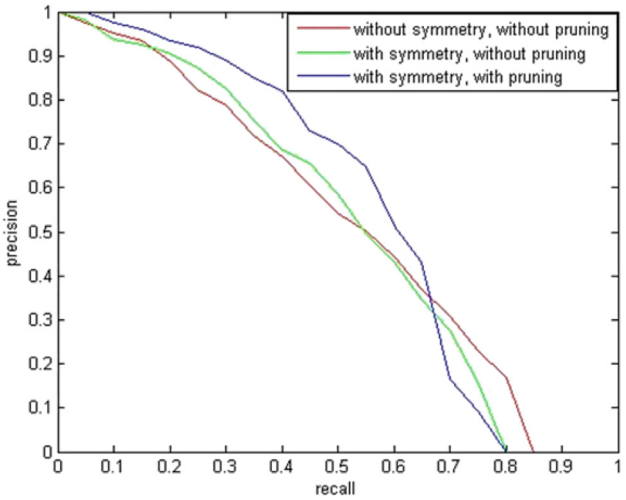


Fig. 10 – The average precision-recall curves of the different computational combinations tested on the labelled dataset. The curve is generated by examining different thresholds of the detection confidence supplied by the SVM classifier. At high confidence thresholds, only the most confident detections are returned, yielding a very high precision but a small recall. Lowering the threshold increases the recall but introduces more false positives and so decreases precision. The average curve is simply the average of precision values of different curves at the same levels of recall. Note how for most recall values (≤0.7) the precision when using both pruning and symmetry is improved significantly.

Visually examining the detected symmetry planes, many examples revealed that the SwissRanger depth camera provided very noisy measurements under real conditions in greenhouses and that without highlight pruning from the RGB image much higher false positive rates were sustained. With better range sensors, the results are expected to provide much improvement even if pruning is not used (though there is little incentive to do so). To verify this hypothesis, an experiment was set up to remove the effects of such noise but remain as close as possible to the original dataset. An image containing foliage, but with no visible peppers was used, to which was added several instances of a synthetic peppers inside the foliage. The synthetic peppers, generated from a 3D model, were placed in different poses,

Table 1 – Mean average precision for different component combinations in our detector. Values are rounded to two digits after the decimal point.

Mean average precision	Without symmetry	With symmetry
without pruning	0.52	0.51
with pruning	0.53	**0.55**

Bold value emphasizes the evaluation result of the algorithm including all suggested components.

Table 2 – Average run-times in seconds for different component combinations in out detector.

Average run-time (s)	Without symmetry	With symmetry
without pruning	152	345
with pruning	91	197

The times were calculated on a 32-bit system, with 3 GB of RAM, and an Intel® Core™ i5760 processor (2.8 GHz). Performance is not real time but at this research stage no code optimisation, parallelism, or the use of GPU were employed.

Please cite this article in press as: Barnea, E., et al., Colour-agnostic shape-based 3D fruit detection for crop harvesting robots, *Biosystems Engineering* (2016), http://dx.doi.org/10.1016/j.biosystemseng.2016.01.013

with the non-visible part (the part occluded from the range sensor) removed. The algorithm was tested with and without symmetry detection (and without pruning in both cases) and the average precision was calculated for both cases. This time the algorithm incorporating symmetry indeed improved the performance, effectively doubling the mean average precision from 0.027 to 0.043 and confirming our hypothesis that beyond a certain level of noise (as was the case with the SwissRanger depth camera) the ability to detect the symmetry plane reliably is diminished.[3]

The system was implemented in C++, using the Point Cloud Library (Rusu & Cousins, 2011) and OpenCV (Bradski, 2000). The average run-times of the different component combinations in the detector can be seen in Table 2.

5. Conclusions

A fruit detection algorithm has been presented that is agnostic to colour and provides much robustness to shape variations and occlusions. It can be used for various tasks with agricultural robots where accurate localisation in space is required but specifically for automatic grasping and harvesting of fruit. In order to provide this novel functionality our algorithm consists of a rapid pruning phase based on visual highlights that are detected by seeking their prototypical signature on image gradients. This is then followed by local symmetry detection in range data, and the representation of shape features relative to this detected symmetry in order to obtain partial post invariance. Finally, a classifier is used to classify measured data into fruits or background (e.g., foliage) in the presence of occlusions, where the estimated symmetry provides partial pose estimation as well. This approach was evaluated on a challenging sweet pepper dataset in conjunction with a real research platform for selective harvesting (FP7 cRops project). It showed how the combination of pruning and symmetry estimation improves upon standard classifiers that do not include these components.

Acknowledgements

We wish to thank the many partners of the cRops project for their help in collecting the field data and making the sensors and robotic manipulator and grippers available to us. This research was funded the European Commission in the 7th Framework Programme (CROPS GA no. 246252). We also thank the generous support of the Frankel fund at the Computer Science department and the Helmsley Charitable Trust through the Agricultural, Biological and Cognitive Robotics Initiative, both at Ben-Gurion University of the Negev.

[3] We note that the results of the control test are not comparable to those in Table 1, as the detectors in this experiments were trained on a much smaller subset of the data (where *pose* ground truth, a much more difficult task for human annotators was carried out also) and tested on a single image with several synthetic peppers.

REFERENCES

Adato, Y., & Ben-Shahar, O. (2011). Specular flow and shape in one shot. In *British Machine Vision Conference* (pp. 1–11).
Adato, Y., Vasilyev, Y., Zickler, T., & Ben-Shahar, O. (2010). Shape from specular flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 32*(11), 2054–2070.
Aldoma, A., Vincze, M., Blodow, N., Gossow, D., Gedikli, S., Rusu, R., et al. (2011). Cad-model recognition and 6DOF pose estimation using 3D cues. In *International Conference on Computer Vision Workshops (ICCV Workshops)* (pp. 585–592).
Barnea, E., & Ben-Shahar, O. (2014). Depth based object detection from partial pose estimation of symmetric objects. In *Proceedings of the European Conference on Computer Vision* (pp. 377–390).
Beckmann, P., & Spizzichino, A. (1963). *The scattering of electromagnetic waves from rough surfaces*. Pergamon Press.
Bishop, C. M. (2007). *Pattern recognition and machine learning*. New York: Springer-Verlag.
Bo, L., Lai, K., Ren, X., & Fox, D. (2011). Object recognition with hierarchical kernel descriptors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1729–1736).
Bradski, G. (2000). The opencv library. *Dr. Dobb's Journal of Software Tools, 25*, 120–126.
Carson, C., Thomas, M., Belongie, S., Hellerstein, J., & Malik, J. (1999). Blobworld: a system for region-based image indexing and retrieval. In *Visual information and information systems* (pp. 509–517). Springer.
Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology, 2*(27), 1–27.
Chi, Y., & Ling, P. (2004). Fast fruit identification for robotic tomato picker. In *ASAE Annual International Meeting* (p. 1). American Society of Agricultural and Biological Engineers.
Cristianini, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods*. New York, NY, USA: Cambridge University Press.
Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 886–893).
Do Carmo, M. P., & Do Carmo, M. P. (1976). *Differential geometry of curves and surfaces* (Vol. 2). Englewood Cliffs: Prentice-hall.
Edan, Y., Rogozin, D., Flash, T., & Miles, G. (2000). Robotic melon harvesting. *IEEE Transactions on Robotics and Automation, 16*, 831–835.
Everingham, M., Van Gool, L., Williams, C., Winn, J., & Zisserman, A. (2010). The Pascal visual object classes (voc) challenge. *International Journal of Computer Vision, 88*, 303–338.
Felzenszwalb, P., McAllester, D., & Ramanan, D. (2008). A discriminatively trained, multiscale, deformable part model. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1–8).
Hannan, M., & Burks, T. (2004). Current developments in automated citrus harvesting. In *ASAE Annual International Meeting*.
Harrell, R., Slaughter, D., & Adsit, P. (1989). A fruit-tracking system for robotic harvesting. *Machine Vision and Applications, 2*, 69–80.
Hinterstoisser, S., Holzer, S., Cagniart, C., Ilic, S., Konolige, K., Navab, N., et al. (2011). Multimodal templates for real-time detection of textureless objects in heavily cluttered scenes. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 858–865).
Janoch, A., Karayev, S., Jia, Y., Barron, J., Fritz, M., Saenko, K., et al. (2011). A category-level 3-D object dataset: putting the Kinect to work. In *Consumer depth cameras for computer vision* (pp. 141–165). London: Springer.
Jimenez, A., Ceres, R., & Pons, J. (2000a). A survey of computer vision methods for locating fruit on trees. *Transactions of the ASAE-American Society of Agricultural Engineers, 43*(6), 1911–1920.

Please cite this article in press as: Barnea, E., et al., Colour-agnostic shape-based 3D fruit detection for crop harvesting robots, *Biosystems Engineering* (2016), http://dx.doi.org/10.1016/j.biosystemseng.2016.01.013

Jimenez, A., Ceres, R., & Pons, J. (2000b). A vision system based on a laser range-finder applied to robotic fruit harvesting. *Machine Vision and Applications, 11*, 321–329.

Kapach, K., Barnea, E., Mairon, R., Edan, Y., & Ben-Shahar, O. (2012). Computer vision for fruit harvesting robots – state of the art and challenges ahead. *International Journal of Computer Vision and Robotics, 3*, 4–34.

Kim, B., Xu, S., & Savarese, S. (2013). Accurate localization of 3D objects from RGB-D data using segmentation hypotheses. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 886–893).

Lai, K., Bo, L., Ren, X., & Fox, D. (2011). A large-scale hierarchical multi-view RGB-D object dataset. In *Proceedings of the IEEE International Conference on Robotics and Automation* (pp. 1817–1824).

Lai, K., Bo, L., Ren, X., & Fox, D. (2012). Detection-based object labeling in 3D scenes. In *Proceedings of the IEEE International Conference on Robotics and Automation* (pp. 1330–1337).

Lee, S., & Bajcsy, R. (1992). Detection of specularity using color and multiple views. In *Proceedings of the European Conference on Computer Vision* (pp. 99–114).

Lee, S., & Liu, Y. (2012). Curved glide-reflection symmetry detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 34*, 266–278.

Levkowitz, H. (1997). *Color theory and modeling for computer graphics, visualization, and multimedia applications* (Vol. 402). US: Springer Science & Business Media. Springer.

Lin, Z., & Davis, L. (2008). A pose-invariant descriptor for human detection and segmentation. In *Proceedings of the European Conference on Computer Vision* (pp. 423–436).

Lin, S., Li, Y., Kang, S., Tong, X., & Shum, H. (2002). Diffuse-specular separation and depth recovery from image sequences. In *Computer Vision European Conference on Computer Vision* (pp. 210–224). Springer.

Lin, S., & Shum, H. (2001). Separation of diffuse and specular reflection in color images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Vol. 1, pp. I–341). IEEE.

Loy, G., & Eklundh, J. (2006). Detecting symmetry and symmetric constellations of features. In *Proceedings of the European Conference on Computer Vision* (pp. 508–521).

Mallick, S., Zickler, T., Belhumeur, P., & Kriegman, D. (2006). Specularity removal in images and videos: a PDE approach. In *Proceedings of the European Conference on Computer Vision* (pp. 550–563). Springer.

Mallick, S., Zickler, T., Kriegman, D., & Belhumeur, P. (2005). Beyond Lambert: reconstructing specular surfaces using color. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Vol. 2, pp. 619–626).

Mitra, N., Pauly, M., Wand, M., & Ceylan, D. (2012). Symmetry in 3D geometry: extraction and applications. *Computer Graphics Forum, 32*(6), 1–23.

Monta, M., & Namba, K. (2003). Three-dimensional sensing system for agricultural robots. In *Proceedings of the IEEE International Conference on Advanced Intelligent Mechatronics* (Vol. 2, pp. 1216–1221).

Nayar, S. K., Fang, X.-S., & Boult, T. (1997). Separation of reflection components using color and polarization. *International Journal of Computer Vision, 21*, 163–186.

Netz, A., & Osadchy, M. (2011). Using specular highlights as pose invariant features for 2D-3D pose estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 721–728). IEEE.

Osadchy, M., Jacobs, D., & Ramamoorthi, R. (2003). Using specularities for recognition. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 1512–1519). IEEE.

Park, M., Lee, S., Chen, P., Kashyap, S., Butt, A., & Liu, Y. (2008). Performance evaluation of state-of-the-art discrete symmetry detection algorithms. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1–8).

Park, J., & Tou, J. (1990). Highlight separation and surface orientations for 3-D specular objects. In *Pattern recognition* (Vol. 1, pp. 331–335). IEEE.

Podolak, J., Shilane, P., Golovinskiy, A., Rusinkiewicz, S., & Funkhouser, T. (2006). A planar-reflective symmetry transform for 3D shapes. *ACM Transactions on Graphics, 25*, 549–559.

Raviv, D., Bronstein, A., Bronstein, M., & Kimmel, R. (2007). Symmetries of non-rigid. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 1–7).

Redondo-Cabrera, C., López-Sastre, R., Acevedo-Rodriguez, J., & Maldonado-Bascón, S. (2012). Surfing the point clouds: selective 3D spatial pyramids for category-level object recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3458–3465).

Rosen, J. (2011). *Symmetry discovered – Concepts and applications in nature and science*. Dover Publications.

Rusu, R. B., & Cousins, S. (2011). 3D is here: Point Cloud Library (PCL). In *Proceedings of the IEEE International Conference on Robotics and Automation* (pp. 1–4).

Saenko, K., Karayev, S., Jia, Y., Shyr, A., Janoch, A., Long, J., et al. (2011). Practical 3-D object detection using category and instance-level appearance models. In *IEEE International Workshop on Intelligent Robots and Systems* (pp. 1817–1824).

Sato, Y., & Ikeuchi, K. (1994). Temporal-color space analysis of reflection. *Journal of the Optical Society of America A, 11*, 2990–3002.

Shen, H., & Cai, Q. (2009). Simple and efficient method for specularity removal in an image. *Applied Optics, 48*, 2711–2719.

Shotton, J., Sharp, T., Kipman, A., Fitzgibbon, A., Finocchio, M., Blake, A., et al. (2013). Real-time human pose recognition in parts from single depth images. *Communication of the ACM, 56*, 116–124.

Silberman, N., & Fergus, R. (2011). Indoor scene segmentation using a structured light sensor. In *IEEE International Conference on Computer Vision Workshops (ICCV Workshops)* (pp. 601–608).

Spinello, L., & Arras, K. (2011). People detection in RGB-D data. In *IEEE International Workshop on Intelligent Robots and Systems* (pp. 3838–3843).

Tang, S., Wang, X., Lv, X., Han, T., Keller, J., He, Z., et al. (2012). Histogram of oriented normal vectors for object recognition with a depth sensor. In *Proceedings of the Asian Conference on Computer Vision* (pp. 525–538). Springer Berlin Heidelberg.

Tan, R., & Ikeuchi, K. (2005a). Reflection components decomposition of textured surfaces using linear basis functions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Vol. 1, pp. 125–131). IEEE.

Tan, R., & Ikeuchi, K. (2005b). Separating reflection components of textured surfaces using a single image. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 27*, 178–193.

Tan, P., Quan, L., & Lin, S. (2006). Separation of highlight reflections on textured surfaces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Vol. 2, pp. 1855–1860). IEEE.

Thrun, S., & Wegbreit, B. (2005). Shape from symmetry. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 1824–1831).

Vasilyev, Y., Zickler, T., Gortler, S., & Ben-Shahar, O. (2011). Shape from specular flow: is one flow enough?. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2561–2568).

Wertheimer, M. (1923). Untersuchungen zur lehre von der gestalt. *Psychologische Forschung, 4*, 301–350.

Zhao, P., & Quan, L. (2011). Translation symmetry detection in a fronto-parallel view. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1009–1016).

Please cite this article in press as: Barnea, E., et al., Colour-agnostic shape-based 3D fruit detection for crop harvesting robots, *Biosystems Engineering* (2016), http://dx.doi.org/10.1016/j.biosystemseng.2016.01.013