Computer vision for fruit harvesting robots – state of the art and challenges ahead

Keren Kapach

Department of Industrial Engineering and Management,
Ben-Gurion University,
P.O. Box 653, Beer-Sheva, 84105, Israel
E-mail: kapach@bgu.ac.il

Ehud Barnea and Rotem Mairon

Computer Science Department,
Ben-Gurion University,
P.O. Box 653, Beer-Sheva, 84105, Israel
E-mail: barneaeh@cs.bgu.ac.il
E-mail: rotemra@cs.bgu.ac.il

Yael Edan

Department of Industrial Engineering and Management,
Ben-Gurion University,
P.O. Box 653, Beer-Sheva, 84105, Israel
E-mail: yael@bgu.ac.il

Ohad Ben-Shahar*

Computer Science Department,
Ben-Gurion University,
P.O. Box 653, Beer-Sheva, 84105, Israel
E-mail: ben-shahar@cs.bgu.ac.il
*Corresponding author

Abstract: Despite extensive research conducted in machine vision for harvesting robots, practical success in this field of agrobotics is still limited. This article presents a comprehensive review of classical and state-of-the-art machine vision solutions employed in such systems, with special emphasis on the visual cues and machine vision algorithms used. We discuss the advantages and limitations of each approach and we examine these capacities in light of the challenges ahead. We conclude with suggested directions from the general computer vision literature which could assist our research community meet these challenges and bring us closer to the goal of practical selective fruit harvesting robots.

Keywords: agricultural computer vision; agrovision; agrobotics; fruit harvesting robots.

Reference to this paper should be made as follows: Kapach, K., Barnea, E., Mairon, R., Edan, Y. and Ben-Shahar, O. (2012) 'Computer vision for fruit harvesting robots – state of the art and challenges ahead', *Int. J. Computational Vision and Robotics*, Vol. 3, Nos. 1/2, pp.4–34.

Biographical notes: Keren Kapach received her Masters degree (MSc) in 2007 and Bachelors degree (BSc) in Electrical and Computer Engineering in 2004, both from Ben-Gurion University, BeerSheva, Israel, where she is currrenly a PhD candidate. Her research interests include performance characterisation in machine vision applications and shape modelling.

Ehud Barnea is studying towards his MSc in the Department of Computer Science at Ben Gurion University of the Negev, where he completed his BSc studies. His research is in computer vision with a focus on target detection with shape prior.

Rotem Mairon graduated from the department of Computer Science in Ben Gurion University of the Negev in 2011, where he currently pursues his MSc studies. His research interests lie in the field of computer vision, particularly in the domains of visual attention and saliency detection.

Yael Edan received her BSc in Computer Engineering and her MSc in Agricultural Engineering from Technion; and her PhD in Engineering from Purdue University. She is a Professor in the Department of Industrial Engineering and Management. Her research includes robotic and sensory performance analysis; systems engineering of robotic systems; robotic control of dynamic tasks; sensor fusion and selection algorithms; robotic gripper analysis and design; and multi-robot and human-robot control and collaboration models. In addition, she has made major contributions in the introduction and application of intelligent automation and robotic systems to the field of agriculture with several patents.

Ohad Ben-Shahar received his BSc and MSc from Technion and his PhD from Yale University, all in Computer Science. He is an Associate Professor in the Department of Computer Science at Ben-Gurion University and the Director of the Interdisciplinary Computational Vision Lab. His main area of research is in computational vision and image analysis, where he is focusing primarily on issues related to the differential geometrical foundations of perceptual organisation and early vision. His work is explicitly multidisciplinary and his computational research is often endowed by investigations into human perception, visual psychophysics, and computational neuroscience of biological vision.

1 Introduction

Advanced agricultural automation has increased productivity many folds by reducing manual labour and production costs, increasing yield and quality, and enabling better control over environmental implications. Despite these developments, numerous agricultural tasks are still being handled manually, challenging the consistently shrinking (and increasingly more expensive) agricultural labour force with physically hard, repetitive, and time-consuming operations that are still too sophisticated for present-day robotic systems.

A critical aspect in successful agricultural robots (henceforth, *agrobots*) is their ability to process sensory information, and in particular, their capacity to analyse and interpret visual input. Indeed, the coupling of visual data with proper machine vision algorithms may facilitate numerous operations and advance agricultural automation to new levels. However, the challenges associated with machine vision in severely unconstrained environments like those encountered in agricultural settings are countless: objects of various colours, shapes, sizes, textures, and reflectance properties; highly unstructured scenes with large degree of uncertainty; ever-changing illumination and shadow conditions; severe occlusions; and the sheer complexity of the typical unstructured agricultural scene, are only part of the problems that such a machine vision system must face. It is no surprise then that present-day success is still limited, leaving agriculture as an important frontier of applied computer vision.

While machine vision in agrobotic systems (henceforth, *agrovision*) is yet to reach its full potential, many applications have been developed for various tasks in the fields, orchards, and greenhouses. Among these are autonomous navigation and obstacles avoidance (Astrand and Baerveldt, 2005; Wei et al., 2005; Zhao and Jiang, 2011), precision and selective spraying (Berenstein et al., 2010; Tellaeche et al., 2008); weed detection (Slaughter et al., 2008), yield estimation (Chinchuluun and Lee, 2006; Qiao et al., 2005), seedling planting (Huang and Lee, 2010) and ripeness and quality evaluation (Yongjie et al., 2010). However, perhaps the most prevalent application has been fruit detection (Bulanon et al., 2002), where the goal is

1 to detect the presence of individual fruits

2 to discriminate them from the rest of the scene (leaves, branches, sky, etc.)

3 to localise them in space.

All these detected targets are then used in order to facilitate the interaction of the fruit with robotic manipulators and end effectors for further handling and physical processing, and in particular for harvesting operations. Interestingly, although much attention was given to these three visual processing problems during the last 30 years, no selective harvesting robot has ever reached commercial maturity. While this unfortunate outcome cannot be blamed solely on the failure of machine vision to handle the challenges posed by the unconstrained and unstructured agricultural environment, it is undeniable that successful machine vision is critical for achieving high detection rates of ripe fruit in real time, all of which are mandatory preconditions for a selective harvesting agrobot to be efficient and cost-effective.

With the identification of the above as a bottleneck task on which progress in agrobotical research and systems greatly depends, in this paper we present a comprehensive review of the state-of-the-art in fruit detection and localisation for harvesting robots. Inspired by an early survey by Jimenez et al. (2000b), here we systematically summarise the related literature from the last two decades, but in order to facilitate a comprehensive outlook, and to provide useful reference for researchers who seek to explore the relevant literature, the bulk of our review is organised around three main themes or criteria - the sensory configuration used by the various systems, the visual cues employed, and the class of machine vision algorithm put to work. With this division, researchers can also use this review as a reference guide for a particular type of solution they may have in mind for future systems. For each such approach we also discuss the limitations of present-day solutions and later on we expand on the challenges ahead more generally. We conclude with suggested directions for the research community to follow in order to meet the goal of practical (perhaps even commercial) selective fruit harvesting robots.

2 Imaging sensors

Although this paper focuses on machine vision, and therefore on sensory data acquired by cameras, it is important to acknowledge that even within this relatively narrow sensory mode, the variability of sensor configurations and types is still large, ranging from a single grey level camera in entry-level systems, to combinations of hyperspectral cameras with non-visual sensors in high end systems. To set the background properly, here we quickly review the main configurations that can be found in the literature.

2.1 Single camera

The sensors employed in most previous studies are standard BW or colour CCD cameras, typically positioned on the body of the robot or the main platform to provide a single view on the scene being analysed (Bulanon et al., 2001, 2002; Bulanon and Kataoka, 2010; Okamoto and Lee, 2010; Zhao et al., 2005). Less frequent are systems with a camera placed on the robot's end effector (Hayashi et al., 2002; Ling et al., 2004) while in still fewer cases the system employs both types of cameras to enjoy global view of the scene *and* a gripper-centred close-up view on specific targets (Edan et al., 2000; Feng et al., 2008; Van Henten et al., 2002; Yuan et al., 2010).

2.2 Calibrated stereo

Naturally, multiple cameras can provide greater information than single ones, and when they form a calibrated pair, their stereo configuration can be used to extract depth information on the imaged objects via triangulation and the analysis of the disparity between corresponding points (Hartley and Zisserman, 2000). Indeed, the use of calibrated stereo in harvesting agrobotics has greatly increased in recent years (Jiang et al., 2008a; Kitamura et al., 2008; Kondo et al., 2008; Kong et al., 2010; Takahashi et al., 2002; Yuan et al., 2010). To the best of our knowledge, multiview configurations involving *more than* two cameras (e.g., Kang et al., 2008) were not used in harvesting robots, although multiple stereo pairs have been employed, mainly to improve target visibility and to handle extreme illumination variations (Plebe and Grasso, 2001).

2.3 Vision and range sensors

Acknowledging that depth information that is obtained purely visually may suffer inaccuracies, some researchers have endowed visual input with range sensors in order to acquire depth more directly so that better fruit detection is obtained (Jimenez et al., 2000a; Monta and Namba, 2003). When reviewed in this paper, however, such configurations will be examined for their vision components only.

2.4 Spectral imaging

With the development of sensor and spectroscopy technology, spectral imaging has become a popular means for recognition of objects based on their different reflectance in selected wavelengths. Naturally, this approach may provide significant advantage when targets and non-targets (e.g., fruits and foliage) have the same apparent colours (Kane and Lee, 2006, 2007; Kondo et al., 1996; Van Henten et al., 2002; Yuan et al., 2010). Especially effective for discrimination purposes in this context are the 970 nm and 850 nm near infrared (NIR) wavelengths due to significant difference in water absorption (Kondo et al., 1996).

2.5 Hyperspectral imaging

Extending both standard colour (RGB) imaging and spectral imaging from other selected spectral bands, hyperspectral imaging is an emerging technology that (in the spectral resolution limit) provides the complete spectral signature for each pixel in the visual field of the camera. Clearly, with the overwhelming amount of additional information available, better decisions can be made regarding fruit detection and classification (Okamoto and Lee, 2009; Safren et al., 2007). This, however, comes at a costly price, both in acquisition time (at the order of minutes per image) and processing time. Hence, at the present state of the technology, it is typically used for offline processing and for preprocessing of data for the identification of selective spectral channels whose processing may provide useful decisions in real time.

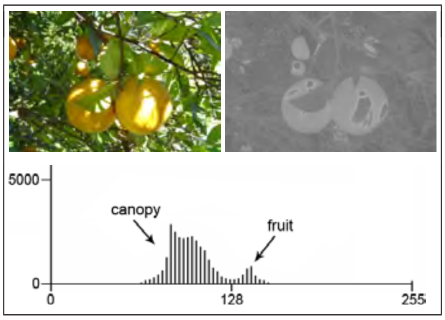
3 Visual cues

The type of image sensor aside, many machine vision algorithms are also characterised by the visual cue, or cues, on which they operate. In an attempt to exploit every possible type of information in the complex agricultural environment, computer vision algorithms in harvesting robots have tried to use a variety of visual cues and properties. Before turning to discuss algorithms in the next section, the present section quickly surveys the literature along this dimension (see also Figure 1).

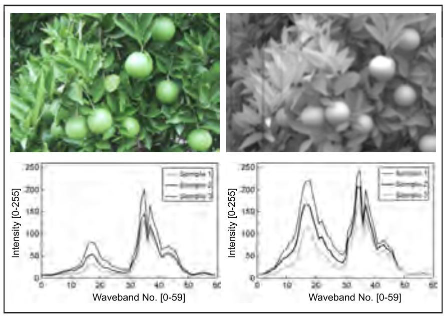
3.1 Colour

Since fruits, and ripe fruit in particular, tend to have different colours than the foliage and branches around them, *colour* becomes one of the most popular visual cues used in harvesting robots that employ machine vision. Typically, colour is used in the RGB representation (Arima et al., 2003; Bin et al., 2010; Jiang et al., 2008a; Okamoto and Lee, 2010), though other colour spaces , both standard (e.g., Annamalai et al., 2004; Jiang et al., 2008b; Kondo et al., 2008; Regunathan and Lee, 2005; Tarrio et al., 2006) and ad hoc (Feng et al., 2008), have been employed also. Naturally, colour may not provide a solution when the targets and their surrounding have similar look (Okamoto and Lee, 2009, 2010; Van Henten et al., 2002), and unless dealt with explicitly, can suffer from uncontrolled or changing illumination and shadows (Bulanon et al., 2002, 2009; Hannan and Burks, 2004). Some of these difficulties are presented in Figure 2.

Figure 1 Typical visual cues used for target detection in harvesting robots, (a) a typical citrus fruit can be separated from the background by colour only, and here this is done by calculating the proportion of the red channel in each pixel (reproduced from Hannan et al., 2007) (b) spectral reflectance, the reflectance in selected (top right) or all wavelengths (top left), may be used to distinguish fruit (in this case, green citrus) from background when they share the same observed colour, the graphs below, show the spectrum of old leaf vs. green fruit and indicate that specific wavebands can be used to distinguish between them (reproduced from Okamoto and Lee, 2009) (c) thermal response allows to highlight citrus fruit whose colour, potentially an indicative cue, is very sensitive to illumination conditions (reproduced from Bulanon et al., 2009) (d) the smooth skin of the fruit reflects a different visual texture than the background, and can be utilised to distinguish it from the foliage, this property is measured here via the density of measured edge elements in unit area (reproduced from Okamoto and Lee, 2010) (see online version for colours)

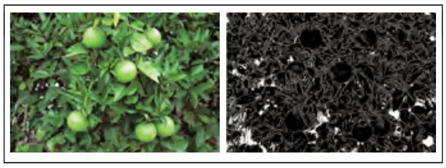


(a)



(b)

(c)

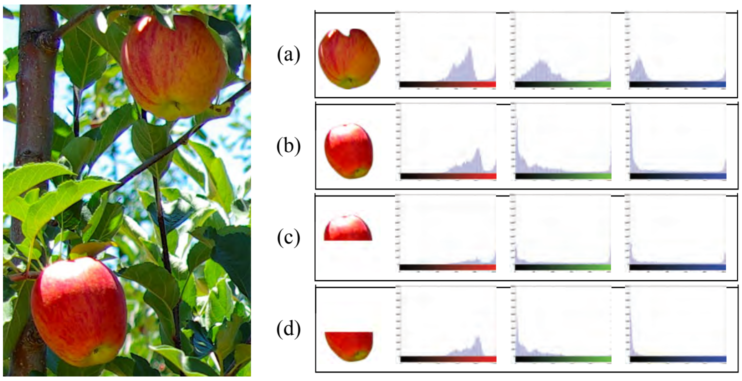


(d)

3.2 Spectral reflectance

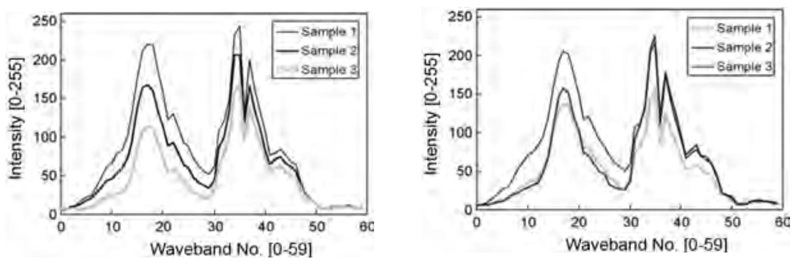
colour analysis can be applied very conveniently for non-green fruits. However, when the chromatic differences between the fruit and its background foliage are more subtle (say in green peppers, pears, or cucumbers), it becomes more difficult to use raw colour as a discriminating factor. Still, as is well known from metamerism (Wyszecki and Stiles, 2000), objects of similar colours do not necessarily have the same spectral signature, and in some cases they reflect quite differently in selected discrete channels, either inside or outside the visible light range. Hence, spectral reflectance can constitute an effective discriminatory factor and indeed it is being used increasingly more often in harvesting robots.

Figure 2 Difficulties involved in colour as a visual cue: although the two apples hold the same perceptual colour (see also the discussion about colour constancy in Section 5.1), illumination variations even in the same scene greatly affect their measured colour components histograms (here in RGB space), note the [(a) and (b)] differences between different objects in the same scene, which correspond to the two apples) and [(c) and (d)] differences within the same object also, which correspond to the two parts of the lower apple) (see online version for colours)



Ultimately, the entire spectral signature would be used for analysis after being measured by a hyperspectral imaging device (Okamoto and Lee, 2009; Safren et al., 2007). To maintain real time performance, however, spectral reflectance is often only obtained from *selected* spectral channels via standard cameras equipped with narrow-band filters (Van Henten et al., 2002; Yuan et al., 2010) or by laser modules operating in specific wavebands (Tanigaki et al., 2008). For example, Figure 1(b) (lower right graph) presents the spectral reflectance of a green fruit and shows how it reflects predominantly in certain wavelength channels.

Figure 3 Advantages and limitations of spectral response as a discriminatory cue (reproduced from Okamoto and Lee, 2009), (a) spectral response of green citrus shows how it reflects predominantly in certain wavelength channels, a fact that can be used for discrimination (b) despite its strength, spectral signature cannot always resolve targets from background



(a) (b)

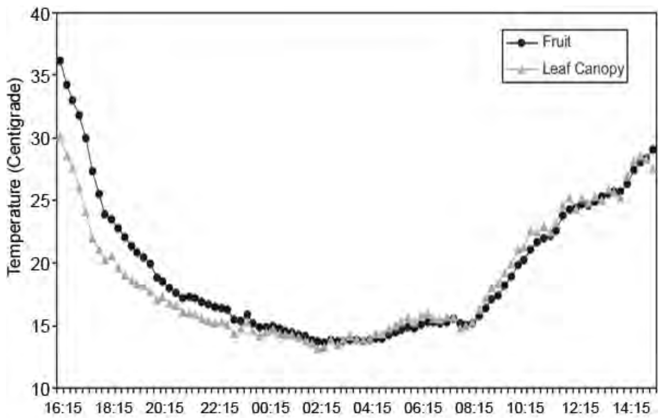
Note: As can be appreciated in comparison to Figure 3(a), young leaves often reflect similarly to green fruit, making it more difficult to use the spectral response as a discriminatory cue.

Whether one uses the entire spectrum of selected spectral channels, the significant potential in exploiting this visual cue is not without limits. First, spectral response is still sensitive to illumination conditions (Van Henten et al., 2002; Yuan et al., 2010). Furthermore, it is not impossible that both the fruit and the background foliage will have the same spectral signature under the operating illumination, as shown in Figure 3 for the green fruit and the young leaves. Hence, spectral response alone cannot always provide discrimination. Moreover, despite its strengths, spectral reflectance cannot resolve issues that relate to the *spatial* organisation of the visual data, such as difficulties in detection due to occlusions.

3.3 Thermal response

Highly related to spectral reflectance is the thermal response of objects, i.e., their emitted radiation in the infrared range where it is strongly affected by both the temperature and the emissivity of materials, typically in the spectral range of 9–14 μm (Bhanu and Pavlidis, 2004). Implied by the black body radiation law (Kittel and Kroemer, 1980), such infrared radiation is emitted by all objects above absolute zero in a manner increasing with temperature. Hence, thermography makes it possible to 'see' the environment with or without active (visible or non-visible) illumination and to sense variations in temperature in and between objects. Figure 1(c) presents an example of citrus fruit and a corresponding thermal image.

Figure 4 Thermal profile of the canopy and the fruit during a 24-hour period (reproduced from Bulanon et al., 2009)



Note: The graph shows both the variability of the thermal response and how foreground and background objects sometimes become indistinguishable by this cue during the day.

Since leaves, unlike fruits, accumulate significantly less heat and emit it for a shorter time, thermography provides an excellent approach for target detection in harvesting machines (Bulanon et al., 2009). However, since the thermal response is sensitive to the illumination (exposure to sunlight) and heat accumulation, fruit on different parts of the tree might respond differently and complicate the exploitation of this vital source of information (Bulanon et al., 2009; Stajnko et al., 2004). Furthermore, detection problems may arise during significant parts of the day when fruit and leaf canopy have the same response (Figure 4). Naturally, issues related to the spatial organisation of the visual data and the effects of occlusions are not addressed by this visual cue either.

3.4 Texture

While colour and spectral reflectance are purely local (point-wise) properties, virtually all objects in the physical world occupy extended spatial area, and hence can be represented with non-local descriptions. Texture, the repeated visual pattern that covers surfaces and regions (either regularly or stochastically), is perhaps the first visual cue that goes beyond purely local cues to describe the appearance of small image patches (Nalwa, 1993; Forsyth and Ponce, 2002). In agricultural settings it can be used as a discriminatory factor between different types of objects, and in particular, between fruits and their surroundings (Okamoto and Lee, 2010; Rakun et al., 2011; Zhao et al., 2005). An example for texture cue can be seen in Figure 1(d). The smooth skin of the fruit (a texture property) is utilised to distinguish fruit from foliage by using edge detection: While few edges are detected on the fruit's skin due to its smoothness, many edges are detected on the leaf area. Hence, when the skin of the fruit is indeed smooth, patches of significant size and low edge density could represent the target quite robustly (Okamoto and Lee, 2010).

Clearly, texture could serve as a particularly effective cue when colour is not discriminatory enough, and it is usually more stable than reflective properties under illumination variations. Still, being confined spatially, texture must be combined with additional visual cues to handle global spatial confounds like occlusion (Rakun et al., 2011). Naturally, the descriptors used must also be powerful enough in order to handle cases where the differences in texture appearance between the targets and the background are finer than, for example, the first order distribution of edges.

3.5 Shape

Great difficulty in detecting fruit in unstructured environments arises from the extreme variations in illumination conditions (Harrell et al., 1989) which cause extreme appearance variations. One global visual cue that could be less susceptible to illumination is the shape of the target, and although it is more computationally demanding to extract and analyse, shape is becoming increasingly more popular in harvesting robots (Chi and Ling, 2004; Edan et al., 2000; Hannan and Burks, 2004; Hayashi et al., 2002; Jimenez et al., 2000b; Kong et al., 2010; Ling et al., 2004; Liu et al., 2011; Okamoto and Lee, 2010; Rakun et al., 2011; Yuan et al., 2010; Zhang and Zhang, 2008). At its essence, shape implies a particular spatial relationship between the geometrical atoms (points, occluding contours, surfaces) that make up a coherent physical object. Since fruits are practically rigid objects, their shape relationship remains invariant in the three dimensional world. Although few invariances persist under perspective projection (Nalwa, 1993; Forsyth and Ponce, 2002), shape remains a strong visual cue in the 2D image plane as well as under illumination variations (see Figure 5).

Figure 5 Shape as a visual cue, (a) the shape of a fruit (in this case, peach) tends to persist even under extreme variation in appearance due to illumination, compare top and bottom images (reproduced from Liu et al., 2011) (b) despite robustness to illumination variations, shape is extremely sensitive to variations due to occlusions, note how the circular projection of the apples changes drastically when occluded (c) shape exhibits large variability *within* each class, the contours on the right show different shapes that the projected image of sweet peppers might take (see online version for colours)

(a) (b) (c)

Despite its strengths, shape can be difficult to extract and analyse, a time consuming task which was impossible to achieve in standard computational hardware until recent times. Shape is also extremely sensitive to occlusions [see Figure 5(b)] and in most cases, it exhibits large variability even *within* classes of targets [see Figure 5(c)]. Hence, in most previous work where shape is used explicitly, analysis employs very simple shape models and is used with fruits whose image projection can be reasonably modelled as spheres (Chi and Ling, 2004; Edan et al., 2000; Hannan and Burks, 2004; Jimenez et al., 2000b; Kong et al., 2010; Ling et al., 2004; Liu et al., 2011; Okamoto and Lee, 2010; Rakun et al., 2011; Zhang and Zhang, 2008). Notable exceptions in this context are studies on eggplants (Hayashi et al., 2002) and cucumbers (Yuan et al., 2010).

3.6 Fusion of visual cues

As could be expected, a single visual cue rarely represents the target object in a satisfactory manner except in extreme cases. In harvesting robots, attempts to detect fruits using a single visual cue typically encounter problems due to illumination variations, spatial occlusions, and appearance variations. However, since each visual cue represents different aspects of the target, it is reasonable to hope that one visual cue could compensate for the representational limitations or flaws of the others. Hence, fusing several cues together may provide increased performance altogether, an approach that has been employed in computer vision in general (Zheng and Xue, 2009) and in agricultural settings in particular (Patel et al., 2011). While most studies in this review indeed employ such an approach (see Section 4), here we only note that fusion of visual cues can be done at different levels of representation, and in particular, at the image level (Bulanon et al., 2009) or at the algorithm level (Wachs et al., 2010).

4 Image analysis algorithms

With imaging sensors and visual cues relevant for target detection in harvesting robots briefly reviewed, in the bulk of this paper we discuss the algorithms, methods and computational approaches used in this unique domain of application. To facilitate more constructive presentation of the technical content, this review is organised by the type of algorithms employed, rather than chronologically or by agricultural criteria (e.g., fruit kind). We believe that this organisation better serves to identify patterns in the literature and provides useful reference for future research in the field.

4.1 Elementary methods for image segmentation

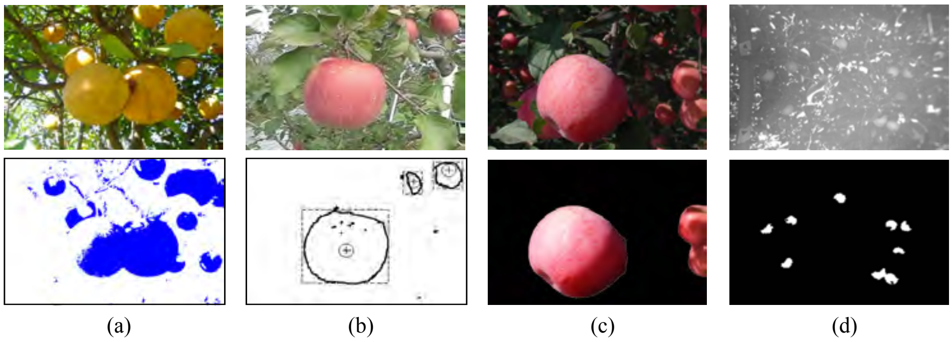
Detection of suitable targets for harvesting operations can be modelled as segmentation of the image to fruits and background. Connected components in the segmented image could then be considered as harvesting targets (assuming no occlusions). Employing the most elementary methods for segmentation, a vast number of studies have approached this problem via thresholding the visual cue (typically, colour) map at proper values (Annamalai et al., 2004; Arima et al., 2003; Bin et al., 2010; Bulanon et al., 2001, 2009; Edan et al., 2000; Feng et al., 2008; Hannan and Burks, 2004; Hayashi et al., 2002; Jiang et al., 2008a, 2008b; Liu et al., 2011; Okamoto and Lee, 2009, 2010; Rakun et al., 2011; Stajnko et al., 2004; Tanigaki et al., 2008). Similar approach has been used on other visual cues as well, such as spectral reflectance (Okamoto and Lee, 2009) and texture (Zhao et al., 2005).

As is well known from the general computer vision literature, predefined global thresholding is prone to fail in most scenarios (Haralick and Shapiro, 1985; Nalwa, 1993), which promoted several researchers to propose adaptive thresholding methods also (Bulanon et al., 2002; Chi and Ling, 2004; Kane and Lee, 2007; Ling et al., 2004; Yuan et al., 2010) where the threshold is automatically adjusted to the ambient illumination conditions. Interestingly, to the best of our knowledge, no attempt was made to use spatially-varying local thresholding in the context of agrovision despite its decades-long presence in the general computer vision literature (e.g., Weszka, 1978).

While adaptive thresholds may provide improved results within the domain of global thresholding (see Figure 6), the high variance which is typical of the unstructured agricultural scene implies that one can expect such algorithms little more than coarse and inaccurate segmentation, an observation that encouraged other researchers to employ other elementary segmentation methods as well, such as edge detection (Zhao et al., 2005), region merging (Safren et al., 2007), and region growing (Kong et al., 2010). In these cases too, one visual cue (e.g., texture or spectral reflectance) was used to define homogeneity of regions. Selected results are illustrated in Figure 6.

One additional problem of elementary methods for segmentation is their ignorance of all shape information or expectations. Indeed, unless dealt with explicitly, these methods are likely to provide segments which include clusters of fruits, rather than individual fruits [Figures 6(a), 6(c) and 6(d)]. For some applications, like spraying, this may not pose a major problem. For harvesting robots, however, this missing component is critical. Notable attempts to handle this problem without shape information have used watershed segmentation (Regunathan and Lee, 2005; Safren et al., 2007), where the grey scale image is considered a topographical surface and boundaries are defined as the curves that separate basins of flooding [Figure 6(d)].

Figure 6 Selected results using elementary methods for image segmentation, (a) predefined global thresholding (reproduced from Hannan et al., 2007) (b) adaptive global thresholding (reproduced from Bulanon et al., 2002) (c) region growing (reproduced from Kong et al., 2010) (d) watershed segmentation (reproduced from Safren et al., 2008) (see online version for colours)



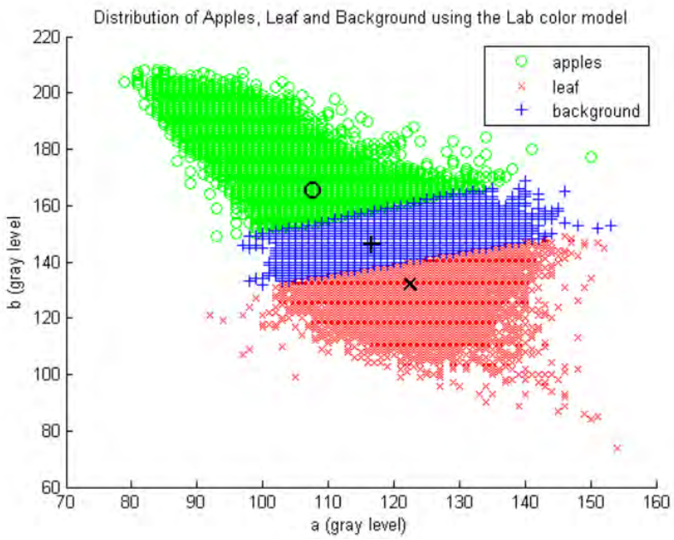
4.2 Clustering

Clustering, as a form of unsupervised learning approach (Marsland, 2009), has also been employed in machine vision algorithms for harvesting robots to partition the image into targets (fruits) and background, especially when multiple visual cues are fused together. Although more advanced clustering methods have been developed, most clustering approaches in fruit detection algorithms have employed the classical k-means algorithm (Marsland, 2009) in various colour spaces, exploiting the fact that the number of clusters in this application is known in advance (Bulanon et al., 2004b; Bulanon and Kataoka, 2010; Chinchuluun and Lee, 2006; Wachs et al., 2010).

One of the major aspects of clustering techniques is the choice of a distance measure between data points in a feature space, while one typically defines it based on an L_p norm (where p is usually selected as 1, 2, or ∞ for the Manhattan, Euclidean, or the maximum norms, respectively). Euclidean distance is sensitive to the scaling of the variables involved and has no means of taking into account correlated variables. Some studies in the harvesting robots literature have employed a clustering method based on Mahalanobis distance (Zhen et al., 2007). In settings involving non-spherically symmetric distributions of data points, as normally obtained when mapping images of natural environments into a certain feature space, the Mahalanobis distance is better adapted than the Euclidean distance, since it takes into account the covariance among the variables in calculating distances. With this measure, the problems of scale and correlation inherent in the Euclidean distance are no longer an issue.

Since basic clustering only replaces more elementary methods for image segmentation, it exhibits several similar limitations and drawbacks. In particular, it is sensitive to illumination conditions and it often needs to cope with feature points that do not cluster or separate well into clusters (Figure 7). Furthermore, clustering of local feature points is intrinsically agnostic to shape aspects and hence, unless combined with other methods with shape-explicit considerations, it is unlikely to provide a practical solution to selective harvesting robots.

Figure 7 Feature point distribution of leaves, apples, and background points in a typical agricultural scene (reproduced from Wachs et al., 2010) (see online version for colours)



Note: While the colours represent clustering results, the lack of clear division into clusters is evident.

4.3 Template matching

Template matching is a technique for recognising portions of a given image that match (typically in terms of appearance) with a specific template pattern. The matching process is usually done by moving the template across the image, while performing a computation that aims to determine how well the template matches the image in each position. Typically, this computation is based on similarity measures such as cross-correlation and sum of squared differences (SSD). Additional similarity measures are based on specific features of the image such as edges and corners. Template matching is useful in contexts where the diversity of the target object is small enough. Thus, considering the high variability characterising natural environments, it is not surprising that this technique has been used only rarely in the context of harvesting robots. One notable example is due to Zhang and Zhang (2008) who applied template matching to a binary image of blobs representing target objects. This reduced the variability of the target objects when they were compared with a perfect circle structure as a template. The matching was done using a joint transform correlation, where the template and binary image were placed side by side forming a joint image as input to the cross correlation process. The peaks in the output cross correlation spectrum corresponded to the cross correlation of the joint image with itself and the cross correlation of the template with a target object. A blob was declared a fruit according to the ratio between the two peaks.

4.4 Shape inference

As mentioned in Section 3, shape is a major visual cue that harvesting robots should exploit, especially when fruit handling must be selective and individual. As a fundamental problem in computer vision, shape inference is concerned with finding a shape that best fits the geometric evidence measured from the image (Forsyth and Ponce, 2002). Choices of such evidence include geometric atoms like points, lines, hyperplanes etc. and the inference process can involve a variety of mechanisms such as voting (see next subsection), statistical inference, or optimisation (deterministic or stochastic).

Within this realm, shape can be inferred either globally or locally. Global shape inference requires a definition of a shape model for the anticipated targets, and the use of fitting techniques to locate instances of the model under expected observable transformations. Local shape inference usually involves the inference of local shape descriptors and the accumulation of enough compatible evidence to indicate the presence of a particular shape instance. The difficulties in both cases lie primarily in the construction of good models for the fruits at hand, and global shape inference also suffers from the computational cost involved in detecting instances of the model in the image. Hence, in practice, the majority of shape inference methods have confined their domain of application to spherical fruits, where both the model and its detection are perhaps the simplest possible. Furthermore, most related papers have focused on *local* inference techniques whose realisation can be done with simpler computational and algorithmic means. A notable exception is the work by Jimenez et al. (2000a), where in addition to local inference processes, a least square sphere fitting procedure was applied to range data. When applying local shape inference, various techniques have been employed. Plebe and Grasso (2001), for example, have devised an adaptive edge fitting process to detect spherical shapes by estimating the centre point and radius of each portion of an edge. In their work, edges were grouped into closed curves, and each closed curve was labelled as a separate object. For each portion of an edge, the curvature and the radius of curvature were calculated according to the length of the arc of edge and the angle corresponding to the arc. These data were then used to estimate the position of the hypothetical circle that goes through the edge. When a sharp variation in the position of the estimated centre is detected, the mean value of the positions of the centres and of the radii of all evidence thus far is calculated and declared as a fruit centre and radius, after which the evidence for a new fruit is beginning to accumulate.

Somewhat differently, Jimenez et al. (2000a) proposed an approach based on the generation of a set of primitives that are characteristic to spherical objects: contour and crown pixels, convex regions and reflectivity-based regions. Due to their ability to cover the whole surface with little overlapping, these primitives are considered useful in accumulating evidence in recognising spherical fruits. In addition, since they detect different areas of the sphere, they are found appropriate in dealing with occlusions. The extracted primitives are used to estimate sphere parameters and a degree of confidence over that estimation. Contour and crown primitives are fitted through circular Hough transform (cf., Section 4.5) while convex and reflectance primitives are used to initialise a least square sphere-fitting process. Each one of the extracted primitives is used to generate a partial hypothesis about the existence of a sphere. At the last step, the four partial hypotheses are integrated to generate a final hypothesis and the hypotheses not having a sufficient evidence value are rejected.

Finally, it is worth mentioning the shape-related approach employed by Okamoto and Lee (2010) where the edge map of the input image was analysed through a geometrical template consisting of a circular region and an outer ring both having predefined dimensions. The ratio between the number of edge points included in these two components was then used to discriminate and locate fruits from the background, though clearly this method is limited to a single circular instance.

4.5 Voting

Voting is a computational technique in which each local visual evidence in the image votes for all possible global interpretations it could arise from. Casting and accumulating all these votes (from all evidences) in an 'interpretation space' results in a distribution of votes that could indicate which interpretation is consistent with most evidence and hence should be selected. In vision, this technique has been used almost invariably for the detection of shapes and patterns.

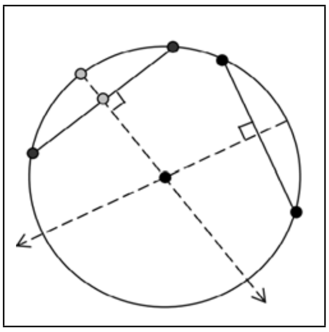
One popular voting technique formulated in the early days of computer vision is the Hough transform (Ballard, 1981; Illingworth and Kittler, 1988). Although formulated for lines, a proper variation of the Hough transform is the circular Hough transform (CHT) whose goal is to find circular patterns. The CHT is used to transform a set of points in an image space into a set of accumulated votes in a parameter space, where each point represents one possible instance of a circle. Votes are accumulated in an accumulator array for all parameter combinations and the array elements that contain the highest number of local maxima of votes indicates the presence of the shape. In agrovision, CHT was used to detect spherical fruits as oranges (Jimenez et al., 2000a), apples (Wachs et al., 2010) and coconuts (Rizon et al., 2005).

The Hough transform and the CHT are relatively expensive in computation, and hence several techniques were used to improve their complexity. For example, expectations about the observed radius of fruit were used to limit the parameter space, either globally (Jimenez et al., 2000a) or dynamically for each candidate (Wachs et al., 2010). In other cases, the number of votes per image point was reduced using edge direction information (Jimenez et al., 2000a; Rizon et al., 2005), and detection of peaks in the parameter space was enhanced using the back-transformation method (Jimenez et al., 2000a), which is suited for complex images where the boundaries are partially missing, obscured, or distorted. The back-transform strategy is used for interpreting the resulting accumulator array in the parameter space in a simplified way which increases the detection accuracy (Gerig, 1987).

Since the CHT remains computationally expensive despite such improvements, faster alternatives have been proposed and applied. Attempting to detect tomatoes, Ling et al. (2004) and Chi and Ling (2004) observed that the interception of perpendicular bisectors of any two chords of a circle is the centre of the circle (Figure 8). Hence, in these studies, every intersection point between two perpendicular bisectors was declared as potential centre point and was voted in an accumulator array. The peaks in the array are chosen to be the centre points of the tomatoes.

Interestingly, although extensions of the Hough transform have been proposed for more general shapes (Ballard, 1981), to our best knowledge these voting techniques was never used for non-spherical fruits.

Figure 8 Voting one shape via shape properties (reproduced from Ling et al., 2004)



Note: The centre of a circle could be determined from an interception of the perpendicular bisectors of any two different chords.

4.6 Machine learning

Machine learning, a branch of computer science which deals with the design and analysis of algorithms that improve their performance based on observable data, is often used for problems whose definitions or solutions lack well defined formalism. Ever since its inception in the early days of computer science, machine learning has evolved into a distinguished set of approaches such as decision trees, evolutionary computation, and Bayesian networks. However, only three main approaches have penetrated the community of target detection for harvesting robots: clustering (cf., Section 4.2), artificial neural networks (ANN), and support vector machines (SVM).

An ANN is a formal computational model that is used frequently in machine learning and is inspired by the rough structure of neural circuitry in biological nervous systems (Gurney, 1997). Used in conjunctions with all learning paradigms (e.g., unsupervised, supervised, reinforcement), ANN have been quite popular in machine vision for harvesting robots as well. In particular, ANN has been proposed for classification between different elements in the scene such as fruit vs. background (Bulanon et al., 2004a; Regunathan and Lee, 2005). In such cases the input values (usually local colour features) are fed to the input layer of the network while the output layer provides a binary categorisation. Training such networks is usually done in a supervised fashion by providing fruit and background examples (Plebe and Grasso, 2001) to build a look-up table that maps each point in the selected colour space to its corresponding class (fruit, background). More elaborate use of ANN was proposed by Wachs et al. (2010) who fused three ANN classifiers, one for each of the three colour spaces L * a * b, HSV, and RGB, in order to leverage the advantages that each one may provide. First, each of the three ANNs classifies the pixels of each image patch (subwindow) into several predefined classes (apple, leaves, branches, ground and sky). Then, a majority vote among the networks determines the final class that is associated with the patch.

As mentioned above, in addition to ANN, another popular learning-based approach used in target detection in agrobotics is SVM – a supervised learning method for classification and pattern recognition. SVM analysis performs classification by constructing an N-dimensional hyperplane that optimally separates classes of data into two categories. In the case of SVM, optimality is defined by maximising the distance between the separating hyperplane to the nearest neighbour in each of the separated classes, a notion also known as the margin (Cristianini and Shawe-Taylor, 2000).

In target detection for agrobotics, SVM has been used to classify objects with prototype texture against other possible patterns. This method has been applied, for example, for the detection of apples (Rakun et al., 2011). Apple detection was also done via LS-SVM, a version of SVM which solves a set of linear equations (instead of optimising a quadratic function of several variables in the classical SVM) based on several colour and shape feature values (Kong et al., 2010).

5 Challenges ahead and possible directions

The review and literature summary presented above suggest several trends that characterise the work on computer vision for fruit harvesting robots. In general, different studies in this area appear rather separated, little reused or developed progressively, often ad hoc, and almost always quite disconnected from progress offered in the general computer vision literature. This, however, should not come as much surprise. Unfortunately, agricultural robotic systems are complicated integrated systems in which particular modules are often required to address very specific and applicative needs in real time. As a result, much of the solutions offered employ only simple (even simplistic) computer vision tools where the emphasis goes for more informative sensory methods rather than for more sophisticated analysis techniques. While the former should not be neglected, of course, we believe that the age of super fast computers offers new opportunities in terms of the latter, with which the main challenges of precise, autonomous, and selective agrobotics may be addressed more successfully. In what follows we discuss several of these challenges more closely.

5.1 Reconsidering classical cues

Undoubtedly, the use of colour as a visual cue masters the literature on fruit detection for harvesting robots. In most cases, there is a good reason for this. A ripe tomato clearly stands out from the green foliage by its colour, as are yellow sweet peppers or purple grapes. But clearly, selective agrobotics may need to deal with green fruit also (cucumbers, green apples, pears, lime, white grapes, green peppers, etc.), and may also need to handle unripe or immature fruit (e.g., for diseases detection or for spraying *during* the growing season). Furthermore, since reflected light depends critically on the incident light, illumination conditions affect measured colour in a significant way.

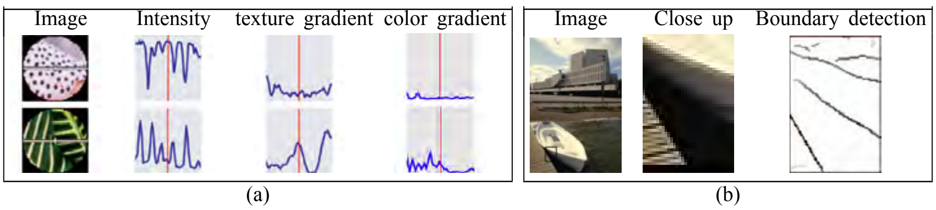
Following these observations, it is clear that using colour alone in a straight forward manner is likely to provide poor results in many cases. The challenge is therefore twofold. First, one may attempt to find visual signatures that are very robust to (and ultimately, independent of) illumination conditions, as is the case with human perception (see Figure 9 or the two apples in Figure 2). These capacities are well known in the human visual system as *colour constancy* (Palmer, 1999) and have recently become a focus of interest in the computer vision community as well (Barnard et al., 2002; Gehler et al., 2008; Gijsenij and Gevers, 2007; van de Weijer et al., 2007).

Figure 9 Colour constancy in human vision (see online version for colours)

Notes: Although acquired in drastically different illuminations, the extreme differences in the pepper appearance are largely (if imperfectly) compensated by human colour constancy mechanisms. The fruit remains ‘red’, the peduncle ‘green’, and the background ‘white’. A more realistic example is illustrated in Figure 2.

Second, although some research has already acknowledged this, much more effort must be placed on fusing colour with other traditional visual cues, and in particular, with texture. Again, this last visual cue is central in the general computer vision literature, and as is illustrated in Figure 10, it has taken a significant role in general segmentation algorithms (e.g., Jain and Farrokhnia (1991); Malik et al. (2001); Shotton et al. (2009); Yang et al. (2008)). It is likely that making the proper use of this cue in a fashion tailored to images from the agricultural domain may improve target detection performance significantly.

Figure 10 Boundaries detection by fusing brightness, colour and texture cues (reproduced from Martin et al., 2004), (a) texture gradient measures can provide useful information about perceptual texture boundaries without responding within coherent texture regions (b) combined with other visual cues, boundary detection can become robust in natural settings, and may improve performance on agricultural images also (see online version for colours)



5.2 Practical use of hyperspectral data

When traditional visual cues fail, machine vision for fruit harvesting robotics resort increasingly more often to spectral response of the target and background. This approach of enriching the sensory data acquired from the scene provides many advantages, but also presents several challenges, some of which are technical.

For example, hyperspectral imaging necessarily provides a wealth of information. Regardless of how much of this data is needed, the mere volume of the data poses a challenge in storage and processing (for example, a three megapixel colour image, which would normally consumes 9 MB in raw format and 1.3 MB in typical JPEG, might turn to 90 MB if acquired hyperspectrally in 10 nm spectral channels, and much larger if non-visible light is considered also). This challenge increases further with the emergence of real time hyperspectral cameras Cao et al. (2011) and the need to process these images as they stream directly from the camera. What might become necessary then are fast methods for dimension reduction often used in computer vision to handle large descriptors and representations (Papageorgiou et al., 1998; Wang and Paliwal, 2003; Yu and Yang, 2001).

Since currently *hyper*spectral imaging and processing cannot be done in real time, when spectral response is still desired, designers of agrovision systems tend to resort to *multi*spectral imaging from carefully selected channels which were found to provide good discrimination in offline tests. These selections are often ignorant of possible variations in illumination or assume controlled illumination altogether. Hence, better preprocessing techniques are needed to analyse hyperspectral input under variable illumination to yield the optimal multispectral filtering that provides robust detection rates under variable illumination. This type of problems may well be informed by work in the remote sensing community, where hyperspectral anomaly detection is a central problem (Manolakis et al., 2009).

5.3 The use of shape

As suggested in Section 4.4, although shape information has been leveraged in machine vision studies for harvesting robots, most proposed solutions have been characterised by local inference procedures while focusing on spherical shapes only (whose image projections are circles). The challenges ahead, however, call for significantly stronger approaches, perhaps involving new tools altogether.

Consider the typical image from a red pepper plantation/greenhouse shown in Figure 11(a). Clearly, given the significant differences in colour between the foliage and the fruits, the latter may be detected in a straight forward fashion, perhaps even robustly to illumination variations. Hence, the conclusion ‘there are sweet pepper fruits in front of the robot’ can indeed be made with most existing fruit detection algorithms.

Figure 11 (a) Red pepper under typical occlusion (b) shape fragments (c) entire shape (see online version for colours)

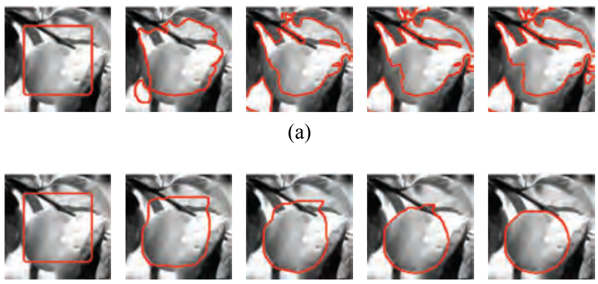
(a) (b) (c)

But clearly, for a harvesting robot, the type of conclusion as above is at best partial, or more reasonably phrased-uninformative (since moving along the aisle, the robot is likely to encounter fruits in almost every frame). A harvesting robot requires eventual physical access and manipulation of *whole* fruits and therefore needs answers to questions such as ‘*how many* fruits do I see in front of me’, ‘*where* are the fruits in the image’, and ‘*what is the pose* of each fruit I see’. In other words, even if one assumes perfect appearance-based detection using visual cues from the image, in a typical agricultural scenario one will obtain only *fragments* of the targets due to occlusion from the foliage and/or other fruits [Figure 11(b)]. In most selective agrobotic applications, the system must bridge the gap between these fragments to whole fruits [Figure 11(c)], which in our view is the main challenge that our research community faces today. Currently an unsolved problem, the solution will most likely involve aspects of *shape*.

When it comes to using shape for grouping visual fragments into wholes, several directions are possible. Most naturally, one may attempt shape *fitting*, as indeed was used few times in the past (see Section 4.4). However, these past proposals were relatively rare and devoted solely to circular shapes (in the image plane). The challenge to more general shapes involves both efficient shape modelling, and shape matching, neither of which is a trivial task.

While some attempts to consider non-spherical shapes in the agrobotics community have been made (Morimoto et al., 2000; Paulus and Schrevens, 1999), one may get important insights on new directions to address this problem from research in the general computer vision literature, and in particular, from research on face modelling (e.g., Rein-Lien and Jain, 2001; Tao et al., 2008) or even whole body modelling (e.g., Allen et al., 2003; Dekker et al., 1999). Perhaps more important are methods developed in the general computer vision literature for using shape *priors* in segmentation of objects under occlusions and/or geometric deformations. For example, while the use of active contours has been very popular in object detection and segmentation (Caselles et al., 1995; Chan and Vese, 2001; Cohen, 1991; Cohen and Kimmel, 1999; Kass et al., 1988; Kimmel and Bruckstein, 2003; Osher et al., 2003; Xu and Prince, 1998), recent work has integrated it with prior knowledge on the expected shape, yielding segmentation of objects under occlusion (Chen and Radke, 2009; Fang and Chan, 2007; Foulonneau et al., 2009; Gastaud et al., 2004), as demonstrated in Figure 12.

Figure 12 Segmentation with active contours and shape priors, implemented via one of the classical active contour schemes Chan and Vese (2001) under the level set framework, here we demonstrate that (a) segmentation without shape prior is prone to fail (b) endowing it with a shape prior component (in this case, a simple circular shape) may be able to cope not only with the shading variations but also with the fragmentation and occlusion (see online version for colours)



(a)

(b)

While these methods have yet to exhibit real-time performance, combining them with progress in GPU implementations of active contour models (e.g., Rumpf and Strzodka, 2001) could allow agrovision systems to enjoy both worlds.

5.4 Human vision inspired approaches

The use of shape in agrovision need not be limited to global shape models since local models can provide significant constraints towards global shape inference also. Interestingly, human vision provides ample evidence in this direction, both in terms of fragment grouping and recognition of occluded objects, and in terms of shape completion.

Consider the visual stimulus in Figure 13(a). Most observers would testify that under the occluders lie rectangles rather than more bow ties, as might be implied by the context. Here the human visual system applies basic Gestalt principles of good continuation (Wertheimer, 1955) to complete the missing parts of the occluding contours in a very particular way, a type of operation that has been modelled computationally in various ways (Ben-Yosef and Ben-Shahar, 2010a,b; Horn, 1983; Kimia et al., 2003; Mumford, 1994; Ullman, 1976) which could serve agrobotics in a meaningful and practical manner [Figure 13(b)].

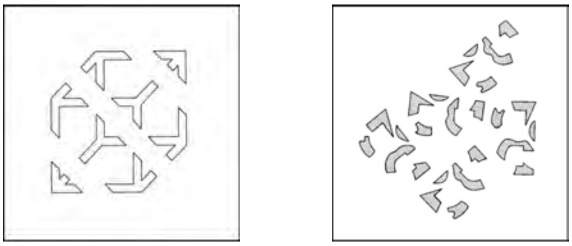
Figure 13 Boundary completion in human and machine, (a) despite the visual context, human observers tend to see rectangles behind the occluders, rather than additional bow ties (b) using completion principles inspired by human vision, here a computational algorithm (Ben-Yosef and Ben-Shahar, 2010a, 2010b) completes an occluded segment in the tomato using information about the occlusion points only (see online version for colours)

(a) (b)

Consider now the classical visual stimuli in Figure 14, which constitutes collections of fragments not much different in nature from Figure 11(b). It might be possible (though still unlikely) that a human observer could identify instances of objects that these fragments are part of should it be known a priori which objects he or she ought to look for. What is clear is that without such knowledge, perception of familiar objects is virtually impossible in this image.

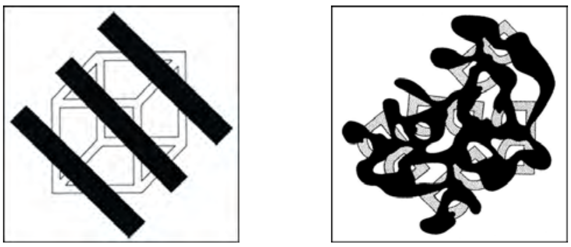
However, consider now in Figure 15 the same set of fragments when the occluder *is* present in the scene. Somehow, our visual system is now able to group the fragments correctly to yield a perception of the global objects. While the processes that bridge this gap are not fully understood yet, gaining insights from visual completion and perceptual organisation in the human visual system may provide new opportunities for progress on this exciting and difficult problem in agrovision.

Figure 14 Shape completion and recognition from visual fragments, (a) can you recognise the shape that these fragments come from? (reproduce from Kanizsa, 1979)
(b) these fragments belong to a collection of very familiar objects. Without the occluder, perception of these objects is nearly impossible (reproduce from Bregman, 1981)



(a) (b)

Figure 15 Shape completion and recognition from visual fragments, repeated from Figure 14 but this time *with* the occluders, perception of the objects whose fragments are observed is easy and immediate, (a) reproduced from Kanizsa (1979) and
(b) Bregman (1981)

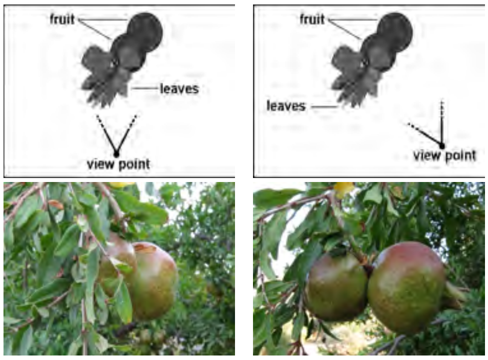


(a) (b)

5.5 Active vision

The issue of occlusions, and the perception of complete objects from partial visual information, can be handled not only passively but in active fashion also. When a human observer (say a worker in the greenhouse) encounters a scene like in Figure 16(a), his first instinct would be to shift his head (and therefore, eyes) in order to obtain a different view of the scene in which new parts of the objects are revealed (perhaps on the expense of parts already observed in the past). This *active vision* approach could well be exploited computationally for the benefit of agrovisual systems, and algorithms should be explored to answer questions such as 'where is my next optimal viewpoint given the information I have so far' or even for more basic tasks such as segmentation. To do so effectively, one must incorporate planning, as well as inference of three dimensional properties of the scene (Blake, 1992).

Figure 16 Active vision provides more information than static vision can, (a) the initial view shows a fruit occluded by leaves and occluding another fruit (b) moving the camera to a carefully selected viewpoint completely reveals both fruits (see online version for colours)



(a) (b)

5.6 Going beyond 2D

As can be appreciated from the literature, the state-of-the-art in agrovision research has focused almost exclusively on 2D image analysis (Chi and Ling, 2004; Edan et al., 2000; Hannan and Burks, 2004; Jimenez et al., 2000a; Kong et al., 2010; Ling et al., 2004; Liu et al., 2011; Okamoto and Lee, 2010; Zhang and Zhang, 2008). Conspired to constrain research in this way were the sheer complexity of the problems due to the unconstrained environment and the limited computational power that could be leveraged with available hardware. However, with autonomous selective agriculture at the front line of agrobotics, the new applications that emerge clearly require a significant leap in the type of machine vision employed, and in particular, one of the main challenges to come is 3D image understanding.

Consider again a harvesting robot operating in an apple orchard [cf., Figure 17(a)]. At the end of the processing loop awaits a harvesting manipulator, which must grasp the physical fruit in a very specific configuration before the correct harvesting sequence is initiated. Hence, optimal performance *demands* the estimation of the 3D pose of the fruit prior to grasping, including the recovery of the peduncle and perhaps the estimation of 3D shape of nearby rigid obstacles (like branches). Given the type of visual inputs involved [cf., Figure 17(a)], the challenge appears nearly impossible with existing tools, but in the spirit of Section 3.5, it may be addressed with a combination of 3D shape modelling with estimation procedures endowed with shape priors. While the general computer vision community has started to consider such complicated tasks in controlled conditions (Dambreville et al., 2008; Dhome et al., 1989; Rosenhahn and Sommer, 2004; Zerroug and Nevatia, 1995), it is up to the agrovision community to extend it to more realistic scenarios incorporating fragmentation (due to occlusion) and spurious data.

Clearly, some instances of the problem may be easier than others (relatively speaking). Consider the case of a harvesting robot operating in an apple orchard which was grown under a specific dilution policy which guarantees a single fruit in a bud

[Figure 17(a)]. For a harvesting robot it would be constructive to estimate the 3D pose of each apple, and since apples may be modelled reasonably well as spheres or ellipsoids, the segmentation of each apple in the 2D image (using traditional cues and 2D shape analysis, c.f., Section 3.5) and the detection of its upper or bottom part based on appearance may lead to results of the sort illustrated in Figure 17(b). Clearly, at the expected level of occlusions and shape variation one might encounter in a sweet pepper plantation, the same task there may be much more difficult. One way or another, next to robust handling of occlusions, handling of 3D shape is likely to become a major challenge in future agrovision research for autonomous harvesting robots.

Figure 17 3D pose estimation for agrobotics, (a) a harvesting robot might need more than just segmentation of the image regions where fruit may be (b) rather, a complete 3D pose estimation of the sort shown here (shape model fitted to image data and axis points along the centre of the fruit towards the peduncle) might be required for proper grasping and picking operations, the result shown here is purely illustrative and is *not* a result of any existing algorithm (see online version for colours)

(a) (b)

5.7 *Performance evaluation*

A key aspect in machine vision algorithms, especially in the applicative domain, is the evaluation of their performance. Unfortunately, however, most papers in the literature reviewed here neither report performance measures nor conduct a comparative evaluation. At best, performance is reported on small sets of test sequences and virtually never in a comparative manner to other algorithms. Furthermore, no attempt is done to test algorithms in similar field conditions to previous work, hence rendering any retrospective comparison using published data of little value.

Given the current mode of activity, the machine vision researchers in the agrobotics community face two main challenges in terms of performance evaluation. First, a common benchmark dataset must be crafted for all crops of interest under most important environmental and growing conditions. Second, clear, subjective, and quantitative evaluation measures must be defined, and ground truth data pertaining to these measures must be associated with the benchmark dataset. All these can then allow the evaluation of algorithms both relative to each other and against absolute desired performance. For example, when it comes to fruit detection, a large set of images of various crops, at different growing stages, different illumination conditions, and a range of occlusion levels, should be prepared and associated with ground truth detection maps.

These maps can then be used to evaluate algorithms for hits (true positives) and misses (false negative), as well as for false positives and localisation and shape errors, all in a statistically meaningful way. This is by no means a trivial task, but one that should be given priority to facilitate more systematic research and reusable ideas.

6 Summary

Computer vision for agrobotics, and especially for selective harvesting robots, is an exciting research domain with challenges that few other applicative disciplines can offer. Here we reviewed more than two decades of progress on this problem, identified several trends and the major limitations, described the main challenges ahead and proposed possible directions to explore in order to make autonomous agrobotics a reality.

Acknowledgements

This research was funded in part by the European Commission in the 7th Framework Programme (CROPS GA no. 246252) and the Israeli Ministry of Agriculture. The authors also thank the generous support of the Frankel fund and the Paul Ivanier Centre for Robotics Research at Ben-Gurion University.

References

Allen, B., Curless, B. and Popović, Z. (2003) 'The space of human body shapes: reconstruction and parameterization from range scans', in *Proceedings of SIGGRAPH*.

Annamalai, P., Lee, W. and Burks, T. (2004) 'Color vision system for estimating citrus yield in real-time', in *ASABE Annual International Meeting*, August.

Arima, S., Shibusawa, S., Kondo, N. and Yamashita, J. (2003) 'Traceability based on multi-operation robot; information from spraying, harvesting and grading operation robot', in *Proceedings of the IEEE International Conference on Advanced Intelligent Mechatronics*, July, Vol. 2, pp.1204–1209.

Astrand, B. and Baerveldt, A. (2005) 'A vision based row-following system for agricultural field machinery', *Mechatronics*, Vol. 15, No. 2, pp.251–269.

Ballard, D. (1981) 'Generalizing the Hough transform to detect arbitrary shapes', *Pattern Recognition*, Vol. 13, No. 2, pp.111–122.

Barnard, K., Cardei, V. and Funt, B. (2002) 'A comparison of computational color constancy algorithms. I: methodology and experiments with synthesized data', *IEEE Transactions on Image Processing*, Vol. 11, No. 9, pp.972–984.

Ben-Yosef, G. and Ben-Shahar, O. (2010a) 'A biologically-inspired theory for non-axiomatic parametric curve completion', in *Proceedings of the Asian Conference on Computer Vision*.

Ben-Yosef, G. and Ben-Shahar, O. (2010b) 'Minimum length in the tangent bundle as a model for curve completion', in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Berenstein, R., Shahar, O., Shapiro, A. and Edan, Y. (2010) 'Grape clusters and foliage detection algorithms for autonomous selective vineyard sprayer', *Intelligent Service Robotics*, Vol. 3, No. 4, pp.233–243.

Bhanu, B. and Pavlidis, I. (2004) *Computer Vision Beyond the Visible Spectrum*, Springer, England.

Bin, L., Maohua, W. and Ning, W. (2010) 'Development of a real-time fruit recognition system for pineapple harvesting robots', in *ASABE Annual International Meeting*, June.

Blake, A. (1992) *Active Vision*, The MIT Press, Cambridge, MA, USA.

Bregman, A.S. (1981) 'Asking the "what for" question in auditory perception', in M. Kubovy and J.R Pomerantz (Eds.): *Perceptual Organization*, pp.99–118, Lawrence Erlbaum Associates.

Bulanon, D.M. and Kataoka,T. (2010) 'Fruit detection system and an end effector for robotic harvesting of Fuji apples', *International commission of Agricultural and Biosystems Engineering Journal*, Vol. 12, No. 1, pp.203–210.

Bulanon, D.M., Kataoka, T., Zhang, S., Ota, Y. and Hiroma, T. (2001) 'Optimal thresholding for the automatic recognition of apple fruits', *ASAE Annual Meeting*, Vol. 8.

Bulanon, D.M., Kataoka, T., Ota, Y. and Hiroma, T. (2002) 'A segmentation algorithm for the automatic recognition of Fuji apples at harvest', *Biosystems Engineering*, Vol. 83, No. 4, pp.405–412.

Bulanon, D.M., Kataoka, T., Okamoto, H. and Hata, S. (2004a) 'Determining the 3-d location of the apple fruit during harvest', in *Proceedings of the Automation Technology for Off-Road Equipment Conference*, October.

Bulanon, D.M., Kataoka, T., Okamoto, H. and Hata, S. (2004b) 'Development of a real-time machine vision system for the apple harvesting robot', in *Proceedings of the SICE Annual Conference*, August.

Bulanon, D.M., Burks, T. and Alchanatis, V. (2009) 'Image fusion of visible and thermal images for fruit detection', *Biosystems Engineering*, Vol. 103, No. 1, pp.12–22.

Cao, X., Tong, X., Dai, Q. and Lin, S. (2011) 'High resolution multispectral video capture with a hybrid camera system', in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Caselles, V., Kimmel, R. and Sapiro, G. (1995) 'Geodesic active contours', *International Journal of Computer Vision*, Vol. 22, No. 1, pp.61–79.

Chan, T.F. and Vese, L.A. (2001) 'Active contours without edges', *IEEE Transactions on Image Processing*, Vol. 10, No. 2, pp.266–277.

Chen, S. and Radke, R. (2009) 'Level set segmentation with both shape and intensity priors', in *Proceedings of the IEEE International Conference on Computer Vision*, September.

Chi, Y. and Ling, P. (2004) 'Fast fruit identification for robotic tomato picker', in *ASABE Annual International Meeting*, August.

Chinchuluun, R. and Lee, W. (2006) 'Citrus yield mapping system in natural outdoor scenes using the watershed transform', in *ASABE Annual International Meeting*, July.

Cohen, L.D. and Kimmel, R. (1999) 'Global minimum for active contour models: a minimal path approach. *International Journal of Computer Vision*, Vol. 24, No. 1, pp.57–78.

Cohen, L.D. (1991) 'On active contour models and balloons', *Computer Vision, Graphics and Image Processing*, Vol. 53, No. 2, pp.221–218.

Cristianini, N. and Shawe-Taylor, J. (2000) *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*, Cambridge University Press, New York, NY, USA.

Dambreville, S., Sandhu, R., Yezzi, A. and Tannenbaum, A. (2008) 'Robust 3d pose estimation and efficient 2d region-based segmentation from a 3d shape prior', in *Proceedings of the European Conference on Computer Vision*.

Dekker, L., Douros, I., Buston, B.F. and Treleaven, P. (1999) 'Building symbolic information for 3d human body modeling from range data', in *Second International Converence on 3-D Digital Imaging and Modeling*.

Dhome, M., Richetin, M., Lapreste, J-T. and Rives, G. (1989) 'Determination of the attitude of 3D objects from a single perspective view', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 11, No. 12, pp.1265–1278.

Edan, Y., Rogozin, D., Flash, T. and Miles, G. (2000) 'Robotic melon harvesting', *IEEE Transactions on Robotics and Automation*, December, Vol. 16, No. 6, pp.831–835.

Fang, W. and Chan, K. (2007) 'Incorporating shape prior into geodesic active contours for detecting partially occluded object', *Pattern Recognition*, Vol. 40, No. 8, pp. 2163–2172.

Feng, G., Qixin, C. and Masateru, N. (2008) 'Fruit detachment and classification method for strawberry harvesting robot', *International Journal of Advanced Robotic Systems*, November, Vol. 5, No. 1, pp.41–48.

Forsyth, D. and Ponce, J. (2002) *Computer Vision: A Modern Approach*, Prentice Hall, Upper Saddle River, NJ, USA.

Foulonneau, A., Charbonnier, P. and Heitz, F. (2009) 'Multi-reference shape priors for active contours', *International Journal of Computer Vision*, Vol. 81, No. 1, pp.68–81.

Gastaud, M., Barlaud, M. and Aubert, G. (2004) 'Combining shape prior and statistical features for active contour segmentation', *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 14, No. 5, pp.726–734.

Gehler, P., Rother, C., Blake, A., Minka, T. and Sharp, T. (2008) 'Bayesian color constancy revisited', in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Gerig, G. (1987) 'Linking image-space and accumulator-space: A new approach for object-recognition', in *Proceedings of the IEEE International Conference on Computer Vision*.

Gijsenij, A. and Gevers, T. (2007) 'Color constancy using natural image statistics', in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.1–8.

Gurney, K. (1997) *An Introduction to Neural Networks*, CRC Press, Bristol, PA, USA.

Hannan, M. and Burks, T. (2004) 'Current developments in automated citrus harvesting', in *ASABE Annual International Meeting*, August.

Hannan, M., Burks, T. and Bulanon, D.M. (2007) 'A real-time machine vision algorithm for robotic citrus harvesting', in *ASABE Annual International Meeting*, June.

Haralick, R.M. and Shapiro, L.G. (1985) 'Image segmentation techniques', *Computer Vision, Graphics and Image Processing*, Vol. 1, No. 1, pp.100–132.

Harrell, R., Slaughter, D. and Adsit, P. (1989) 'A fruit-tracking system for robotic harvesting', *Machine Vision and Applications*, Vol. 2, No. 2, pp.69–80.

Hartley, R. and Zisserman, A. (2000) *Multiple View Geometry in Computer Vision*, Cambridge University Press, New York, NY, USA.

Hayashi, S., Ganno, K., Ishii, Y. and Tanaka, I. (2002) Robotic harvesting system for eggplants. *Japan Agricultural Research Quarterly*, Vol. 36, No. 3, pp.163–168.

Horn, B.K.P. (1983) 'The curve of least energy', *ACM Transactions on Mathematical Software*, Vol. 9, No. 4, pp.441–460.

Huang, Y. and Lee, F. (2010) 'An automatic machine vision-guided grasping system for phalaenopsis tissue culture plantlets', *Computers and Electronics in Agriculture*, Vol. 70, No. 1, pp.42–51.

Illingworth, J. and Kittler, J. (1988) 'A survey of the Hough transform', *Computer Vision, Graphics and Image Processing*, Vol. 44, No. 1, pp.87–116.

Jain, A.K. and Farrokhnia, F. (1991) 'Unsupervised texture segmentation using Gabor filters', *Pattern Recognition*, Vol. 24, No. 12, pp.1167–1186.

Jiang, H., Peng, Y., Shen, C. and Ying, Y. (2008a) 'Study of area-based stereovision method for locating tomato in greenhouse', in *ASABE Annual International Meeting*, July.

Jiang, H., Peng, Y. and Ying, Y. (2008b) 'Measurement of 3-d locations of ripe tomato by binocular stereo vision for tomato harvesting', in *ASABE Annual International Meeting*, July.

Jimenez, A., Ceres, R. and Pons, J. (2000a) 'A vision system based on a laser range-finder applied to robotic fruit harvesting', *Machine Vision and Applications*, Vol. 11, No. 6, pp.321–329.

Jimenez, A., Ceres, R. and Pons, J. (2000b) 'A survey of computer vision methods for locating fruit on trees', *Transactions of the ASABE*, Vol. 43, No. 6, pp.1911–1920.

Kane, K. and Lee, W. (2006) 'Spectral sensing of different citrus varieties for precision agriculture', in *ASABE Annual International Meeting*, July.

Kane, K. and Lee, S. (2007) 'Multispectral imaging for in-field green citrus identification', in *ASABE Annual International Meeting*, June.

Kang, Y., Lee, C. and Ho, Y. (2008) 'An efficient rectification algorithm for multi-view images in parallel camera array', in *Proceedings of 3DTV Conference*.

Kanizsa, G. (1979) *Organization in Vision: Essays on Gestalt Perception*, Praeger Publishers. West Port, NY, USA.

Kass, M., Witkin, A. and Trezopoulos, D. (1988) 'Snakes: Active contour models', *International Journal of Computer Vision*, Vol. 1, No. 4, pp.321–331.

Kimia, B.K., Frankel, I. and Popescu, A.M. (2003) 'Euler spiral for shape completion', *International Journal of Computer Vision*, Vol. 54, Nos. 1–3, pp.159–182.

Kimmel, R. and Bruckstein, A.M. (2003) 'Regularized Laplacian zero crossing as optimal edge integrators', *International Journal of Computer Vision*, Vol. 53, No. 3, pp.225–243.

Kitamura, S., Oka, K., Ikutomo, K., Kimura, Y. and Taniguchi, Y. (2008) 'A distinction method for fruit of sweet pepper using reflection of led light', in *Proceedings of the SICE Annual Conference*, August.

Kittel, C. and Kroemer, H. (1980) *Thermal Physics*, W.H. Freeman, San Francisco, USA.

Kondo, N., Monta, M. and Fujiura, T. (1996) 'Fruit harvesting robots in Japan', *Advances in Space Research*, Vol. 18, Nos. 1–2, pp.181–184.

Kondo, N., Yamamoto, K., Yata, K. and Kurita, M. (2008) 'A machine vision for tomato cluster harvesting robot', in *ASABE Annual International Meeting*, July.

Kong, D., Zhao, D., Zhang, Y., Wang, J. and Zhang, H. (2010) 'Research of apple harvesting robot based on least square support vector machine', in *Proceedings of the International Conference on Electrical and Control Engineering*, June, pp.1590–1593.

Ling, P., Ehsani, R., Ting, K., Chi, Y., Ramalingam, N., Klingman, M. and Draper, C. (2004) 'Sensing and end-effector for a robotic tomato harvester', in *ASABE Annual International Meeting*, August.

Liu, Y., Chen, B. and Qiao, J. (2011) 'Development of a machine vision algorithm for recognition of peach fruit in a natural scene', *Transactions of the ASABE*, Vol. 54, No. 2, pp.695–702.

Malik, J., Belongie, S., Leung, T. and Shi, J. (2001) 'Contour and texture analysis for image segmentation', *International Journal of Computer Vision*, Vol. 43, No. 1, pp.7–27.

Manolakis, D., Lockwood, R., Cooley, T. and Jacobson, J. (2009) 'Is there a best hyperspectral detection algorithm?', in *Proceedings of the SPIE*.

Marsland, S. (2009) *Machine Learning: An Algorithmic Perspective*, Chapman and Hall/CRC, Boca Raton, Florida, USA.

Martin, D.R., Fowlkes, C.C. and Malik, J. (2004) 'Learning to detect natural image boundaries using local brightness, color, and texture cues', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 26, No. 1, pp.530–549.

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

Morimoto, T., Takeuchi, T., Miyata, H. and Hashimoto, Y. (2000) 'Pattern recognition of fruit shape based on the concept of chaos and neural networks', *Computers and Electronics in Agriculture*, Vol. 26, No. 2, pp.171–186.

Mumford, D. (1994) 'Elastica in computer vision', in B. Chandrajit (Ed.): *Algebric Geometry and it's Applications*, Springer-Verlag.

Nalwa, V.S. (1993) *A Guided Tour of Computer Vision*, Addison-Wesley Publishing, Reading, MA, USA.

Okamoto, H. and Lee, W. (2009) 'Green citrus detection using hyperspectral imaging', *Computers and Electronics in Agriculture*, Vol. 66, No. 2, pp.201–208.

Okamoto, H. and Lee, W. (2010) 'Machine vision for green citrus detection in tree images', *Environment Control in Biology*, Vol. 48, No. 2, pp.93–99.

Osher, S., Paragios, N. and Kimmel, R. (2003) 'Fast edge integration', in *Geometric Level Set Methods in Imaging, Vision, and Graphics*, pp.59–77, Springer New York.

Palmer, S.E. (1999) *Vision Science: Photons to Phenomenology*, The MIT Press, Cambrdige, MA, USA.

Papageorgiou, C., Oren, M. and Poggio, T. (1998) 'A general framework for object detection', in *Proceedings* of the IEEE International Conference on Computer Vision.

Patel, H., Jain, R. and Joshi, M. (2011) 'Fruit detection using improved multiple features based algorithm', *International Journal of Computer Applications*, Vol. 13, No. 2, pp.1–5.

Paulus, I. and Schrevens, E. (1999) 'Shape characterization of new apple cultivars by Fourier expansion of digitized images', *International Journal of Computer Applications*, Vol. 72, No. 2, pp.113–118, JAER.

Plebe, A. and Grasso, G. (2001) 'Localization of spherical fruits for robotic harvesting', *Machine Vision and Applications*, Vol. 13, No. 2, pp.70–79.

Qiao, J., Sasao, Shibusawa, S., Kondo, N. and Morimoto, E. (2005) 'Mapping yield and quality using the mobile fruit grading robot', *Biosystems Engineering*, Vol. 90, No. 2, pp.135–142.

Rakun, J., Stajnko, D. and Zazula, D. (2011) 'Detecting fruits in natural scenes by using spatial-frequency based texture analysis and multiview geometry', *Computers and Electronics in Agriculture*, Vol. 76, No. 1, pp.80–88.

Regunathan, M. and Lee, W. (2005) 'Citrus fruit identification and size determination using machine vision and ultrasonic sensors', in *ASABE Annual International Meeting*, July.

Rein-Lien, H. and Jain, A. (2001) 'Face modeling for recognition', in *Proceedings of the IEEE International Conference on Image Processing*.

Rizon, M., Yazid, H., Saad, P., Shakaff, A. and Saad, A. (2005) 'Object detection using circular Hough transform', *American Journal of Applied Sciences*, Vol. 2, No. 12, pp.1606–1609.

Rosenhahn, B. and Sommer, G. (2004) 'Pose estimation of free-form objects', in T. Pajdla and J. Matas (Eds.): *Proceedings of the European Conference on Computer Vision*, pp.414–427, Lecture Notes in Computer Science, Springer Berlin/Heidelberg.

Rumpf, M. and Strzodka, R. (2001) 'Level set segmentation in graphics hardware', in *Proceedings of the IEEE International Conference on Image Processing*.

Safren, O., Alchanatis, V., Ostrovsky, V. and Levi, O. (2007) 'Detection of green apples in hyperspectral images of apple-tree foliage using machine vision', *Transactions of the ASABE*, Vol. 50, No. 6, pp.2303–2313.

Shotton, J., Winn, J., Rother, C. and Criminisi, A. (2009) 'Textonboost for image understanding: multi-class object recognition and segmentation by jointly modeling texture, layout, and context', *International Journal of Computer Vision*, Vol. 81, No. 1, pp.2–23.

Slaughter, D., Giles, D. and Downey, D. (2008) 'Autonomous robotic weed control systems: a review', *Computers and Electronics in Agriculture*, Vol. 61, No. 1, pp.63–78.

Stajnko, D., Lakota, M .and Hocevar, M. (2004) 'Estimation of number and diameter of apple fruits in an orchard during the growing season by thermal imaging', *Computers and Electronics in Agriculture*, Vol. 42, No. 1, pp.31–42.

Takahashi, T., Zhang, S. and Fukuchi, H. (2002) 'Measurement of 3-d locations of fruit by binocular stereo vision for apple harvesting in an orchard', in *ASABE Annual International Meeting*, July.

Tanigaki, K., Fujiura, T., Akase, A. and Imagawa, J. (2008) 'Cherry harvesting robot', *Computers and Electronics in Agriculture*, Vol. 63, No. 1, pp.65–72.

Tao, D., Song, M., Li, X., Shen, J., Sun, J., Wu, X., Faloutsos, C. and Maybank, S. (2008) 'Bayesian tensor approach for 3-d face modeling', *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 18, No. 10, pp.1397–1410.

Tarrio, P., Bernardos, A., Casar, J. and Besada, J. (2006) 'A harvesting robot for small fruit in bunches based on 3-d stereoscopic vision', in *Proceedings of the World Congress Conference on Computers in Agriculture and Natural Resources*, July.

Tellaeche, A., BurgosArtizzu, X., Pajares, G., Ribeiro, A. and Fernandez-Quintanilla, C. (2008) 'A new vision-based approach to differential spraying in precision agriculture', *Computers and Electronics in Agriculture*, Vol. 60, No. 2, pp.144–155.

Ullman, S. (1976) 'Filling in the gaps: the shape of subjective contours and a model for their creation', *Biological Cybernetics*, Vol. 25, No. 1, pp.1–6.

van de Weijer, J., Schmid, C. and Verbeek, J. (2007) 'Bayesian color constancy revisited', in *Proceedings of the IEEE International Conference on Computer Vision*, pp.1–8.

Van Henten, E., Hemming, J., van Tuijl, B., Kornet, J., Meuleman, J., Bontsema, J. and van Os, E. (2002) 'An autonomous robot for harvesting cucumbers in greenhouses', *Autonomous Robots*, Vol. 13, No. 3, pp.241–258.

Wachs, J., Stern, H., Burks, T. and Alchanatis, V. (2010) 'Low and high-level visual feature based apple detection from multi-modal images', *Precision Agriculture*, Vol. 11, No. 6, pp.717–735.

Wang, X. and Paliwal, K. (2003) 'Feature extraction and dimensionality reduction algorithms and their applications in vowel recognition', *Pattern Recognition*, Vol. 36, No. 10, pp.2429–2439.

Wei, J., Rovira-Mas, F., Reid, J. and Han, S. (2005) 'Obstacle detection using stereo vision to enhance safety of autonomous machines', *Transactions of the ASABE*, Vol. 48, No. 6, pp.2389–2397.

Wertheimer, M. (1955) 'Laws of organization in perceptual forms', in W.D. Ellis (Ed.): *A Source Book of Gestalt Psychology*, pp.1–88, Routledge & Kegan Paul Ltd.

Weszka, J. (1978) 'A survey of threshold selection techniques', *Computer Graphics and Image Processing*, Vol. 7, No. 2, pp.259–265.

Wyszecki, G. and Stiles, W. (2000) *Color Science: Concepts and Methods, Quantitative Data and Formulae*, Wiley-Interscience, New York, USA.

Xu, C. and Prince, J.I. (1998) 'Snakes, shapes, and gradient vector flow', *IEEE Transactions on Image Processing*, Vol. 7, No. 3, pp.359–369.

Yang, A., Wright, J., Ma, Y. and Sastry, S. (2008) 'Unsupervised segmentation of natural images via lossy data compression', *Computer Vision and Image Understanding*, Vol. 110, No. 2, pp.212–225.

Yongjie, C., Masateru, N., Feng, G., Kenji, H., Osamu, K. and Masafumi, M. (2010) 'An automatic machine vision-guided grasping system for phalaenopsis tissue culture plantlets', *Computers and Electronics in Agriculture*, Vol. 70, No. 1, pp.42–51.

Yu, J. and Yang, H. (2001) 'A direct LDA algorithm for high-dimensional data – with application to face recognition', *Pattern Recognition*, Vol. 34, No. 10, pp.2067–20701.

Yuan, T., Wei, L., Qingchun, F. and Junxiong, Z. (2010) 'Spectral imaging for greenhouse cucumber fruit detection based on binocular stereovision', in *ASABE Annual International Meeting*, June.

Zerroug, M. and Nevatia, R. (1995) 'Pose estimation of multi-part curved objects', in *IEEE International Symposium on Computer Vision*.

Zhang, F. and Zhang, N. (2008) 'Applying joint transform correlator in tomato recognition', in *ASABE Annual International Meeting*, July.

Zhao, C. and Jiang, G. (2011) 'Baseline detection and matching to vision-based navigation of agricultural robot', *Computer and Computing Technologies in Agriculture*, Vol. 344, No. 1, pp.138–148.

Zhao, J., Tow, J. and Katupitiya, J. (2005) 'On-tree fruit recognition using texture properties and color data', in *IEEE International Workshop on Intelligent Robots and Systems*, August.

Zhen, L., Hong, T., Wei-xing, W. and Shu-ran, S. (2007) 'Automatic detection of growing orange fruits by machine vision', in *ASABE Annual International Meeting*, June.

Zheng, N. and Xue, J. (2009) *Statistical Learning and Pattern Analysis for Image and Video Processing*, Springer, London, England.