A Fruit Recognition Method for Automatic Harvesting

L. Yang, J. Dickinson, Q. M. J. Wu, S. Lang

Abstract — *This paper presents a method to detect and recognize mature tomato fruit clusters on a complex-structured tomato plant containing clutter and occlusion in a tomato greenhouse for automatic harvesting purpose. A color stereo vision camera (PGR BumbleBee2) is applied as the vision sensor. The proposed method performs a 3D reconstruction with the data collected by the stereo camera to create a 3D environment for further processing. The Color Layer Growing (CLG) method is introduced to segment the mature fruits from the leaves, stalks, background and noise. Target fruit clusters can then be located by depth segmentation. The experimental data was collected from a tomato greenhouse and the method is justified by the experimental results. Our experiments included severe self and stereo occlusion, which wasn't both included in the previous work.*[1]

Index Terms — Fruit Recognition, Object Localization, Color Segmentation, 3D Reconstruction.

I. INTRODUCTION

A. BACKGROUND

Currently, labour costs are one of the largest costs incurred in modern greenhouse operations, making up about 35% of operational costs, and this pushes the growers to constantly look for alternatives to replace intensive human operations with automation. Before automation of harvesting or inspection tasks can be considered a reliable method of identifying fruit clusters on plants is required.

In this paper, we propose using color and depth segmentation in a 3D reconstructed environment to identify the position of mature fruit clusters with occlusion in a complex-structured tomato plant. This novel method provides us with the ability to detect mature fruit clusters with severe occlusion including self and stereo occlusion. The application can be expanded to other fruits.

The remainder of this section briefly outlines the methods previously used in horticultural operations. Section 2 introduces our method in detail before Section 3 presents the experimental results. Section 4 closes by summarizing our conclusions and makes recommendations for future work.

[1] This work was supported by the University of Windsor and Integrated Manufacturing Technology Institute, National Research Council

L. Yang and Q. M. J. Wu are with the Electrical and Computer Engineering, University of Windsor, Windsor, ON, Canada (e-mail: yang111@uwindsor.ca; jwu@uwindsor.ca).

J. Dickinson and S. Lang are with the Integrated Manufacturing Technology Institute, National Research Council, London, ON, Canada (john.dickinson@nrc-cnrc.gc.ca; sherman.lang@nrc-cnrc.gc.ca)

B. PREVIOUS WORK

A good review of fruit recognition approaches can be found in [12] though a few representative works are described here. Selective harvesting requires the determination of the location of ripe fruit clusters, which are assessed based on four basic characteristics: intensity, color, shape and texture.

Sites[1] and Slaughter and Harrel[2] used intensity/color based methods to recognize mature fruit. In their work, a thresholding technique is employed to obtain a binary image; afterwards, filter smoothing is applied to eliminate noise and irrelevant details. Then the recognition of ripe fruit is made either based on the size of the segmented region or feature recognition.

Whitaker[3] and Benady[4] applied shape information to detect fruit. In their work, a Circular Hough Transform is applied to binary edge images to obtain a matrix of votes indicating the candidates for being the center of a melon, thus detect arc segments in the images, however their approach is less applicable in a complex structured plant because of false positives created by other parts of the plant.

Qiu and Shearer[5] and Purdue University and The Volcani Center[6] combined shape and texture to recognize fruits due to the fact that some fruit has textures different from its leaves. Texture analysis has been used to locate some specific fruit, such as broccoli in their work.

Grand D'Esnon[7] developed a vision system for the MAGALI robot, which applied three color CCD cameras and three different filters (950, 650 and 550 nm) to obtain three different intensity images. Ratio features are then used to decide which pixels belong to a fruit or to a leaf based on a preliminary spectral property study.

The AID robot vision system [8] recognized oranges by color and intensity gradient characteristics. Recognition is accomplished by matching gradient image data with a previously stored object model.

Finally, Jun Zhao, Joel Tow and Jayantha Katupitiya [13] applied color (redness measure), texture properties and shape (circle fitting) with a Laplacian filter to locate fruits.

II. METHODS

A. 3D RECONSTRUCTION

In our work, a BumbleBee2 [10] color stereo camera is applied as vision sensor and the data flow goes as Fig. 1.

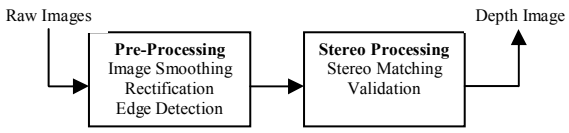


Fig. 1 3D reconstruction diagram

Pre-processing prepares the raw images for stereo processing. A low-pass filter is applied to smooth the images in order to rectify the images. Rectification is the process of correcting input images for the distortions of the lenses. Since the stereo cameras are placed horizontally, rectification aligns rows of pixels to make searching along rows for stereo matching practical. Edge detection is applied to allow matching on the changes in the brightness. This feature is useful because the auto gain of two cameras may lead to different absolute brightness values of the same pixel while the change in the intensity stays nearly constant. Thus difference comparisons are also applicable in environments where the lighting conditions change significantly.

The stereo processing module applies the Sum of Absolute Differences algorithm [11] for stereo matching.

$$\min_{d\,\min}^{d\,\max} \sum_{i=-m/2}^{m/2} \sum_{j=-m/2}^{m/2} | I_{right}[x+i][y+j] - I_{left}[x+i+d][y+j] \qquad (1)$$

where d_{max} and d_{min} are the minimum and maximum disparities; m is the mask size; I_{right} and I_{left} are the right and left images.

There are a number of parameters which are used to determine the type of depth image produced. Disparity range is the range of pixels that the stereo algorithm searches in order to find the best match, where zero pixel corresponds to an infinitely far away object. In our case, disparity only appears in x (columns) because the two cameras are vertically aligned, resulting in close to zero vertical disparity. Correlation mask is a square neighborhood around the pixel that the system is trying to match in the second image. The impact of various sized masks is described in the Table 1.

TABLE 1
MASK SIZE SELECTION

	Large mask	Small mask
Disparity maps	Denser and smoother	Sparser and noisier
Depth discontinuities	Less precise	Better localization
Computational cost	Larger	Smaller

In some cases, such as occlusions and lack of texture, it is impossible to establish correspondence between portions of images. In order to avoid incorrect measurements, Texture, Uniqueness and Surface Validation are introduced.

Texture validation determines whether disparity values are valid based on levels of texture in the correlation mask. Texture validation is based on the amount of edge strength in the stereo mask. Its main purpose is to reject regions which have little or no texture. Without sufficient texture information, the system is incapable of producing correct matches between left and right views.

Uniqueness validation determines whether the best match for a particular pixel is significantly better than other matches within the correlation mask, where a threshold judges if the correlation result is strong enough to declare the pixel valid.

Surface Validation removes the so called "spikes" in a disparity image. This noise is difficult to remove with standard filtering techniques as it is not zero-mean, random, evenly distributed or Gaussian in nature. For large connected regions of pixels in a disparity image surface validation is a method to validate of the pixels based on the assumption that they must belong to a likely physical surface in the image. The method is based on the attributes of these errors: the error regions are locally stable but not large and are characterized by sharp disparity discontinuities at all borders. In the disparity image, pixels are connected to form a region if they are neighbors and their absolute disparity difference value is within a certain threshold. After segmenting the disparity image into connected regions, any region less than a given size is regarded as a spark and removed from the disparity image The method is described in detail in the following steps:

i = L ; j = L ;
For all j ∈ N(i)
If | $d_j - d_i$ | ≤ 1
Then group i and j
Count number of pixels with Label L
If number > t (threshold)
Then it is valid

Where i is any given pixel, L is a surface label, N(i) is a neighborhood of pixels around i and d_i is the disparity value at location i. Entire surfaces are invalidated from the disparity image if the number of pixels that have a given label do not pass a threshold.

Triangulation is a classical method to calculate 3D XYZ position based on the disparity map we retrieve from the previous steps [11].

B. COLOR SEGMENTATION (CLG)

Image segmentation is a key step towards object recognition in image analysis. Common methods include edge-detection, region-growing and clustering techniques. Region growing methods can be classified as local and global techniques. Our method combines the advantages of local (simplicity and quickness) and global (robustness and accuracy) techniques. By bottom-up local layer growing and repeating distance measurements at higher levels we connect regions of the same level by applying the global aggregated distance and color similarity information.

A basic concept in this approach is islands. An island of level 0 simply denotes a single pixel and an island of level 1 is a set of level 0 pixels in orthogonal topology. Likewise, an island of level n+1 is a set of islands of level n.

Color Layer Growing (CLG) is developed based on Color structure code (CSC) [9] which can segment an image by linking its homogeneous regions and splitting the regions by the similarity in color. CLG segmentation algorithm operates essentially in the following phases: preprocessing, initialization and linking. In the preprocessing phase, color band selection is accomplished by a color band filter and noise suppression is implemented by a nonlinear filter. In an initialization phase the image is partitioned into small color

regions within an island of level 1. These small color regions are growing in the linking phase to complete regions.

In preprocessing, as mature fruit in the greenhouse has its specific color band, filtering the color within this specific color band can eliminate the complicated plant structure, including stalks, stems, leaves, etc. The remaining image after filtering might consist of target tomato fruit clusters, background components and noise.

Symmetric Nearest Neighbor (SNN) is applied for noise cancellation. We have 8 neighbors for a center point, constituting four center symmetric pairs (c1,c8), (c2,c7), (c3,c6) and (c4,c5). From each pair we choose the pixel whose color value is closer to the central pixels color value. The central pixel is then replaced by the equally weighted average of the three determined neighbors. Let vi = (ri,gi,bi), $1 \le i \le 4$, be the three color vectors with,

v1 = f(c1), if $\| f(c0)-f(c1) \| \le \| f(c0)-f(c8) \|$
= f(c8), otherwise
v2 = f(c2), if $\| f(c0)-f(c2) \| \le \| f(c0)-f(c7) \|$
= f(c7), otherwise
v3 = f(c3), if $\| f(c0)-f(c3) \| \le \| f(c0)-f(c6) \|$
= f(c6), otherwise
v4 = f(c4), if $\| f(c0)-f(c4) \| \le \| f(c0)-f(c5) \|$
= f(c5), otherwise

$$f(c0)new = \frac{1}{4}\left(\sum_{1}^{4} ri, \sum_{1}^{4} gi, \sum_{1}^{4} bi\right) \quad (2)$$

In the initialization phase, regions grow with every 9 pixels. Such a region growing consists of those pixels of level 0 that are neighbored and whose mutual color distance lies below a certain threshold, thus forming a level 1 island. This operation is a purely local operation, independently processed for each island. Instead of starting with one seed pixel, the CLG starts in all islands of the preprocessed image. The result of the initialization phase is a set of new regions of level 1, with each one describing a small color patch. In the following linking phase these small color patches are checked for color similarity and continuity to grow hierarchically to complete color segments.

At the beginning of linking phase, elements of level 1 are given the same label if they are physically connected. Then close islands of level n (n>1) are linked to new islands of level n+1. Color similarity and island distance within island structures become the process to determine which islands of the same level should be grouped into a higher level. Only regions within the sub-regions of a higher level region are candidates to be connected. The color similarity between two candidate regions r1 and r2 and the region distance are both measured to decide whether they should be connected.

The color similarity measure employs a color predicate P. The color features are in RGB color space. Let c(r) be the mean of all color contents within that region. According to the color distance of the two regions, r1 and r2 are connected if the color distance P(c(r1), c(r2)) < t for some threshold. While color similarity is one criterion to decide whether two regions should be connected, region center distance is the other factor, wherein thresholds for selecting proper candidates vary according to the different level they belong to.

Those color regions which are too small are not considered for the further layer growing, meaning only large region (over certain threshold) can participate in region growing to form higher level layer. Islands that do not find any partner for linking on level n and are under certain size are discarded from the image. The linking operations are repeated for all islands on every level, starting from level 1 until the size of the region is large enough to be considered as the target object. Thus, color segmentation results in connected regions with similar color.

C. DEPTH SEGMENTATION

Color segmentation can result in a few target candidates, but the process of converting a 2D pixel to a 3D position in space may not be very accurate, since the pixel position of the target is a mean value.

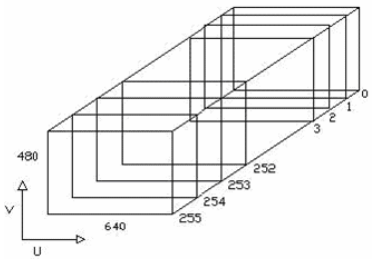


Fig. 2 Image plane levels

Depth Segmentation, a combination of 3D reconstruction and color segmentation we are introducing in this section, is able to get a 3D point-cloud of the target fruits with little noise. Consider a 3D picture as 480 pixels high, 640 pixels wide and 256 (0-255) levels in depth. Referring to Fig. 2, a depth map is visualized as 256 separate images and each depth image has information about the image at that depth.

Shape based recognition is efficient and accurate by using depth histogram representation. The first step is to derive depth histograms of objects of interest. The statistical distribution of the depth histogram of targeted fruit can be used to characterize the shape of target objects. A fruit sphere has most the points in its center depth level and fewer points on its front which faces the camera. Color segmentation can be used to isolate the pixels representing fruit in the 3D point cloud. Then, according to the shape property of the fruit spheres, points scattered on different levels roughly appear in normal distributions as shown in Fig. 3, where there are two separate target fruit clusters at different distances. The two peaks are exactly where the fruit clusters are located. This reduces the object position detection problem to a local maxima detection problem.

Given the noise in the distance histogram curve, it is divided into groups and a local maxima is found for each group to yield the distance of each cluster from the camera.

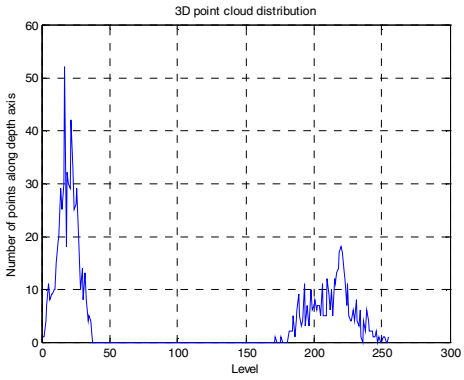


Fig. 3 3D point cloud depth histogram

III. EXPERIMENTAL RESULTS

A commercial tomato greenhouse has many long rows of plants between which carts can roll on rails to support workers doing maintenance and harvesting tasks. Usually the cart is designed with the capacity to raise and lower the worker and move forward and backward along the tracks. During our experiments the cart height was adjusted to about 55cm and the mature fruit clusters were between 60 and 100 cm in height. Since the distance between two rows is around 155cm, the camera was mounted on the cart at three different distances (close, medium and far, around 32~35cm, 50~53cm and 65~68cm respectively) for every single scene to test the robustness of the method.

The experiments were run during when the tomato plants were quite mature in April 2007 in a tomato greenhouse in South Western Ontario. The gathered data consists of raw images, stereo input images and calibration files. The image resolution used was 600*480.

In Fig. 4, there is one mature and one immature fruit cluster in both the left and right images. The fruit on the right bottom corner in Fig. 4(b) does not appear in Fig. 4(a) because of stereo occlusion. With depth filter and color segmentation, the mature fruit cluster was located at 34.55cm with (X, Y, Z) = (-6.16, -0.37, 34) cm as depicted in Fig. 4(d).

In Fig. 5, there are more mature fruit clusters at greater distances with the center cluster being around 65-68cm away. The results indicate that our method is capable of dealing with multiple clusters in the image. The calculated distances of the three clusters were 81.78cm, 67.98cm and 75.96cm (starting from left). And the 3D positions were (-43.67, 1.35, 69.13) cm, (-10.4, 15.91, 70.27) cm, (24.04, 15.91, 70.27) cm respectively.

Table 2 and Fig. 6 evaluate the accuracy and robustness of our method at different distances in various scenarios (stereo/self occlusion). The max and min distances in Fig. 6 are the distance range of target fruit clusters, and the calculated distance (blue line) falls in between the two measured lines, which means that our method can recognize the fruit clusters and correctly calculate their positions in the real world.

Fig. 4 Mature fruits at 32~35cm distance with stereo occlusion (a) raw left image (b) raw right image (c) CLG color segmentation (d) fruit cluster in the image with depth filter

Fig. 5 Mature fruits at 65~68cm distance with stereo occlusion (a) raw left image (b) raw right image (c) CLG color segmentation (d) fruit cluster in the image without depth filter

IV. CONCLUSIONS

In this paper, we propose a novel method to recognize mature fruit and locate cluster positions for greenhouse automatic harvest applications. To recognize fruit, 3D reconstruction eliminates stereo occluded objects and we use CLG to segment fruits from the background in a 3D reconstructed environment. A depth filter is applied to remove irrelevant image information to speed up the processing. A histogram is employed in the depth segmentation to calculate the exact 3D XYZ position of a fruit cluster and thus the 3D position. The selection of a color stereo camera makes the vision sensor not only small in size but also economic in price. We tested the robustness of this method in a real tomato greenhouse with strong sunlight and severe noise, nevertheless we were still able to detect and locate the targets even with stereo and self occlusion. Future work should include integrating the fruit locating system with an automated harvesting system.

TABLE 2
COMPARISON BETWEEN MEASURED DISTANCE RANGE AND CALCULATED DISTANCE

Min Measured Distance (cm)	Max Measured Distance (cm)	Calculated Distance (cm)
32	35	34.55
36	39	38.46
37	40	39.78
38	41	41.08
51	54	53.42
52	55	54.13
53	56	55.43
53	56	55.24
63	66	65.51
65	68	67.98
69	72	71
71	74	73.24

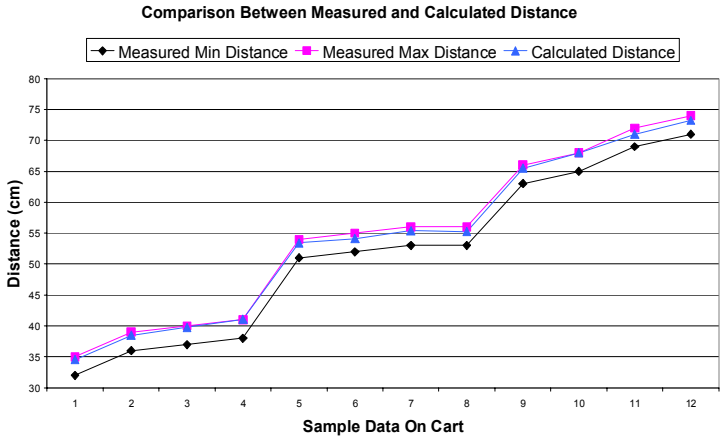


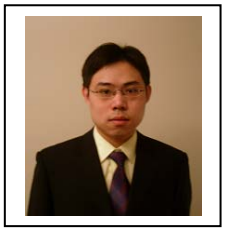
Fig. 6 Comparisons between Measured and Calculated Distance of Fruit Clusters

ACKNOWLEDGEMENTS

Research was conducted in Integrated Manufacturing Technology Institute, National Research Council of Canada (IMTI-NRC) and University of Windsor. Special thanks go to Mr. Shalin Khosla from OMAFRA and cordial help from IMTI-NRC, OCE and OGVG.

REFERENCES

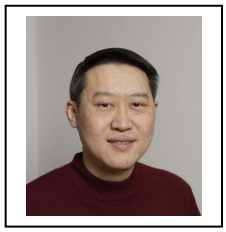
[1]. Sites and Dewilche, Computer vision to locate fruit on a tree, Transactions of the ASAE, Vol.31(1), 257-263 (1988).
[2]. D. Slaughter and R. C. Harrel, Color Vision in Robotic Fruit Harvesting, Transactions of the ASAE, Vol.30(4), 1144-1148 (1987).
[3]. Whitaker, Miles, Mitchell and Gaultney, Fruit location in a partially occluded image, Transactions of the ASAE, Vol. 30(3), 591-597 (1987).
[4]. M. Benady and G. E. Miles, Locating melons for robotic harvesting using structured light, ASAE Paper No.:92-7021 (1992).
[5]. W. Qiu and S.A. Shearer, Maturity assessment of broccoli using the discrete Fourier transform, ASAE Paper No. 91-7005, St. Joseph, MI (1991).
[6]. M. Cardenas-Weber, A. Hetzroni and G.E. Miles, Machine vision to locate melons and guide robotic harvesting, ASAE Paper No. 91-7006 (1991).
[7]. A. Grand D'Esnon, G. Rabatel and R.Pellenc, Magali: A self-propelled robot to pick apples, ASAE paper No. 87-1037, St. Joseph, MI 49085-9659 (1987).
[8]. P. Levi, R. Falla and R. Pappalardo, Image controlled robotics applied to citrus fruit harvesting, Procedures, ROVISEC-VII, Zurich (1988).
[9]. P. Sturm, 3D-color-structure-code-segmentation by using a new non-plainness island hierarchy. IEEE, Vol.2 page(s): 953- 956, Oct. 2004
[10]. Point Grey Research BumbleBee2 Specifications, http://www.ptgrey.com/products/bumblebee2/index.asp
[11]. Introductory Techniques for 3D Computer Vision, Trucco and Verri, Prentice Hall 1998, Page 123-172
[12]. A. R. Jimenez, R. Ceres, and J. L. Pons, "A survey of Computer Vision Methods for Locating Fruit on Trees," ASAE, vol. 43(6): 1911–1920, (2000).
[13]. Zhao, J. Tow, J. Katupitiya, J. , "On-tree fruit recognition using texture properties and color data", 2005 IEEE/RSJ International Conference, page(s): 263- 268, 2-6 Aug. 2005



Linghe Yang received M.A.Sc Degree in Electrical and Computer Engineering from University of Windsor, Windsor, Ontario, Canada in 2007. During the past few years, he has mainly focused on stereo vision, image processing and object detection.

John K. Dickinson (PhD 1999, M. Math 1993, B.Sc. 1991) Research interest is in developing tools to support engineers in the manufacturing and construction industries based on simulation models and optimisation with a particular focus in simulation model development to characterise operations problem diagnosis and improvement.

Jonathan Wu received his Ph.D. degree in electrical engineering from the University of Wales, U.K., in 1990. In 1995, he joined the National Research Council of Canada where he worked as a senior research officer and group leader. He is currently a full Professor in the Department of Electrical and Computer Engineering at the University of Windsor, Canada. He has published over 100 scientific papers in areas of computer vision, neural networks, fuzzy systems, robotics, micro-sensors and actuators, and integrated micro-systems. His current research interests include 3D image analysis, active video object extraction, vision-guided robotics, sensor analysis and fusion, wireless sensor network, and integrated micro-systems. Dr. Wu is a holder of Canada Research Chair in Automotive Sensors and Sensing Systems. He is an Associate Editor for IEEE Transaction on Systems, Man and Cybernetics, Part A. Dr. Wu is also on the Editorial Board of the Journal of Control and Intelligent Systems.



Dr. Sherman Lang is a native of Victoria, B. C., Canada. He obtained B.A.Sc., M.A.Sc. and Ph.D. degrees in Systems Design Engineering from the University of Waterloo. Dr. Lang has held positions with the Laboratory for Biomedical Engineering of the Medical Engineering Section of the Division of Electrical Engineering of the National Research Council of Canada, the Autonomous Systems Laboratory of the Institute for Information Technology of the National Research Council of Canada, and the Department of Manufacturing Engineering and Engineering Management of the City University of Hong Kong. Dr. Lang is a senior research officer with the National Research Council of Canada's Integrated Manufacturing Technologies Institute, located in London, Ontario, Canada, and group leader of the Reconfigurable Manufacturing Group which focuses on reconfigurable machines and reconfigurable assembly processes. His research interests include reconfigurable manufacturing systems, mobile robots, autonomous guided vehicles, mechatronic systems, vision and sensor systems, graph theoretic modeling of mechanisms, parallel kinematic mechanisms, sensor guided intelligent control, system design, mechatronics and manufacturing systems. Dr. Lang is a member of PEO, serves on the Board of Directors of the London District Science and Technology Fair, and served on the Board of Directors of the London Chinese Canadian National Council.