Non-destructive Automatic Leaf Area Measurements by Combining Stereo and Time-of-Flight Images

Yu Song, Chris A. Glasbey, Gerrit Polder, Gerie W.A.M. van der Heijden

Abstract Leaf area measurements are commonly obtained by destructive and laborious practice. This paper shows how stereo and Time-of-Flight (ToF) images can be combined for non-destructive automatic leaf area measurements. We focus on some challenging plant images captured in a greenhouse environment, and show that even the state-of-the-art stereo methods produce unsatisfactory results. By transforming depth information in a ToF image to a localised search range for dense stereo, a global optimisation strategy is adopted for producing smooth results that preserve discontinuity. We also use edges of colour and disparity images for automatic leaf detection and develop a smoothing method necessary for accurately estimating surface area. In addition to show that combining stereo and ToF images gives superior qualitative and quantitative results, 149 automatic measurements on leaf area using our system in a validation trial have a correlation of 0.97 with true values and the root-mean-square error is 10.97 cm^2, which is 9.3% of the average leaf area.. Our approach could potentially be applied for combining other modalities of images with large difference in image resolutions and camera positions.

Keywords Dense Stereo · Time-of-Flight · Leaf Detection · Surface Reconstruction · 3D Measurements

1 Introduction

In our post-genomic world, where we are deluged with genetic information, the bottleneck to scientific progress is often phenotyping, i.e. measuring the observable characteristics of plants and animals. For example, surface

Y. Song · C. A. Glasbey
Biomathematics and Statistics Scotland, The King's Buildings, Edinburgh, EH9 3JZ, UK
E-mail: {y.song,chris.glasbey}@bioss.ac.uk

G. Polder · G.W.A.M. van der Heijden
Biometris, Wageningen UR, PO Box 100, 6700 AC Wageningen, Netherlands
E-mail: {gerrit.polder,gerie.vanderheijden}@wur.nl

area of a leaf is a powerful diagnostic of plant productivity. Current common practice, also known as direct measurement, requires each plant leaf to be manually stripped and fed through a dedicated measuring machine, which is destructive, laborious and time-consuming [23]. Owing to these difficulties, leaf area measurements during the plant growth period as in [46] are often not possible and they have been empirically simulated in plant growth analysis [21,32].

Image analysis has the potential to overcome these problems. The aim is to develop an automated procedure for extracting each leaf individually from an image, followed by reconstruction and measurement of extracted leaves in 3D. Automatic interpretation of plant images remains very difficult. Fig. 1 shows a stereo pair of images of pepper plants captured using the setup shown in Fig. 2, using flash lighting both to cancel the variable effects of natural illumination and permit a fast shutter speed which minimises blur while the camera rig is in motion. Our aim is to recover dense depth information and estimate leaf area in 3D along with other plant characteristics such as stem length or fruit size. This is a challenging task, as surfaces are of complex shape, and there are multiple depths, linear features, occlusions and inconsistent shadows between images. Stereo vision is usually seemingly-effortless for the human eye and brain, but unfortunately still not so for computers.

Recently the use of low-resolution range cameras based on the Time-of-Flight (ToF) principle has received increasing attention, with Kolb *et. al.* [25] providing an overview on techniques and potential applications. The state-of-the-art stereo vision algorithms are known for underperforming in low-textured areas (usually at object centres) and preserving edge discontinuity, which can be complemented by the use of a ToF camera. Augmenting stereo pairs with ToF images can therefore be used for applications aimed of improving recovery of depth information. However, as in Fig. 1, the ToF camera has different field-of-view and position to the colour camera, and the difference in resolution between the colour image and the ToF image is enormous.

These images were collected as part of an EU-funded FP7 project, SPICY (Smart tools for Prediction and Improvement of Crop Yield). The plant breeding industry has contributed greatly to the increased quality and yield of plant products over recent decades. However, to sustain and accelerate this progress, the relationship between genotype and phenotype needs to be better understood. For example, yield is a result of the interaction of many genetic factors, and is also subject to large, extraneous variation. The approach taken in SPICY is to use crop growth models to predict the phenotypic response, with genotype encapsulated in model parameters

[2,20]. Our task in the project is the development of image analysis tools to replace hand measurements for phenotyping over a large range of genotypes in a practical environment.

Fig. 3 shows the main processes of our approach for leaf area estimation. The first step is to recover dense depth information from image pairs, since leaves appear in different sizes in an image at different depths. Without depth estimation, we can only calculate projected area of leaves as in [38]. Leaf detection enables extracting each individual leaf from images, and together with depth estimates, three-dimensional surface of each leaf can then be reconstructed and measured. In effect, we aim to identify individual leaves and then measure them, which is analogous to the manual measurement process.

In Fig. 1, the ratio of pixels between colour and ToF images is 200 : 1. With the development of ToF camera such as a more recent camera used in [1] [1] , image registration[7] in principle could become a possible solution but cluttered scenes as in Fig. 1 are difficult. In this paper, we used a rigid setup allowing projecting a partial and coarse-resolution ToF image into a corresponding colour image as an aid to stereo vision.

Part of our work has been described in [37,42,20]. [37] first described the imaging setup we built, and [42] described the general idea of our work. [20] discussed the practical imaging applications using our methods and how they can be used for quantifying plant characteristics. This paper presents details on camera calibration for combining stereo and ToF images, and a smoothing technique for surface reconstruction. We also describe a edge-based segmention method using the 3D geometry to extract individual leaves from images.

Our imaging setup and our approach in principle can be applied for combining many modalities of images with large difference in image resolutions and camera positions. For example, our imaging setup could have a thermal camera providing temperature information, and high dimensional data from multiple imaging modalities would lead to many practical applications.

2 Relevant work

Dense stereo produces a depth estimate for every pixel in an image, which is required in the first step of our system. One approach to dense stereo is via local descriptors such as SIFT [31], followed by methods such as DAISY [44] and SIFTflow [29], both of which deal with challenging stereo images. Discontinuity preserving results are highly desirable for our application, but no quantitative result addressing this issue was produced in

[1] In [1], the resolutions were 640×480 for colour images and 204×204 for ToF images, and their ratio of pixels is 7.38 : 1.

[29,44]. Though global optimisation methods such as graph cuts [5] can produce edge-preserving results on the Middlebury stereo dataset[39], challenges in the Middlebury stereo dataset are different to these images in our work. Ogale and Aloimonos [34] proposed to use shape in establishing edge-preserving dense correspondence, but surfaces in our images are more complicated than theirs.

Gudmundsson *et. al.* [17] transformed ToF points into colour images by rectification homographies, and then fed them into a hierarchical stereo matching algorithm. Hahne and Alexa [18] demonstrated the combined ToF and stereo method can enhance the depth estimation even without accurate extrinsic calibration. Zhu *et. al.* [47] developed a weighting method combining stereo and ToF data by fixed values, and then used belief propagation to optimise the data. Motivated by this research, we first present a geometric approach to transform points from ToF image coordinates to colour image coordinates, and then derive a localised search range for stereo matching. Despite the simplicity of the ToF transformation, we demonstrate that a global stereo strategy can then be applied and does improve results and preserve discontinuity. Compared with above works [17,18,47], difficult low-resolution ToF images 48×64 were used in this work. Beder *et. al.* [4] also developed a fusion scheme using ToF images in the same resolution as ours except that their images were planar surfaces.

Current ToF and stereo fusion work (e.g. [4,17,18,27]) lack quantitative results on preserving depth discontinuity, with the exception of [47], which used another 3D scanner to produce pixel-by-pixel depth data in an indoor lab environment. In our greenhouse setting it is problemmatic to collect accurate pixel-by-pixel depth data, so we use an indirect method to evaluate the quality of depth estimation for competing methods, by quantifying how well depth-discontinuity is preserved.

Foreground extraction of live video using ToF and colour cameras is proposed by Wang *et. al.* [45] in order to segment a person in foreground from the background. Their challenges are to track and segment the foreground person from a continuous video sequence, while our images have very limited views of foreground objects. Similar applications in this area have also been investigated in [8,16,40].

Regarding leaf measurement, existing work [35,38,43] focuses on collecting images of individual plants that are separately transported from the greenhouse to a controlled imaging environment, or on imaging single leaves against a plain backround [26]. Instead of moving individual plants around the greenhouse, our methodology brings the imaging equipment to the plants. Transporting growing plants is undesirable, because of potential damage to plants that can highly disturb their growth. But more importantly, for many greenhouse crops like pepper and tomato, the plants are simply too large to be transported. Our system measures plants in their own growing environment, and does not require transporting plants. However, by using our system, challenges arise from less controllable lighting conditions and a cluttered scene with large occlusions as shown in Fig. 1 and 2.

3 Setup and Calibration

Every pepper plant has a QR barcode for relating the manual measurements to the automatic measurements, and the plants grow in rows with heating pipes in-between. The maximum height of the plants is about three metres, while the space between rows is only one metre. Four camera rigs were therefore vertically stacked in a trolley known as Spy-See to cover the complete field of view, and [37, 20] provided hardware details.

Our camera rig consists of a colour camera and a ToF camera. The ToF camera is a radio-frequency modulated camera with phase shift detectors (IFM O3D201 PMD camera), with a resolution of 64×48 pixels, while the colour camera has a resolution of 480×1280. The Spy-See setup moves in a straight line on top of rigid heating pipes in the greenhouse and captures overlapping images at a fixed interval (see Fig. 2). The baseline between images is 5 cm, and objects of interest (e.g. leaves) are located between 55 cm and 120 cm away from the camera.

Once assembled, our imaging setup was rigid and fixed. The positions of the colour and ToF cameras were unchanged, so was the capture interval. We therefore performed depth calibration to find both cameras in 3D space relative to each other.

A two-layer board shown in Fig. 4(a) was used for calibration of the colour camera at different distances from the camera. The front layer moved from 40 cm to 120 cm away from the camera in 5 cm steps. We used a simple pinhole camera model for the colour camera [19] as shown in Fig. 4(b). Let (x, y) be the position of a point in colour camera coordinate, and $(x_i, y_i), i = 0, 1, 2 \cdots$ represents the position in image i. Given a point in the world coordinate (X, Y, Z), (x_i, y_i) can be obtained as,

$$x_i = (X - s\,i)\, f \,/\, Z + x_m \tag{1}$$

$$y_i = Y\, f \,/\, Z + y_m \tag{2}$$

where s is the baseline between the two images, which was 5 cm for our setup, and f is the focal point of the colour camera. We set the principal point (x_m, y_m) as the image centre for simplicity, which only produced negligible errors. We only consider horizontal disparity since our imaging setup moved in the horizontal direction only. Given a point identified in two colour images x_0 and x_1, from (1),

$$x_0 = X\,f\,/\,Z + x_m \tag{3}$$

$$x_1 = (X - s)\,f\,/\,Z + x_m \tag{4}$$

Let $d = x_0 - x_1$ be the disparity,

$$d = s\ f\,/\,Z \tag{5}$$

During calibration, multiple depth measurements $\mathbf{Z}$ (e.g. 40 cm to 120 cm in this work) and correspondences in each view $\mathbf{d}$ are used to compute $\hat{f}$ by applying the least squares fitting technique.

$$\hat{f} = \arg\min_{f} \parallel \mathbf{d} - s\,f\,/\,\mathbf{Z} \parallel^2 \tag{6}$$

The centre of the square seen in Fig. 4(a) is used to compute $\mathbf{d}$, and $\mathbf{Z}$ is known for each image. The squares can be identified by linear discriminant analysis and connected-component labelling [13]. We manually labelled squares in two board images as training data, and linear discriminant analysis was subsequently used to segment squares for all the other board images. Since the centre of the square was required, morphological dilation was applied for refining the square shape before connected-component labelling. Fig. 4(c) presents the relationship in (5) and plots $\mathbf{d}$ against $\mathbf{Z}$. Compared to a flat checkerboard used in [28], the two-layer board is also used for the ToF camera, to convert ToF measurements (x', y', z') to world coordinate (X, Y, Z).

Since the relative positions of colour and ToF cameras did not change, a transformation can be established for points in colour image and ToF image. A ToF image maps Z on z' and a near-linear relationship between ToF depth measurements z' and Z was observed as in [47]. As in (1) (2), the focal point f' is still needed for the transformation of $X \rightarrow x'$ and $Y \rightarrow y'$.

$$x'_i = (X - X_0 - s\,i)\, f' \,/\, Z + x'_m \tag{7}$$

$$y'_i = (Y - Y_0)\, f' \,/\, Z + y'_m \tag{8}$$

X_0 and Y_0 are the physical distance in cm between the colour camera and the ToF camera. Although the relationship between the colour camera and the ToF camera is translational in this work, [19] provided information on the homogeneous affine transformation between two cameras.

During ToF camera calibration, we used known Z as a cue to perform thresholding for identifying the centre of each square, and then compute d' using two adjacent ToF images. The same procedure for $\hat{f}$ as in (6) was used to obtain $\hat{f}'$ for the ToF camera. Challenges and error sources in ToF camera calibration have been discussed in [25, 28] ([28] also provided calibration software), and readers can review these works for further information.

4 Combining stereo and Time-of-Flight images

4.1 Dense stereo methods

Dense stereo methods can estimate disparity d for every pixel given a pair of stereo images. However, the pixel consistency assumption is often made for building the correspondence between two images. In our application, we have found that pixel values were not reliable for matching due to changes of perspective, lighting, and noise. To address this issue, the SIFTflow [29] method was chosen, which uses pixel-wise SIFT features between two images instead of pixel values for matching. Complex image pairs across different scenes and object appearances have been shown robustly matched in [29].

The pepper plant images shown in Fig. 1 have very sharp depth edges, and we have observed step changes over 50 pixels between neighbourhood pixels. Although Liu *et. al.* used a simple synthetic image in [29] to demonstrate that the dense SIFT features contain sharp edges with respect to the sharp edges in the original image, there is no close-up on complex scenes to prove that the SIFTflow method can preserve discontinuity. Ogale and Aloimonos [34] examined the implications of shape on the process of finding dense correspondence, and attempted to produce disparities in the form of a piecewise continuous function consistent with the stereo images. Using piecewise constant and piecewise linear shape models, they have presented results on images with slanted planar surfaces as well as a pair of stereo images on some branches of a tree, but no results on curved or nonrigid surfaces common in the pepper plant images have been shown.

Global optimisation methods such as graph cuts and belief propagation have been shown producing satisfactory discontinuity-preserving results on the Middlebury stereo dataset [39]. Since global stereo methods produce better results compared with local stereo methods for combining with ToF information [47], we chose the alpha expansion technique applied in a graph-based energy minimisation framework [5]. The energy cost E given a pixel disparity d is defined as:

$$E(d) = \sum D(d_{(x,y)}) + \sum_{q \in N} V(d_{(x,y)}, d_{(x_q,y_q)}) \tag{9}$$

where N denotes the first-order neighbourhood pixels. For the data term cost D,

$$D(d_{(x,y)}) = \min \left\{ \frac{1}{3} \sum_{c=\{R,G,B\}} \left| I^{(c)}_{(x,y)} - I^{'(c)}_{(x+d_{(x,y)},y)} \right|, T_d \right\} \tag{10}$$

where I and I' represent the intensity value in the pair of colour images. T_d is a truncation constant, and $D(d_{(x,y)})$ is computed for all the possible disparities. For the smoothness term cost V,

$$V(d_{(x,y)}, d_{(x_q,y_q)}) = u_{(x,y,x_q,y_q)} \min \left\{ \left| d_{(x,y)} - d_{(x_q,y_q)} \right|, T_k \right\} \tag{11}$$

where parameter T_k is used to truncate the linear energy. (x_q, y_q) is one of the first-order 4-neighbourhood pixels around (x, y). $u_{(x,y,x_q,y_q)}$ represents static cues in Boykov *et. al.* [5], which was used as an indicator function in this work as:

$$u_{(x,y,x_q,y_q)} = \begin{cases} \alpha_v & \text{if } \sum_{c=\{R,G,B\}} \left| I^{(c)}_{(x,y)} - I^{(c)}_{(x_q,y_q)} \right| > T_e \\ n_v \, \alpha_v & \text{otherwise} \end{cases} \tag{12}$$

α_v is the smoothness cost for intensity edges produced by the thresholding value T_e, and $n_v \, \alpha_v$ is the smoothness cost for surfaces. Both α_v and $n_v \, \alpha_v$ should be set according to the data cost values in (10). (12) gives more smoothness if there is no intensity edge, and therefore achieves edge-preservation by encouraging changes at edges at a cost of α_v and limiting changes on the surface by $n_v \, \alpha_v$.

4.2 Localised search range from ToF image

Given the complexity associated with the pepper plant images for dense stereo methods, a localised search range $[d_{min}, d_{max}]$ derived from the corresponding ToF depth image should improve the estimation accuracy.

Since the ToF image is much coarser in resolution compared to the colour image, the transformation from ToF image coordinates to colour image coordinates alone would only give isolated point depth measurements in the colour image. We therefore treat each ToF pixel as a patch centring around the pixel, and then transform all points in the patch to the colour image (see ToF in Fig. 9 for an example). In effect, this transformation is one of the up-scaling techniques as discussed by Lindner *et. al.* [27] and they provided a biquadratic scheme for this purpose.

Due to different viewing positions of ToF and RGB cameras, there are n ToF measurements for Z $(n \geq 0)$ at location (x, y). If multiple depths were found at (x, y), the minimum value would be chosen, which represents the closest point to the camera. If no measurement of Z is available for (x, y), this would be treated as a missing value. To produce a localised search range $[d_{min}, d_{max}]$ for stereo matching, we used a patch centring around every pixel in the colour image to compute the minimum and maximum depth values. Denote (x, y, Z) as $Z_{(x,y)}$ and the patch as $Z_{(\mathbf{m},\mathbf{n})}$,

$$|\mathbf{m} - x| \leq r, \quad |\mathbf{n} - y| \leq r \tag{13}$$

In effect, this allows mis-alignment up to r pixels when transforming the ToF image to the colour image. The maximum and minimum depths are then converted into disparities as,

$$d_{min(x,y)} = s\,f/\max\{Z_{(\mathbf{m},\mathbf{n})}\} - k \tag{14}$$

$$d_{max(x,y)} = s\,f/\min\{Z_{(\mathbf{m},\mathbf{n})}\} + k \tag{15}$$

The search range is expanded by k pixels (normally $0 \leq k \leq 3$) at each direction to allow for the noise in the ToF estimates. Given a localised search range $[d_{min}, d_{max}]$ for every pixel, a stereo method can then be used to find correspondences between images. In this work, for the data term cost D in (10), if $d_{(x,y)}$ is outside the search range $[d_{min}, d_{max}]$ or $d_{(x,y)}$ is linked to a pixel outside the image, $D(d_{(x,y)})$ is set to the maximum pixel difference value T_d. If the localised search range $[d_{min}, d_{max}]$ is missing, $D(d_{(x,y)})$ is computed for all the possible disparities same as a dense stereo method.

It should also be noted that the search range is small at the object centre due to small depth variation and ToF measurements contribute more to the results, while the opposite can be observed at the depth discontinuities.

4.3 Quality score

Since pixel-by-pixel depth data are unavailable as ground truth for our images and many other applications, we propose a quantitative method accounting for the surface smoothness and the edge sharpness to evaluate estimation results. Leaf has depth edges along its boundary as seen in Fig. 5 and we can label depth edges to quantify how well the result has preserved them. Leaf boundaries shown in Fig. 5 were obtained manually. We only performed this manual edge labelling at this evaluation stage to produce ground truth for depth edges, and neither the ToF transformation nor the stereo method required any intervention after calibration. The area within the leaf boundaries is considered a leaf surface, and the final output consists of two binary images, a surface and an edge image. We applied the Canny edge filter [13] with non-maximum suppression to compute the smoothness of the surface and sharpness of the depth edges. Surface smoothness penalty P_s was calculated for the surface image as,

$$P_s = \overline{M(d)_{(x_s,y_s)}} \tag{16}$$

where M represents the Canny edge filter using a Gaussian that has a standard deviation of 1 and a radius of 1.5 for non-maximum suppression. (x_s, y_s) are surface pixels. Edge sharpness score S_e was calculated for the edge image as,

$$S_e = \overline{(g * M(d))_{(x_e,y_e)}} \tag{17}$$

where $*$ is the 2-dimensional convolution operation [13] and g denotes a Gaussian filter in order to deal with thin and sharp depth edges. In this work, we set the neighbourhood size of the Gaussian filter to 15 and the

Table 1 Steps for automatically detecting frontal leaves.

1. Select the disparity plane nearest to the camera $\max\{d\}$.
2. Select a point (x_c, y_c) from $\max\{d\}$ that has the minimum value of combined edge magnitude $M(I, d, \gamma)$.
3. Use (x_c, y_c) as the seed point, perform region growing method to segment a leaf (x_l, y_l).
4. Set $d_{(x_l,y_l)}$ to the furthest to the camera $\min\{d\}$.
5. Repeat 1 - 4 for the next leaf.

standard deviation to 5. The effects of (16) and (17) can be seen in Fig. 5. A quality score S accounting for the surface smoothness P_s and the edge sharpness S_e was computed as below,

$$S = S_e - P_s \tag{18}$$

The score S penalises displacement between defined depth edges and depth edges by a dense method while requiring the surface to be smooth. S is a relative score that becomes meaningful when comparing two dense methods. Given S calculated for two methods, the stereo result with the lower S score has more blurry depth edges (smaller S_e), more noise on surface (larger P_s), or both. Consequently, for our application in this paper, a dense method with a higher S score is preferred over one with a lower S score.

It should be noted that Sobel edge magnitude can also be used for M to calculate P_s and S_e as in [42]. However, automatic leaf detection described in section 5 uses the same Canny edge filter.

4.4 Parameter tuning

One way of using the quality score in (18) is tuning parameters. Let θ be a set of parameters that is required by a method (e.g. our method combining stereo and ToF or other stereo methods), and the sum of score $\sum S$ on some objects with manually labelled depth edges is used as a quality measure. In this work, we manually labelled three leaves in calibration images. The set of parameters θ that maximises $\sum S$ is considered as the tuned parameter set. Since θ has one or more parameters, we sequentially optimise all possible values in each parameter leading to a local optimisation $\hat{\theta}$. Several iterations defined by the number of parameters in θ are then used to refine $\hat{\theta}$, and the order of parameters is randomised in each iteration.

5 Leaf Detection

Given a colour image I and its corresponding disparity image d, this section describes a general approach for automatically extracting leaf boundaries from images. Let M represent a filter that produces edge magnitude, and $M(I)$ and $M(d)$ are the edge magnitude on the colour image I and the disparity image d respectively. It is possible to use either $M(I)$ or $M(d)$ alone for boundary detection. However, both $M(I)$ and $M(d)$ are complicated and noisy as shown in Fig. 6, which is hard to produce reasonable boundary detection. A combination of $M(I)$ and $M(d)$ can simplify the problem, which is obtained as follows,

$$M(I, d, \gamma) = \gamma\, M(I) + (1 - \gamma)\, M(d) \tag{19}$$

where γ is a weighting coefficient and $0 \leq \gamma \leq 1$. The idea is based on the fact that object boundaries could be enhanced by blending edges existing in the colour and disparity images, despite one of them could be weak in edge magnitude.

We assumed that some leaves are in foreground closer to the camera than other objects like stems, since leaves have to reach out for maximising light interception. Based on this assumption and the combined edge magnitude $M(I, d, \gamma)$, an automated procedure was then developed to detect the nearest frontal leaf. A summary of steps in the procedure is shown in Table 1. The second step attempts to select a point on a leaf surface as the seed point, which should have a small value of combined edge magnitude $M(I, d, \gamma)$. We used a region growing method [15] on edge magnitude produced by the filter M to extract regions, since the method is well-understood and performs well with respect to noise. The criterion we used was to compare edge magnitude of the adjacent pixels near the region borders to the region's mean value. A thresholding parameter T_c acting as a similarity threshold value was used to determine the terminal condition. There are also a number of other alternatives for foreground extraction using ToF and colour cameras [45,8,16,40].

In this paper, we have found that the Canny edge filter with non-maximum suppression described in section 4.3 can be used for the filter M. The Canny edge filter was configured to use a Gaussian that has a standard deviation of 1 and a radius of 1.5 for non-maximum suppression. For calculating edge magnitude on colour image $M(I)$, I is transformed from RGB to CIELAB colour space, and the edge magnitude by the Canny filter is separately calculated for each component and is subsequently summed up. There are other choices of M that can be used (e.g. [3]), and more methods can be found in the Berkeley segmentation benchmark [33].

Fig. 6(e) shows the output of (19) using the Canny filter, and $M(I, d, \gamma)$ emphasises edges found in both $M(I)$ and $M(d)$ that corresponds to the boundary of our interest. Fig. 7 shows automatic leaf detection using our method.

6 Surface Reconstruction

Given an identified leaf (x_l, y_l) in an image with colour and disparity data d, it is possible to reconstruct and measure the leaf shape in 3D by a triangular mesh representation. Vertices in the mesh correspond to pixels in the image which is $(x_l, y_l, d_{(x_l,y_l)})$, and the edges in the mesh are built by connecting nearest neighbouring vertices. The use of a triangular mesh representation allows calculating the surface area in 3D analogous to manual measurement. $(x_l, y_l, d_{(x_l,y_l)})$ is transformed to the world coordinate (X_l, Y_l, Z_l) as shown in section 3, and areas of all triangles are then summed up.

An immediate problem of the triangular mesh representation is the 'rice terrace' effect as shown in Fig. 8(a). This aliasing effect was caused by discretisation in the depth estimates at such a small scale. In this work, baseline s is small for stereo matching and disparity estimates are integers, which cannot cope with the demand for an accurate reconstruction. In addition to the inaccurate visual reconstruction, the 'rice terrace' effect would over-estimate the surface area that is not desirable.

One approach is to develop sub-pixel accuracy stereo as described by [39], but in this work we propose a method to smooth the depth data Z_l at the reconstruction stage after extracting leaves by combining stereo and ToF. We decided to use local regression (LOESS) [6, 30], and it is based on the idea that any function can be well approximated in a small neighbourhood by a low-order polynomial and that simple models can easily be fit to data. A linear LOESS model was used in this work, and it requires a specific smoothing parameter β that is a percentage of the total number of data points. The effect of LOESS smoothing increases with increasing β as shown in Fig. 8. Fig. 8(d) and 8(e) also show the histogram of the residual image to illustrate the effect of smoothing.

β can be defined empirically or by Generalised Cross Validation (GCV) [14]. However, we prefer a smooth surface which would have larger residual on the flat planes of the 'rice terraces' rather than no or small residual, and GCV is not built for correlated errors. An ad-hoc procedure to determine the smoothing parameter β is therefore developed. Since we would like to eliminate the aliasing effects, the residuals would not have a pronounced peak at 0 in histogram as in Fig. 8(d). Denote the histogram count as h_c with bin width b. The number of residuals at 0 is $h_c(0)$, and its two adjacent bins of the histogram are $h_c(-b)$ and $h_c(b)$ respectively. The stopping criterion for increasing the smoothing parameter is,

$$\min\left\{h_{(0)} - h_{(-b)}, h_{(0)} - h_{(b)}\right\}\frac{1}{h_{(0)}} < T_h \tag{20}$$

where T_h is a parameter in percentage that controls the difference of histogram counts between the residual at 0 and its two adjacent bins.

7 Results

7.1 Depth estimation

We compare three dense stereo algorithms with our method on some challenging pepper plant images. Images used for parameter tuning are calibration images, and the others are used as validation images. Once parameter tuning has been performed on calibration images as in section 4.4, none of the methods require intervention in the validation step.

We present results on three sets of images capturing well-developed plants (Plant 1 - 3). Plant 1 acts as the calibration image, and Plant 2 and 3 are validation images. Let SIFTflow, Shape and GC represent methods by Liu *et. al.* [29], Ogale and Aloimonos [34] and Boykov *et. al.* [5] respectively. GC refers to the graph cut method without using ToF, and GC+ToF is our method. By parameter tuning described in section 4.4, SIFTflow was configured with a 5-level pyramid, 5×5 window, $\alpha = 1$ and $\gamma = 0.001$. The α in the Shape method was set to 2. Parameters $T_e, T_d, T_k, \alpha_v, n_v$ for GC were set as $25, 20, 6, 4, 4$. For GC+ToF, the same parameters for GC were used for dense stereo and ToF parameters r and k were set to 10 and 1 respectively. Since these methods are established, readers can see the effects of these parameters by following [29, 34, 5] for SIFTflow, Shape and GC respectively.

Fig. 9 shows qualitative stereo results produced by the four methods on Plant 1. Methods GC and GC+ToF produced results with leaves recognisable from the background. SIFTflow produced smooth results but did not

Table 2 Numerical summary of quality evaluation for Leaf 1, Leaf 2 and Leaf 3. S_e refers to edge sharpness, P_s refers to surface smoothness and S is the quality score. For our application in this paper, we would like to find a method giving the highest S score.

	Leaf 1			Leaf 2			Leaf 3		
	S_e	P_s	S	S_e	P_s	S	S_e	P_s	S
SIFTflow	0.22	0.10	0.12	0.11	0.01	0.10	0.04	0.02	0.02
Shape	0.70	0.44	0.26	0.74	0.20	0.54	0.27	0.10	0.18
GC	0.49	0.20	0.29	0.87	0.04	0.83	0.20	0.04	0.16
GC+ToF	1.34	0.18	1.15	1.31	0.05	1.26	0.36	0.06	0.30

Table 3 Quantitative summary of quality scores for both calibration and validation images. Figures shown here are total quality scores for three leaves in a image, $\sum S$.

	Calibration	Validation	
	Plant 1	Plant 2	Plant 3
SIFTflow	0.24	0.18	0.36
Shape	0.98	1.32	0.45
GC	1.28	1.29	0.51
GC+ToF	2.71	2.04	0.63

preserve discontinuity, while Shape showed the opposite effect. This can be further examined in Fig. 10, which shows a closer view of three leaves. The edge was weak for SIFTflow, although the surface was the most smooth. Method Shape suffered from noises on the surface, and GC failed to produce some depth edges. In comparison, GC+ToF produced the best qualitative results among the four methods.

A summary of quantitative results (S_e, P_s, S) for all three leaves is shown in Table 2. Similar to the findings in the qualitative results above, we see that GC+ToF produced sharp depth edges represented by a high S_e score especially for Leaf 1 and Leaf 2. The ranking of methods produced by the score S is also consistent with the qualitative results for the two leaves. Leaf 3 is in front of another leaf, and the magnitude of depth edges is therefore not as strong as those in Leaf 1 and Leaf 2. GC+ToF produced the best scores S_e and S among the four methods.

This section has shown results on the calibration image (Plant 1) to illustrate the behaviour of the four methods. Furthermore, two sets of results on validation images (Plant 2 - 3) have been produced. Table 3 presents $\sum S$ for Plant 1 - 3. By using ToF as a localised search range, the estimation results were improved by at least 23% measured by the score $\sum S$.

It should be noted that the leaf boundary was manually selected here for comparison purposes, since automatic leaf detection can be difficult for estimates in Fig. 9. The next section will present the results of automatic leaf detection using our method.

7.2 Leaf detection and area measurements

A validation trial using 44 experimental plants (11 plots of 11 genotypes, each plot also has four border plants) was carried out to provide a set of validation images. The 44 experimental plants grew in a standard double-row arrangement, with 22 of them visible from each side of the row by our setup. The validation images were collected first, and four leaves in the foreground of images were then manually detached from every genotype giving 88 leaves in total. The positions of these leaves were annotated on the validation images in order to relate to manual measurements as the ground truth. There were over 600 colour images to annotate, and these 88 leaves were identified in 244 images as 248 separate measurements. We ignored these leaves spanning across two images (i.e. partial view of a complete leaf). Finally, the manual measurements of these 88 leaves were obtained by removing each leaf from its plant and scanning using an industry-precision LI-COR 3100 leaf area meter.

Calibration images for parameter tuning were collected in a different trial. The parameters determined in section 7.1 were used for our GC+ToF method. Based on qualitative results of calibration images, parameters T_c and γ in leaf detection were set to 0.5 and 0.4, and three leaves were automatically detected in every calibration image. Parameters b and T_h in surface reconstruction were empirically set to 0.2 and 0.1.

Our methods successfully extracted three leaves in each image producing 732 separate measurements from 244 images, but only 149 separate measurements on 59 leaves could be linked with the ground truth. This was because the ground truth data were created first, and our methods were fully automated for analysing validation images without accessing any manual annotation. Fig. 11 presents 15 examples randomly selected from the 149 boundary results.

Fig. 12 and Table 4 show the validation results of the 149 automatic measurements against 59 manual measurements with average area of 102.0cm^2. If no smoothing was applied, a lower correlation score and a considerable larger RMSE value were obtained, which is due to the ‘rice terrace’ effect as expected. Using our proposed smoothing method, the correlation between automatic and manual measurements is 0.97 and the RMSE value is 10.97 cm^2.

By averaging the 149 measurements for 59 leaves from different views, the correlation score has increased to 0.98, and the RMSE value is reduced to 9.50 cm^2, i.e. 9.3% of the average leaf area. These estimates have

Table 4 Correlation and Root-Mean-Square-Error (RMSE) results of 149 automatic leaf area measurements in validation data. Pearson correlation coefficient is used in this paper.

	Correlation	RMSE (cm^2)
No Smoothing	0.72	129.45
With Smoothing	0.97	10.97

been found to be of sufficient accuracy for plant breeding by identifying QTLs: positions on chromosomes which correlate significantly with measurements [20].

8 Discussion

Current practice is to measure leaf areas manually by destroying plants, which is also very costly in human time. Our method is non-destructive, and took about 3 minutes to record all images in a single row of plants. Then, average CPU times were 61sec/image for depth estimation, 1sec/image for leaf detection and 44sec/image for leaf area estimation. However, as images were processed off-line this is not critical, and could be speeded up by using more sophisticated or approximate algorithms or parallel processing.

The proposed 3D approach allows automatic measurement of the sizes of pepper leaves in a greenhouse. The setup could be extended to measure leaf sizes of other greenhouse crops such as cucumber or tomato, and in other situations, such as pot plants on a conveyer belt system. Similar approaches could be developed to measure size, orientation and shape of other components of plants, e.g. flowers, fruits, stems, and to exploit other multisensor systems [10,12,36].

Using stereo vision alone to extract individual leaves in scenes of plant structures with many leaves is clearly not sufficient. For occlusions and areas affected by unpredictable illumination, the data term in a global stereo framework (i.e. D in (10)) produces inaccurate energy costs since corresponding pixels are either unavailable or difficult to match. The minimised energy therefore does not represent the desirable results. Using ToF in these situations provides an estimate and reduces ambiguities. Another advantage is that dense stereo can be a super resolution technique for ToF images as discussed by [11,22,41]. Even using the latest ToF camera, the resolution of a ToF camera (320×240) is low compared with a colour camera, and this leads to errors at depth discontinuities. On the other hand, stereo vision is known for underperforming in low-textured areas usually at the object centres, but it can preserve discontinuity as we have presented in this paper (e.g. Fig. 10). Since stereo and ToF complement each other, our methods to combine them can therefore be used for applications aiming to improve the accuracy of measurement.

For our application, edge-preserving disparity result is very important. Since there are many leaves in an image (Fig. 9) and they have the similar appearance as well as large intra-class variation in visual images, it is impossible to extract one leaf from the rest without combining disparity estimation into the colour image (see Fig. 11). As shown in Fig. 11 and 12, our methods were fully automated for the validation images without any manual input (e.g. annotation), and were therefore proven to be effective for extracting reasonable foreground boundaries.

Our current assumption that leaves are in the foreground limits our methods to find only frontal leaves. We successfully collected 732 separate leaf measurements for 11 different genotypes of plants, and [20] presented further results using our methods for 151 genotypes. This was the first viable approach to collect a large number of leaf measurements in a non-destructive manner. In our work, foreground objects near to the cameras were all leaves. In a more general framework, a detection verification procedure, using classification methods such as support vector machine [9], could be developed for detecting target objects that are not in foreground.

Fig. 12 and Table 4 demonstrate the consequence of the 'rice terrace' effect, and highlight the importance of surface smoothing when calculating surface area. Although we presented our method for one pair of stereo images and one ToF image, it is in principle rather straightforward to apply it to multiple colour images and one ToF image, or even to multiple colour and ToF images. Both the correlation score and the RMSE value can be improved by averaging over multiple views, and the use of two or more views should be adopted in practice to reduce occlusions. We hope to build on the work in this paper for combining multiple colour and ToF images, and Kim *et. al.* [24] have shown some promising results on this subject.

9 Conclusions

This paper has presented an automatic approach for non-destructively measuring leaf surface area. Agreement with the ground truth of manually measured leaf areas, shown in Fig. 12, is good, and sufficient for plant breeding purposes [20]. Unlike most existing methods requiring individual plants to be transported to a controlled imaging environment, our work collects measurements from plants in their own growing environment. Three frontal leaves in an image (twelve from four images) were automatically measured, and they produced an estimate of the average leaf area that can be used in plant growth analysis[20].

We have demonstrated that combining stereo and ToF images leads to discontinuity-preserving results, which enable algorithms like Canny filter and region growing segmentation to extract individual leaves from a cluttered scene. We have also highlighted the importance of surface smoothing for calculating surface area, and proposed an adaptive way to choose the smoothing parameter.

Our approach has produced promising results on 149 automatic leaf area measurements in the validation data, which had a correlation score of 0.97 and a RMSE value of 10.97 cm^2 against the manual measurements. By using multiple views for the 59 leaves, the RMSE value reduced to 9.50 cm^2, with a correlation of 0.98.

The idea of combining stereo and ToF images has been proven useful for 3D measurements, and our approach could potentially be applied for combining other modalities of images with large difference in image resolutions and camera positions. Moreover, our automated approach is a major step forward in relation to the current destructive and laborious practice for measuring leaf area.

10 Acknowledgements

This work is part of the Smart tools for Prediction and Improvement of Crop Yield (SPICY) project supported by the European Community and funded by the KBBE FP7 programme. (Grant agreement number KBBE-2008-211347) We also acknowledge Scottish Government funding.

References

1. Alenya, G., Dellen, B., Torras, C.: 3d modelling of leaves from color and tof data for robotized plant measuring. In: International Conference on Robotics and Automation, ICRA 2011, pp. 3408–3414 (2011)
2. Alimi, N., Bink, M., Dieleman, J., Nicola, M., Wubs, M., Heuvelink, E., Magan, J., Voorrips, R., Jansen, J., Rodrigues, P., Heijden, G., Vercauteren, A., Vuylsteke, M., Song, Y., Glasbey, C., Barocsi, A., Lefebvre, V., Palloix, A., Eeuwijk, F.: Genetic and QTL analyses of yield and a set of physiological traits in pepper. Euphytica **190**, 181–201 (2013)
3. Arbelaez, P., Maire, M., Fowlkes, C., Malik, J.: Contour detection and hierarchical image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **33**, 898–916 (2011)
4. Beder, C., Bartczak, B., Koch, R.: A combined approach for estimating patchlets from pmd depth images and stereo intensity images. In: Proceedings of the 29th DAGM conference on Pattern recognition, pp. 11–20. Springer-Verlag (2007)

5. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. IEEE Trans. Pattern Anal. Mach. Intell. **23**, 1222–1239 (2001)
6. Cleveland, W.S., Loader, C.L.: Smoothing by local regression: Principles and methods. In: W. Haerdle, M.G. Schimek (eds.) Statistical Theory and Computational Aspects of Smoothing, pp. 10–49. Springer, New York (1996)
7. Collignon, A., Maes, F., Delaere, D., Vandermeulen, D., Suetens, P., Marchal, G.: Automated multi-modality image registration based on information theory. Information Processing in Medical Imaging pp. 263–274 (1995)
8. Crabb, R., Tracey, C., Puranik, A., Davis, J.: Real-time foreground segmentation via range and color imaging. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pp. 1–5 (2008)
9. Cristianini, N., Shawe-Taylor, J.: An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge University Press (1999)
10. Dhondt, S., Wuyts, N., Inze, D.: Cell to whole-plant phenotyping: the best is yet to come. Trends in Plant Science (in press)
11. Diebel, J., Thrun, S.: An application of markov random fields to range sensing. In: Proceedings of Conference on Neural Information Processing Systems (NIPS), pp. 291–298. MIT Press, Cambridge, MA (2005)
12. Fiorani, F., Schurr, U.: Future scenarios for plant phenotyping. Plant Biology **64**, 267–291 (2013)
13. Glasbey, C.A., Horgan, G.W.: Image analysis for the biological sciences. John Wiley & Sons, Inc., New York, NY, USA (1995)
14. Golub, G.H., Heath, M., Wahba, G.: Generalized cross-validation as a method for choosing a good ridge parameter. Technometrics **21**(2), pp. 215–223 (1979)
15. Gonzalez, R.C., Woods, R.E.: Digital Image Processing, 3rd edn. Prentice Hall (2008)
16. Gudmundsson, S.A., , Pardis, M., Casas, J.R., Sveinsson, J.R., Aanæs, H., Larsen, R.: Improved 3d reconstruction in smart-room environments using tof imaging. Comput. Vis. Image Underst. **114**, 1376–1384 (2010)
17. Gudmundsson, S.A., Aanaes, H., Larsen, R.: Fusion of stereo vision and time-of-flight imaging for improved 3d estimation. International Journal of Intelligent Systems Technologies and Applications **5**(3/4), 425–433 (2008)
18. Hahne, U., Alexa, M.: Combining time-of-flight depth and stereo images without accurate extrinsic calibration. International Journal of Intelligent Systems Technologies and Applications **5**(3/4), 325–333 (2008)
19. Hartley, R.I., Zisserman, A.: Multiple View Geometry in Computer Vision, second edn. Cambridge University Press (2004)
20. van der Heijden, G., Song, Y., Horgan, G., Polder, G., Dieleman, A., Bink, M., Palloix, A., van Eeuwijk, F., Glasbey, C.: SPICY: towards automated phenotyping of large pepper plants in the greenhouse. Functional Plant Biology **39**(11), 870–877 (2012)
21. Heuvelink, E.: Evaluation of a dynamic simulation model for tomato crop growth and development. Annals of Botany **83**, 413–422 (1998)
22. Huhle, B., Schairer, T., Jenke, P., Straíer, W.: Fusion of range and color images for denoising and resolution enhancement with a non-local filter. Comput. Vis. Image Underst. **114**, 1336–1345 (2010)
23. Jonckheere, I., Fleck, S., Nackaerts, K., Muys, B., Coppin, P., Weiss, M., Baret, F.: Review of methods for in situ leaf area index determination: Part i. theories, sensors and hemispherical photography. Agricultural and Forest Meteorology **121**(1-2), 19 – 35 (2004)

24. Kim, Y., Theobalt, C., Diebel, J., Kosecka, J., Micusik, B., Thrun, S.: Multi-view image and tof sensor fusion for dense 3d reconstruction. In: Proceedings of IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops, pp. 1542–1549 (2009)
25. Kolb, A., Barth, E., Koch, R., Larsen, R.: Time-of-Flight Sensors in Computer Graphics. In: M. Pauly, G. Greiner (eds.) Eurographics 2009 - State of the Art Reports, pp. 119–134. Eurographics (2009)
26. Kumar, N., Belhumeur, P.N., Biswas, A., Jacobs, D.W., Kress, W.J., Lopez, I.C., Soares, J.V.B.: Leafsnap: a computer vision system for automatic plant species identification. In: Computer Vision – ECCV 2012, pp 502–516. (2012)
27. Lindner, M., Lambers, M., Kolb, A.: Sub-pixel data fusion and edge-enhanced distance refinement for 2d/3d images. International Journal of Intelligent Systems Technologies and Applications **5**, 344–354 (2008)
28. Lindner, M., Schiller, I., Kolb, A., Koch, R.: Time-of-flight sensor calibration for accurate range sensing. Comput. Vis. Image Underst. **114**, 1318–1328 (2010)
29. Liu, C., Yuen, J., Torralba, A.: Sift flow: Dense correspondence across scenes and its applications. IEEE Trans. Pattern Anal. Mach. Intell. (2010)
30. Loader, C.: Local regression and likelihood (2001). http://cm.bell-labs.com/stat/project/locfit
31. Lowe, D.: Object recognition from local scale-invariant features. In: Proceedings of the Seventh IEEE International Conference on Computer Vision, vol. 2, pp. 1150–1157 (1999)
32. Marcelis, L., Heuvelink, E., Goudriaan, J.: Modelling biomass production and yield of horticultural crops: a review. Scientia Horticulturae **74**(1-2), 83 – 111 (1998)
33. Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Proc. 8th Int'l Conf. Computer Vision, vol. 2, pp. 416–423 (2001)
34. Ogale, A.S., Aloimonos, Y.: Shape and the stereo correspondence problem. Int. J. Comput. Vision **65**, 147–162 (2005)
35. Parsons, N., Edmondson, R., Song, Y.: Image analysis and statistical modelling for measurement and quality assessment of ornamental horticulture crops in glasshouses. Biosystems Engineering **104**(2), 161 – 168 (2009)
36. Pieruschka, R., Poorter, H.: Phenotyping plants: genes, phenes and machines. Functional Plant Biology **39**, 813–820 (2012)
37. Polder, G., van der Heijden, G.W.A.M., Glasbey, C.A., Song, Y., Dieleman, J.A.: Spy-See - Advanced vision system for phenotyping in greenhouses. In: Proceedings of the MINET Conference: Measurement, sensation and cognition, pp. 115–117. National Physical Laboratory (2009)
38. Rajendran, K., Tester, M., Roy, S.J.: Quantifying the three main components of salinity tolerance in cereals. Plant, Cell and Environment **32**(3), 237–249 (2009)
39. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Int. J. Comput. Vision **47**(1-3), 7–42 (2002)
40. Schiller, I., Koch, R.: Improved video segmentation by adaptive combination of depth keying and mixture-of-gaussians. In: The 17th Scandinavian Conference on Image Analysis, SCIA 2011, pp. 59–68. Ystad, Sweden (2011)

41. Schuon, S., Theobalt, C., Davis, J., Thrun, S.: Lidarboost: Depth superresolution for tof 3d shape scanning. In: Proceedings of the IEEE CVPR 2009, pp. 343 – 350 (2009)

42. Song, Y., Glasbey, C.A., van der Heijden, G.W.A.M., Polder, G., Dieleman, J.A.: Combining stereo and Time-of-Flight images with application to automatic plant phenotyping. In: The 17th Scandinavian Conference on Image Analysis, SCIA 2011, pp. 467–478. Ystad, Sweden (2011)

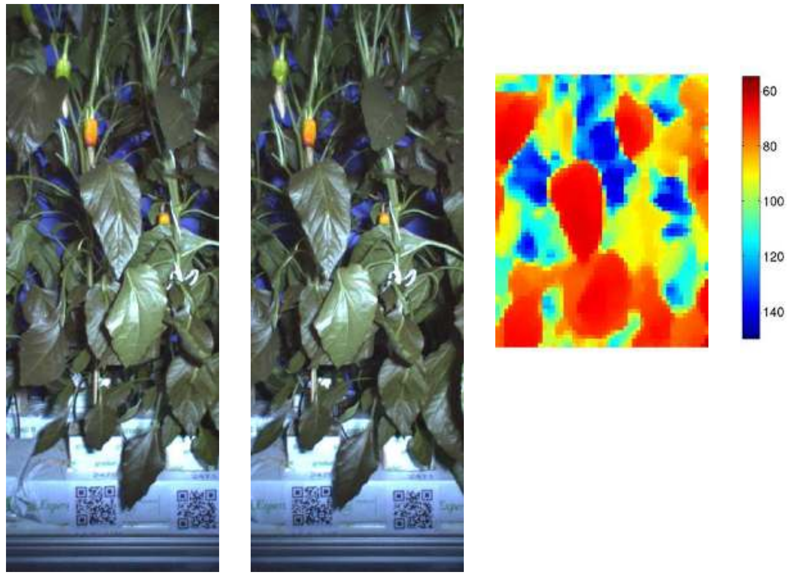
43. Song, Y., Wilson, R., Edmondson, R., Parsons, N.: Surface modelling of plants from stereo images. In: Proceedings of the 6th International Conference on 3-D Digital Imaging and Modeling. Montreal, Canada (2007)

44. Tola, E., Lepetit, V., Fua, P.: Daisy: an efficient dense descriptor applied to wide baseline stereo. IEEE Trans. Pattern Anal. Mach. Intell. **32**(5), 815–830 (2010)

45. Wang, L., Zhang, C., Yang, R., Zhang, C.: TofCut: Towards Robust Real-Time Foreground Extraction Using a Time-of-Flight Camera. In: The Fifth International Symposium on 3D Data Processing, Visualization and Transmission, 3DPVT 2010 (2010)

46. Wubs, A.M., Ma, Y., Hemerik, L., Heuvelink, E.: Fruit set and yield patterns in six capsicum cultivars. HortScience **44**, 1296–1301 (2009)

47. Zhu, J., Wang, L., Yang, R., Davis, J.: Fusion of time-of-flight depth and stereo for high accuracy depth maps. In: Proceedings of the IEEE CVPR 2008, pp. 1–8 (2008)



Colour (L) Colour (R) ToF

Fig. 1 Pepper plant images. Colour (L) and Colour (R) are a stereo pair of images of pepper plants, and Colour (R) is the base image. ToF is the depth image in cm matching the base image. The colour images are 480×1280 in size, while ToF image is only 64×48. The ratio of pixels between colour and ToF images is $200 : 1$.

(a) (b)

Fig. 2 Our system has four camera rigs vertically stacked to capture pepper plants, and each one has colour and ToF cameras and a flash light as seen in (a). Our system also moves around in a greenhouse that is exposed to unpredictable lighting including the sun and reflection as in (b).

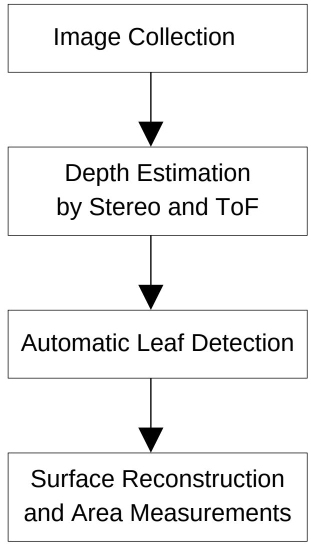


Fig. 3 Overview of main processes of our system for measuring leaf area.

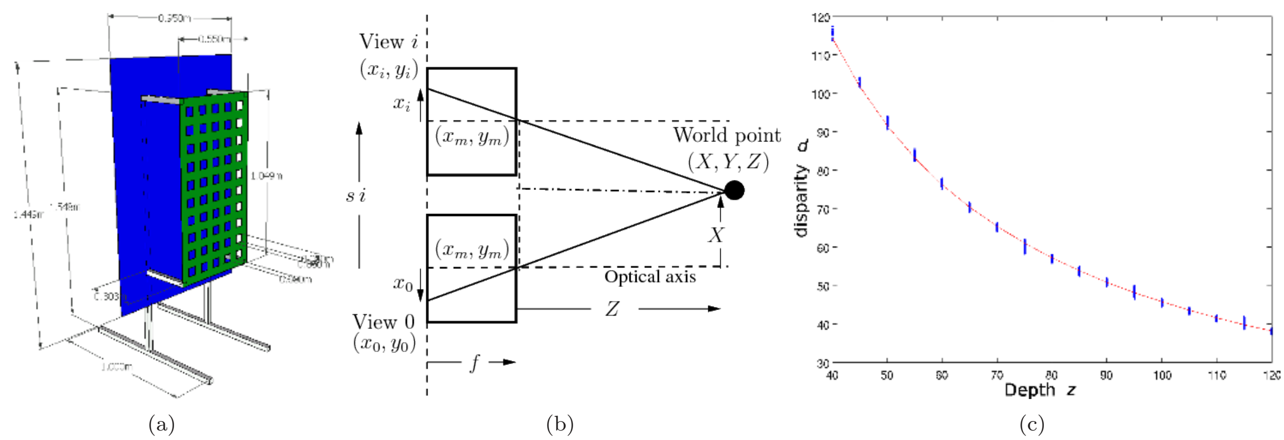


Fig. 4 Camera calibration: (a) diagram of calibration board; (b) diagram illustrating the transformation of a point between colour images and world coordinate. Dashed line represents the optical axis going through the principal point (x_m, y_m). y_i and Y represent axis perpendicular to the page with the same projective properties. (c) plot of the relationship between depth Z in cm and disparities d in pixels for colour camera. Blue dots were disparity measurements d for each Z, and the red line was the fit by (6).

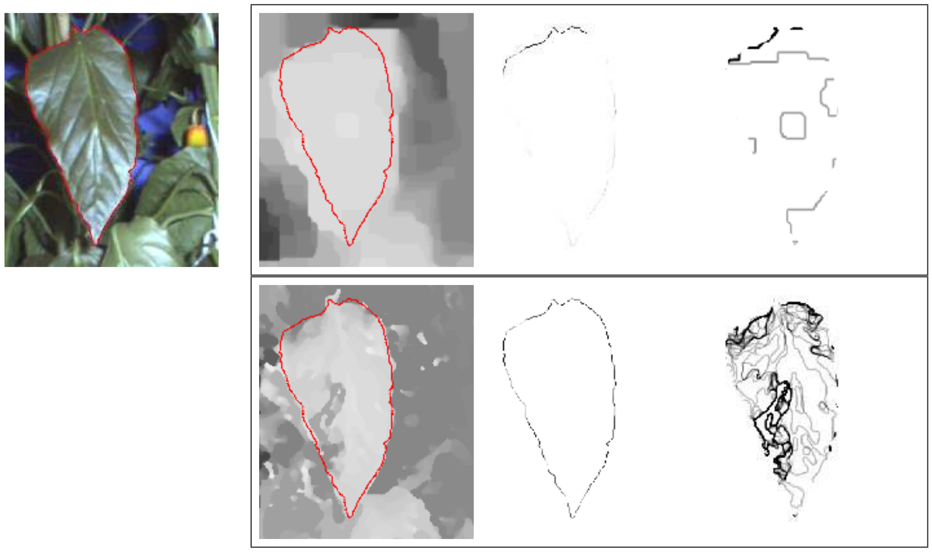


Fig. 5 Examples to illustrate quality scores on two disparity results. The colour image with depth edges plotted in red is shown on the left. The grey values in each panel represent the disparity (left), S_e (middle) and P_s (right) respectively. The disparity maps use a grey value scale of 20-90 pixels black-white. The S_e and P_s images also use a common scale.

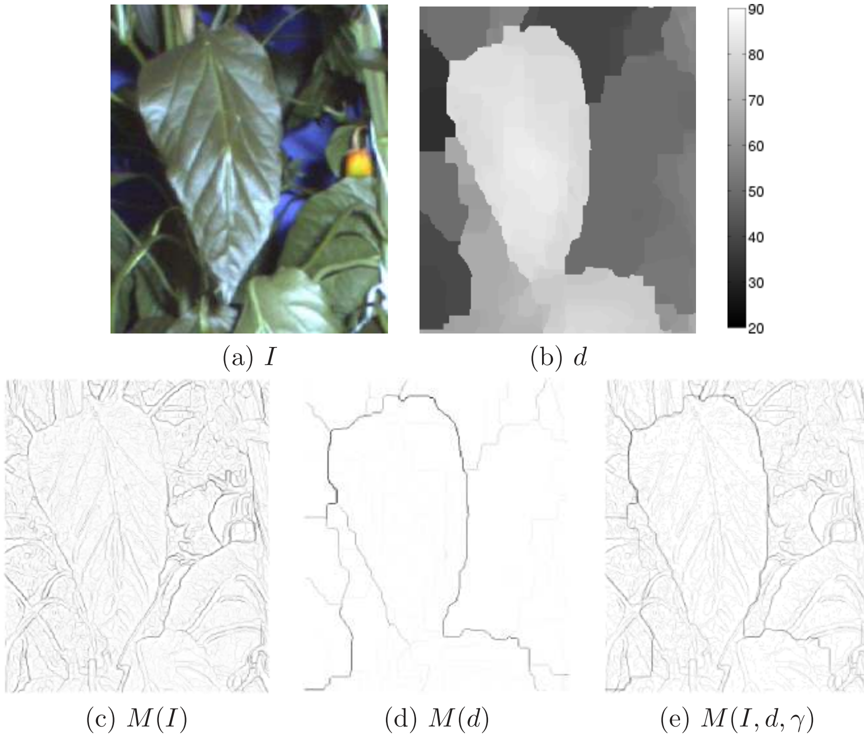
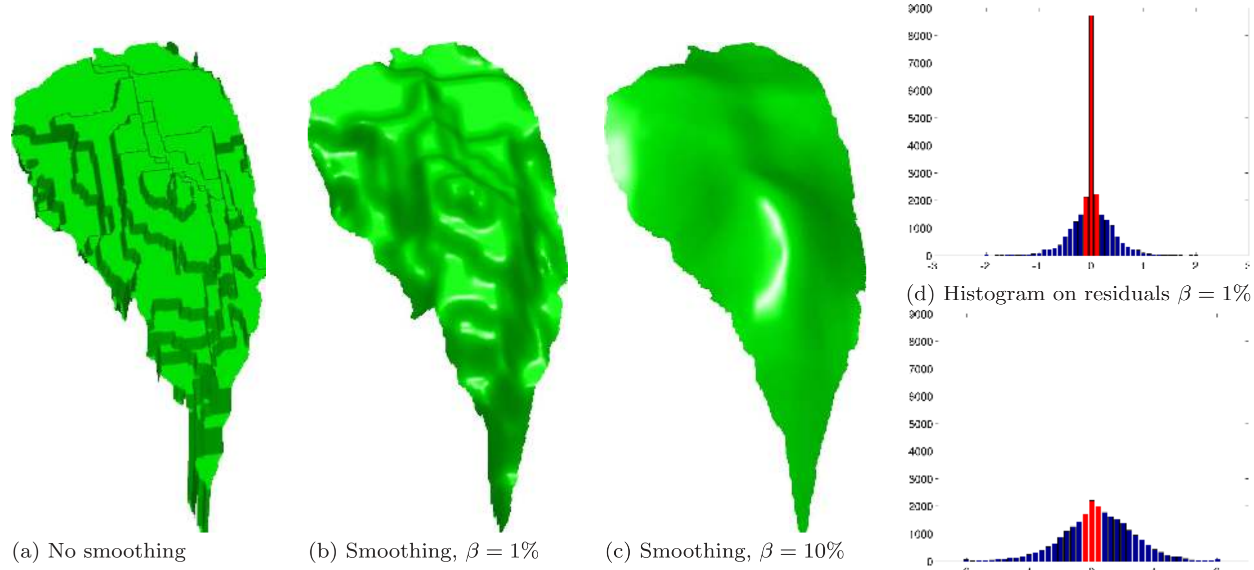


Fig. 6 Edge magnitude of colour and disparity images. (a) and (b) show the colour I and estimated disparity d images respectively. (c) and (d) show edge magnitudes by the Canny filter $M(I)$ and $M(d)$, and (e) shows the combined Canny edge magnitude. For this example, the weight coefficient γ is set to 0.4.

Fig. 7 Comparison between automatic leaf detection and manual selection. Yellow boundaries represent automatic leaf detection by our method and red boundaries represent manual selection.



(a) No smoothing (b) Smoothing, $\beta = 1\%$ (c) Smoothing, $\beta = 10\%$

(d) Histogram on residuals $\beta = 1\%$

(e) Histogram on residuals $\beta = 10\%$

Fig. 8 Reconstruction of a leaf surface with smoothing parameter β. (a) shows reconstruction without smoothing and the 'rice terrace' effect is clear. (b) shows little LOESS smoothing with $\beta = 1\%$, which is also clear from histogram on residuals in (d). (c) shows LOESS smoothing with $\beta = 10\%$. (e) shows histogram on residuals with $\beta = 10\%$, and there is a pattern of changes in the three bins near 0 in red colour compared with (d).

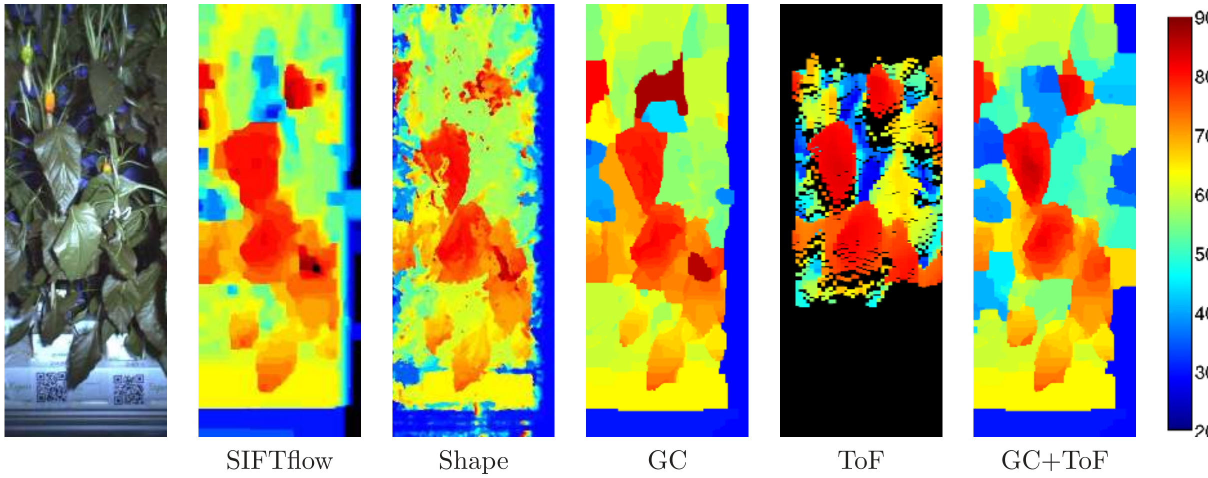


Fig. 9 Disparity results on the 'Plant 1'. SIFTflow, Shape and GC represent methods by Liu *et. al.* [29], Ogale and Aloimonos [34] and Boykov *et. al.* [5] respectively. ToF shows transformed points in colour image coordinates and the black pixels indicate missing ToF information. GC+ToF is our method combining stereo and ToF images.

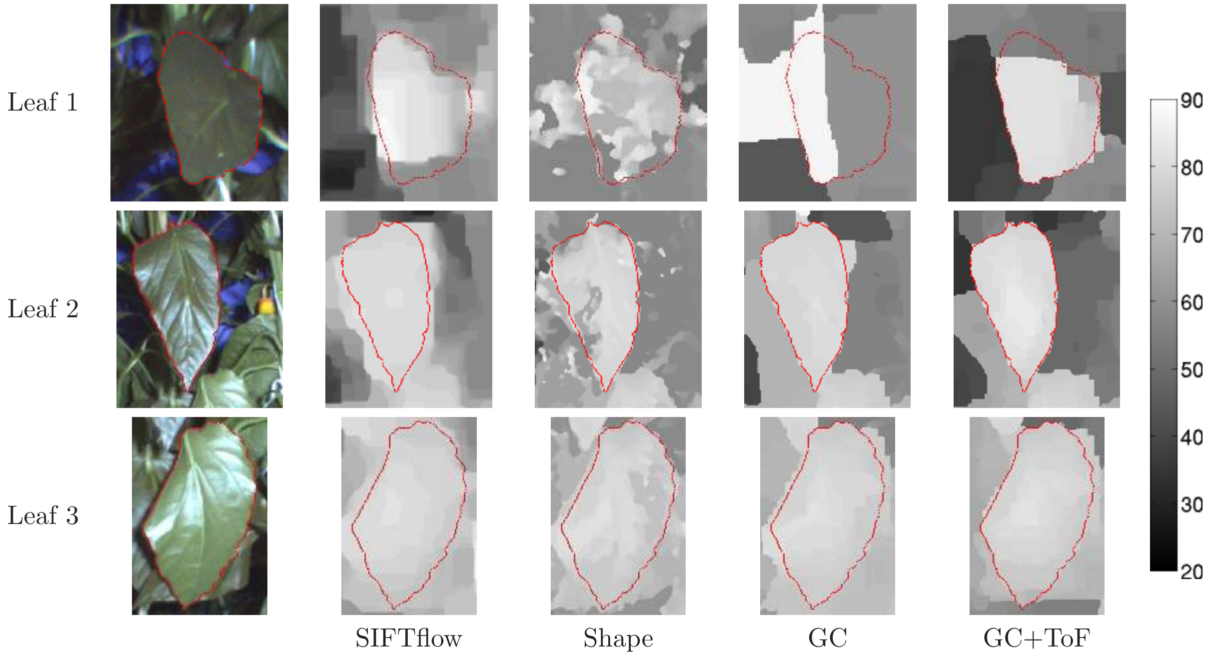


Fig. 10 Results for three leaves from top to bottom in Plant 1. Depth edges are plotted in red.

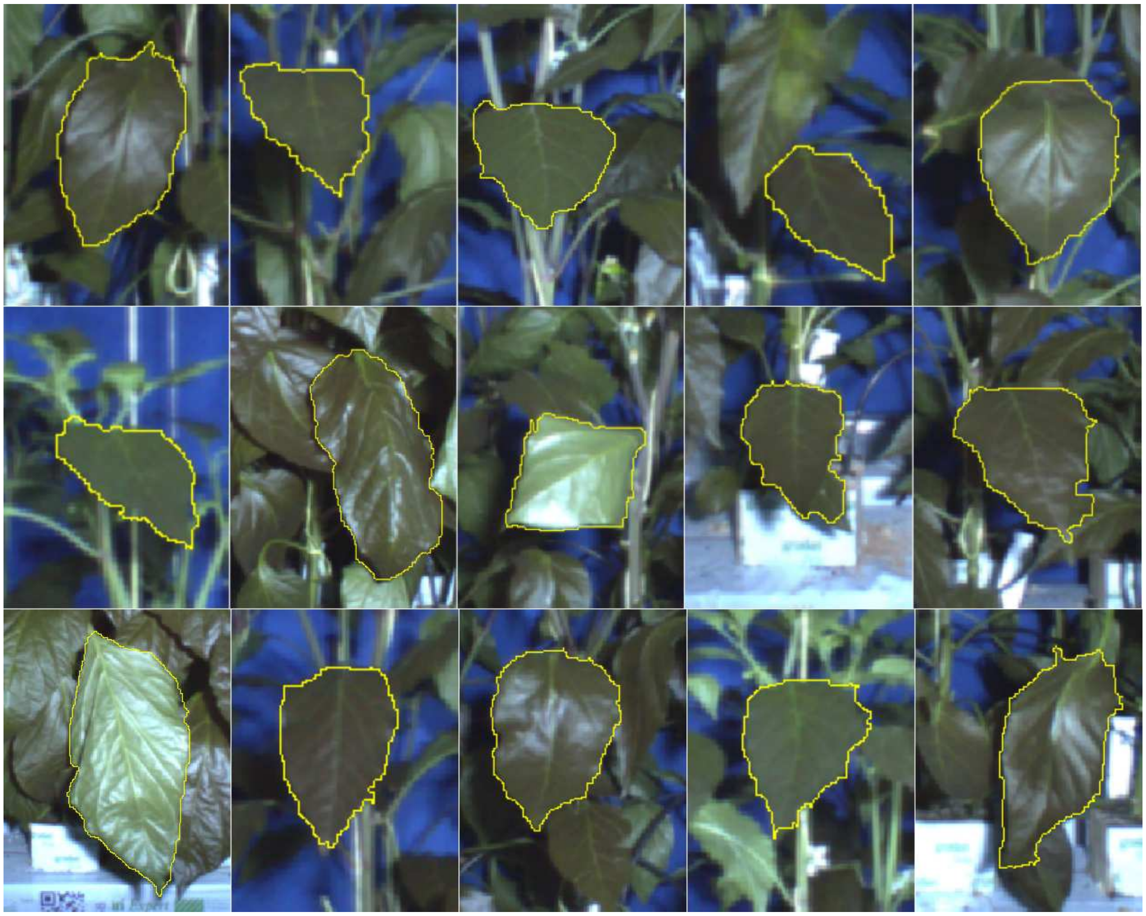


Fig. 11 Results on 15 examples of automatic leaf detection by our method. The yellow boundary outlines an automatically identified leaf. The 15 examples were randomly selected from the possible 149 leaves automatically detected in the validation images.

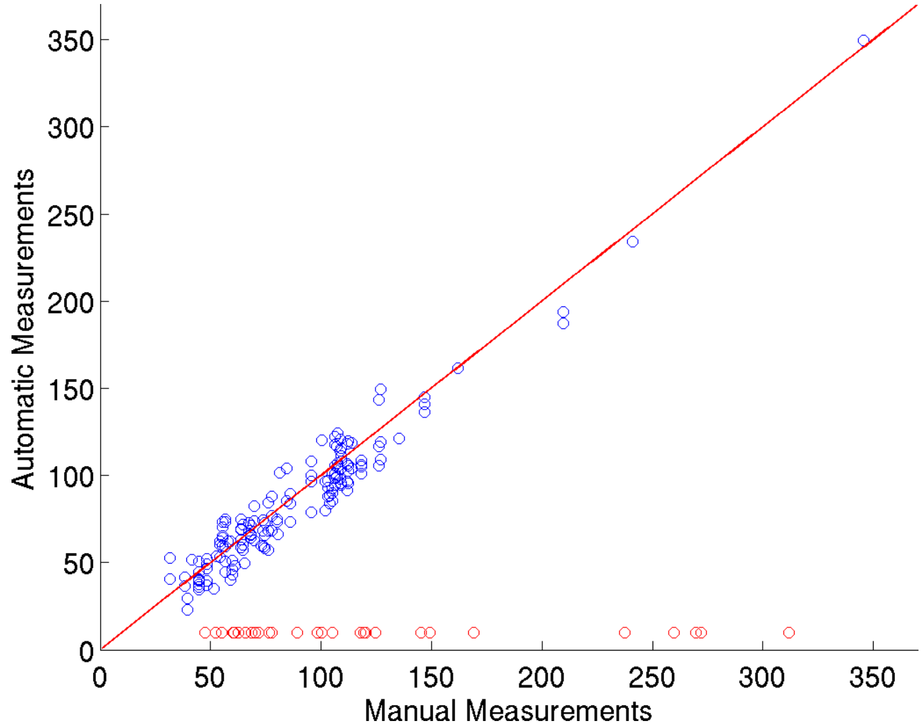


Fig. 12 Plots of 149 automatic leaf area measurements automatically obtained using our system against manual measurements on the validation images. The x-axis is manual measurements in cm^2, and the y-axis is automatic measurements in cm^2. The red line is the 1 : 1 reference line, and the red circles show the 29 leaves that could not be automatically detected.