

Combining Stereo and Time-of-Flight Images with Application to Automatic Plant Phenotyping

Yu Song¹, Chris A. Glasbey¹, Gerie W.A.M. van der Heijden²,
Gerrit Polder², and J. Anja Dieleman³

¹ Biomathematics and Statistics Scotland, The King's Buildings,
Edinburgh, EH9 3JZ, UK
{yu,chris}@bioss.ac.uk

² Biometris, Wageningen UR, PO Box 100, 6700 AC Wageningen, Netherlands
{gerie.vanderheijden,gerrit.polder}@wur.nl

³ Wageningen UR Greenhouse Horticulture, P.O. Box 644,
6700 AP Wageningen, Netherlands
anja.dieleman@wur.nl

Abstract. This paper shows how stereo and Time-of-Flight (ToF) images can be combined to estimate dense depth maps in order to automate plant phenotyping. We focus on some challenging plant images captured in a glasshouse environment, and show that even the state-of-the-art stereo methods produce unsatisfactory results. By developing a geometric approach which transforms depth information in a ToF image to a localised search range for dense stereo, a global optimisation strategy is adopted for producing smooth and discontinuity-preserving results. Since pixel-by-pixel depth data are unavailable for our images and many other applications, a quantitative method accounting for the surface smoothness and the edge sharpness to evaluate estimation results is proposed. We compare our method with and without ToF against other state-of-the-art stereo methods, and demonstrate that combining stereo and ToF images gives superior results.

1 Introduction

In our post-genomic world, where ever increasing volumes of genetic information are obtained at great speed and little cost, the collection of phenotypic information is often a bottleneck to scientific progress. A phenotype is any observable characteristic of an organism such as its shape and height. In stark contrast to genotyping, phenotyping is slow, and expensive in human time. Moreover, measurements are affected by the varying perception and interpretation of different observers. Image analysis has the potential to overcome these problems, but automatic interpretation of images of plants and animals remains very difficult. For example, Figures 1(a) and 1(b) show a stereo pair of images of pepper plants, from which we wish to estimate phenotypic characteristics such as leaf area, stem length or fruit size. This is a challenging task, as surfaces are of complex shape,

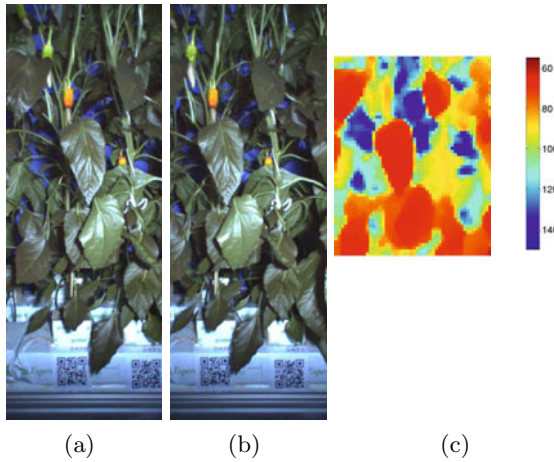


Fig. 1. Plant images: (a) and (b) show a stereo pair of images of pepper plants, (c) is the matching ToF image, which is at much coarser resolution and has been scaled to match (a) and (b)

and there are multiple depths, linear features and occlusions. Further, shadows are inconsistent between images because a flash light attached to the camera was used to offset ambient lighting.

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. Our component of 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, with the first step being recovery of dense depth information from image pairs such as Figures 1(a) and 1(b). This is usually seemingly-effortless for the human eye and brain, but unfortunately still not so for computers!

One approach to dense stereo is via robust point correspondence methods using local feature descriptors such as SIFT [12], followed by methods such as DAISY [17] and SIFTflow [11]. However, no convincing result addressing the issue of preserving discontinuity was produced in [11,17] for complex real world scenes. Global optimisation methods such as graph cuts [3] can produce edge-preserving results on the Middlebury dataset [15], but challenges in the Middlebury dataset are different to these images in our work. Ogale and Aloimonos [13] proposed to use shape in establishing edge-preserving dense correspondence, but their images were mostly planar surfaces.

Recently the use of low-resolution range cameras based on the Time-of-Flight (ToF) principle has received increasing attention. A ToF range camera is an active image sensor using infrared illumination, and distance measure in cm is calculated from the time the light has used for travelling to the object and back. Kolb *et. al.* [9] gave an overview on techniques and applications of ToF images, and these provide an option for improving recovery of depth information by augmenting stereo pairs with partial, coarse resolution, ToF images as in Fig. 1(c). Given the availabilities of high-resolution stereo images taken close to the viewpoint of the ToF sensor, it is natural to combine ToF and stereo results and develop statistical relations between them. In the direction of combining ToF and stereo, Gudmundsson *et. al.* [6] transformed ToF points into colour images by rectification homographies, and then fed them into a hierarchical stereo matching algorithm. Hahne and Alexa [7] demonstrated the combined ToF and stereo method can enhance the depth estimation even without accurate extrinsic calibration. Zhu *et. al.* [18] 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 [6,7,18], challenging low-resolution ToF images 48×64 were used in this work. Beder *et. al.* [1] also developed a fusion scheme using ToF images in the same resolution as ours, but their images were planar surfaces while ours are more complicated.

Current ToF and stereo fusion work (e.g. [1,6,7,10]) lack quantitative results on preserving depth-discontinuity, and most results were qualitative (except [18] which used another 3D scanner to produce pixel-by-pixel depth data). This is partly due to the fact that it is impossible to collect pixel-by-pixel depth data for ground truth. Our images were collected inside a glasshouse unlike the work by Zhu *et. al.* [18] which was done in an indoor lab environment, and the use of a ToF camera can become obsolete given the readily available and accurate depth data. In the situation without pixel-by-pixel depth data, we propose a method to quantify how much depth-discontinuity has been preserved and evaluate the quality of depth estimation for our approach as well as other state-of-the-art stereo methods.

In addition to address challenges raised from our applications, the technical contributions of this paper are: a simple yet effective geometric approach transforming ToF points and producing a localised search range for dense stereo; a global graph-cut strategy using the localised search range with an emphasis on preserving discontinuity; an evaluation method to determine the quality of estimation without pixel-by-pixel depth data. After describing our contributions, this paper presents comparison results on some challenging pepper plant images.

2 Methods

2.1 Setup and Calibration

The camera rig consisted of a colour camera and a ToF camera. The ToF camera is a RF 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 rig known as Spy-See [14] moved in a straight line on top of rigid heating pipes in the glasshouse and captured overlapping images at a fixed interval. The baseline between images was 5 cm, and objects of interest (e.g. leaves) were located between 55 cm and 120 cm away from the camera. Fig. 1(a)-(c) show a pair of stereo colour images and the corresponding ToF image.

Given a rigid and fixed camera setup described above, a two-layer board shown in Fig. 2(a) was used for calibration at different distances from the camera. The front layer moved from 40 cm to 120 cm away from the camera in 5 cm steps, and we used a simple pinhole camera model for the colour camera. Denote s as the baseline distance between images measured in cm and f' as the focal length for the colour camera, and the relationship between the disparity d and the depth z is,

$$d = s f' / z \quad (1)$$

Given multiple depth measurements \mathbf{z} (e.g. 40 cm to 120 cm in this work) and correspondences in each view to compute \mathbf{d} , \hat{f}' can be obtained by applying the least squares fitting technique:

$$\hat{f}' = \arg \min_{f'} \|\mathbf{d} - (s f' / \mathbf{z})\|^2 \quad (2)$$

The centre of the square seen in Fig. 2(a) is used to compute \mathbf{d} , and \mathbf{z} is known for each image. Fig. 2(b) presents the relationship in (1) and plots \mathbf{d} against \mathbf{z} .

2.2 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 [11] 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 [11].

For our application, discontinuity preserving results are highly desirable. The pepper plant images shown in Figures 1(a) and 1(b) 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 [11] 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

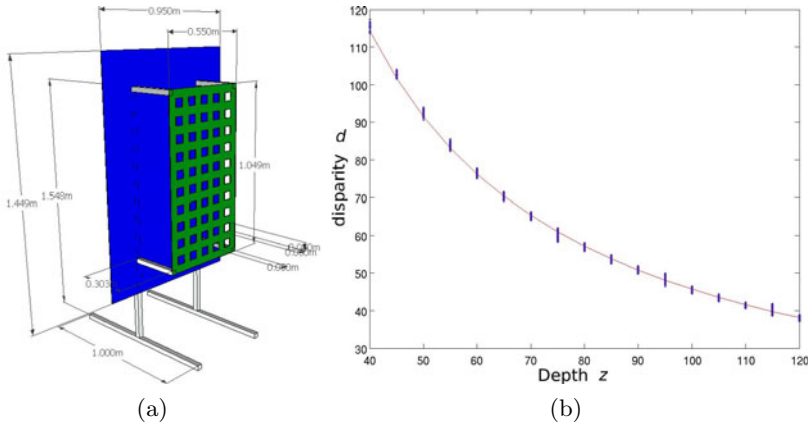


Fig. 2. Calibration: (a) diagram of calibration board, (b) plot of the relationship between depth z in cm and disparities d in pixels for colour camera. Blue dots were disparity measurements d for each z , and the red line was the fit by (2).

SIFTflow method can preserve discontinuity. Ogale and Aloimonos [13] 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 dataset [15]. Since global stereo methods produce better results compared with local stereo methods for combining with ToF information [18], we chose the alpha expansion technique applied in a graph-based energy minimisation framework [3]. The energy cost E given a pixel disparity d is defined as:

$$E(d) = \sum D(d_{(x',y')}) + \sum_{q \in N} V(d_{(x',y')}, d_{(x'_q, y'_q)}) \quad (3)$$

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

$$D(d_{(x',y')}) = \min \left\{ \frac{1}{3} \sum_{c \in \{R,G,B\}} \left| I_{(x',y')}^{(c)} - I'_{(x'+d_{(x',y')}, y')}^{(c)} \right|, T_d \right\} \quad (4)$$

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 ,

$$V(d_{(x',y')}, d_{(x'_q, y'_q)}) = u_{(x',y',x'_q, y'_q)} \min \left\{ \left| d_{(x',y')} - d_{(x'_q, y'_q)} \right|, T_k \right\} \quad (5)$$

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.* [3], which was used as an indicator function in this work as:

$$u_{(x', y', x'_q, y'_q)} = \begin{cases} \alpha_v & \text{if } \sum_{c=\{R,G,B\}} \left| I_{(x', y')}^{(c)} - I_{(x'_q, y'_q)}^{(c)} \right| > 25 \\ n_v \alpha_v & \text{otherwise} \end{cases} \quad (6)$$

α_v is the smoothness cost for intensity edges, and $n_v \alpha_v$ is the smoothness cost for surfaces. The thresholding value 25 was empirically determined from our experiments. Both α_v and $n_v \alpha_v$ should be set according to the data cost values in (4). (6) 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 $n_v \alpha_v$.

2.3 Localised Search Range from ToF Image

Given the complexity associated with the pepper plant images for dense stereo methods, a localised search range derived from the corresponding ToF depth image should improve the estimation accuracy. First of all, a transformation should be established for points in colour image and ToF image. In our experiments, a near-linear relationship between ToF depth measurements z'' and z was observed as in [18]. We used the same procedure for \hat{f}' to obtain \hat{f}'' for the ToF camera, and developed ToF transformation methods in (7)-(10) to tolerate errors in ToF camera. For further information on ToF camera calibration, Kolb *et. al.* [9] briefly discussed error sources and challenges, and Beder and Koch [2] developed a checkerboard method and calibration software. 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 Fig. 3 for an example). In effect, this transformation is one of the up-scaling techniques as discussed by Lindner *et. al.* [10] 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 $z_{(\mathbf{m}, \mathbf{n})}$,

$$|\mathbf{m} - x'| \leq r \quad (7)$$

$$|\mathbf{n} - y'| \leq r \quad (8)$$

In effect, this allows mis-alignment up to r pixels when transforming the ToF image to the colour image. The maximum and minimum depths ($\max\{z_{(\mathbf{m},\mathbf{n})}\}$ and $\min\{z_{(\mathbf{m},\mathbf{n})}\}$) are then converted into disparities as,

$$d_{min(x',y')} = s f' / \max\{z_{(\mathbf{m},\mathbf{n})}\} - k \quad (9)$$

$$d_{max(x',y')} = s f' / \min\{z_{(\mathbf{m},\mathbf{n})}\} + k \quad (10)$$

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. To incorporate the localised search range in a graph-based energy minimisation framework, for the data term cost D in (4), 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.

2.4 Quality Quantification

From Fig. 1(a) and 1(b), we see a few foreground leaves with depth edges present along the leaf boundary. Although pixel-by-pixel depth data were not available, we can label depth edges to quantify how well the result has preserved depth edges. The Canny filter was used to detect intensity edges, and these edges were then manually refined for leaf boundaries (see Fig. 4). Note that we only performed this manual edge refinement 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 was considered a leaf surface, and the final output was a binary image with surface pixels located at (x'_s, y'_s) and edge pixels located at (x'_e, y'_e) . To compute the smoothness of the surface and sharpness of the depth edges, we applied 3×3 Sobel operators in both horizontal and vertical directions, and surface smoothness penalty P_s was calculated as,

$$P_s = \overline{M_{(x'_s, y'_s)}} \quad (11)$$

where M is the edge magnitude by the Sobel operators. Edge sharpness score S_e was calculated as,

$$S_e = \overline{g_{(x'_e, y'_e)}} \quad (12)$$

where g denotes the edge magnitude M convoluted with 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 standard deviation to 5. A quality score S accounting for the surface smoothness P_s and the edge sharpness S_e was therefore computed as below,

$$S = S_e - P_s \quad (13)$$

The score S penalises displacement between defined depth edges and depth edges produced by a dense method while requiring the surface to be smooth.

3 Results

This section compares three dense stereo algorithms with our method on some challenging pepper plant images. Let SIFTflow, Shape and GC represent methods by Liu *et. al.* [11], Ogale and Aloimonos [13] and Boykov *et. al.* [3] respectively. GC refers to the graph cut method without using ToF, and GC+ToF is the method we propose in this paper. Parameters for all the methods were optimised by running the particular method for several iterations through all parameters with different testing orders, and both qualitative and quantitative results were taken into account. For all our experiments, 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_d, α_v, n_v, T_k for GC were set as 20, 4, 4, 6. 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 [11,13,3] for SIFTflow, Shape and GC respectively.

Fig. 3 shows an example of qualitative stereo results produced by the four methods. Methods GC and GC+ToF produced results with leaves recognisable from the background. SIFTflow produced smooth results but did not preserve discontinuity, while Shape was opposite. This can be further examined in Fig. 4, which shows effects of (11) and (12) on a close-up of a leaf (Leaf 1) by the four methods. The edge magnitude 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 best qualitative results among the four methods, and this was verified by two more examples in Fig. 5 (Leaf 2 and Leaf 3).

A summary of quantitative results (S_e, P_s, S) for all three leaves is shown in Table 1. 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 strong as those in Leaf 1 and Leaf 2. GC+ToF still produced the best scores S_e and S among the four methods.

This section has shown results on one example of stereo images (Pepper 1), and two more qualitative results (Pepper 2 and Pepper 3) have been made available¹. Table 2 presents $\sum S$ for all three results. By using ToF as a localised search range, the estimation results were improved by at least 16% measured by the score $\sum S$.

¹ <http://www.bioss.ac.uk/staff/yu/tof>

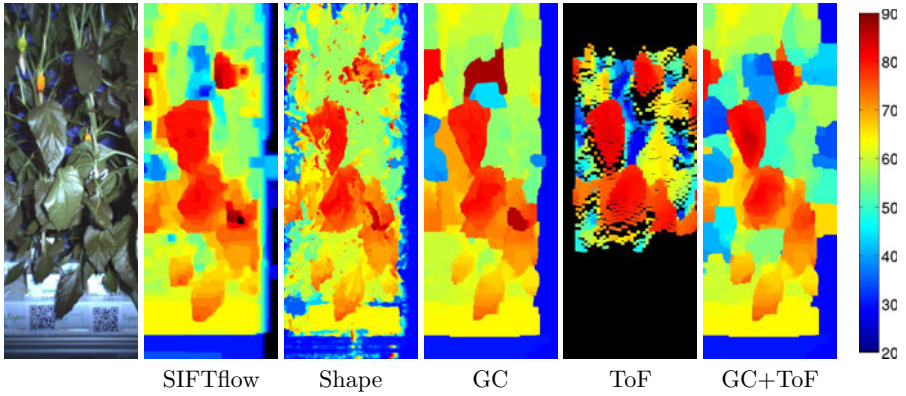


Fig. 3. Disparity results on the ‘Pepper 1’. ToF shows transformed points in colour image coordinates and the black pixels indicate missing ToF information.

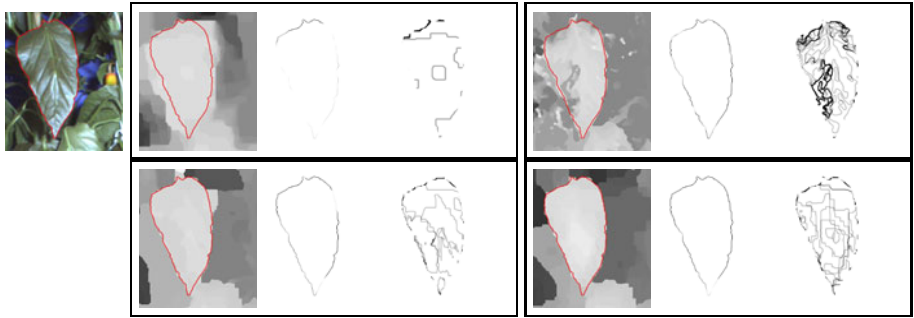


Fig. 4. Quality evaluation for Leaf 1. In every panel, the grey values represent the disparity (left), S_e (middle) and P_s (right) respectively. All four disparity maps use a common scale shown in Fig. 5. Also the S_e and P_s images use a common scale. Left column: base colour image with depth edges plotted in red. Middle column: results by SIFTflow (upper) GC (lower). Right column: results by Shape (upper) GC+ToF (lower).

Table 1. 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.

	Leaf 1			Leaf 2			Leaf 3		
	S_e	P_s	S	S_e	P_s	S	S_e	P_s	S
SIFTflow	4.55	0.81	3.74	9.35	5.04	4.30	3.56	0.94	2.62
Shape	12.66	4.09	8.56	12.81	8.79	4.02	6.52	1.90	4.61
GC	14.27	1.66	12.61	7.80	3.86	3.94	4.76	0.93	3.83
GC+ToF	20.89	2.76	18.13	20.76	6.20	14.56	7.20	1.65	5.54

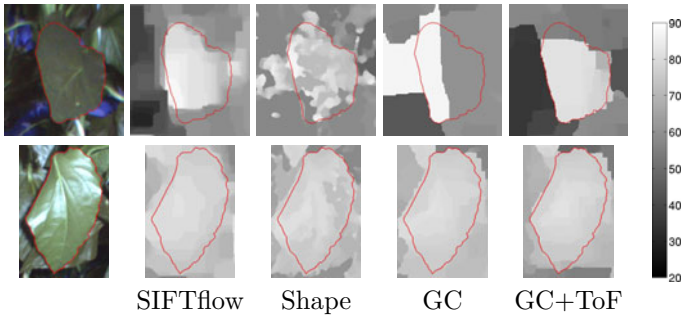


Fig. 5. Results for Leaf 2 and Leaf 3. Upper row: base image and estimates by four methods for leaf 2. Lower row: base image and estimates by four methods for Leaf 3. Depth edges are plotted in red.

Table 2. Quantitative summary of quality evaluation for three image examples. Figures shown here are total quality scores for all leaves in a image, $\sum S$.

	Pepper 1	Pepper 2	Pepper 3
SIFTflow	10.65	3.21	7.80
Shape	17.19	22.65	9.66
GC	20.37	21.03	10.50
GC+ToF	38.22	31.38	12.21

4 Discussion

This paper shows that dense stereo matching is not a trivial task for the pepper plant images collected inside a glasshouse. All three state-of-the-art methods produced unsatisfactory results, but a simple yet effective geometric approach to transform coarse-resolution ToF image together with a global graph-cut strategy can produce smooth results and preserve discontinuity. We have provided both visual and numerical results to demonstrate this. Fig. 6 presents a surface reconstruction for Leaf 1 using depth estimates by the proposed method combining stereo and ToF results, and depths have been recovered for such a complex surface. Although the quality score can quantify the quality of estimation without pixel-by-pixel depth data as shown in the results section, it only considers the surface smoothness and the edge sharpness without comparing the depth values. This was our first attempt on quality quantification that is different to those in [15,18], and we would like to draw the community’s attentions on evaluation methods producing quantitative results of depth-discontinuity preservation.

Combining ToF and stereo offers two main advantages. For occlusions and areas affected by unpredictable illumination, the data term in a global stereo framework (i.e. D in (4)) produces inaccurate energy costs since corresponding pixels are either unavailable or difficult to be matched. Using ToF in these situations provided an estimate and reduced ambiguities. Another advantage is that

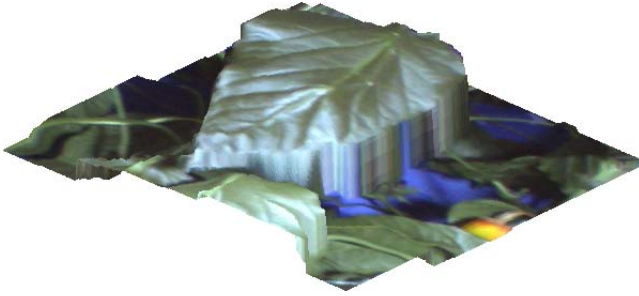


Fig. 6. Surface reconstruction for Leaf 1 using depth estimates by the GC+ToF method combining stereo and ToF results

dense stereo can be a super resolution technique for ToF images as discussed by [4,16], and we have presented discontinuity preserving results by combining ToF and stereo (e.g. Fig. 4). We have not considered environmental effects or measurement uncertainties related to the ToF camera as in [5,9] (e.g. the reflection issue in a cluttered environment and the influence of intensity on depth), which are beyond the scope of this paper.

Although we presented the method for one pair of stereo and one ToF images, 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. As Kim *et. al.* [8] have shown some promising results on this subject, we hope to build on the work in this paper for combining multiple colour and ToF images. Problems faced by stereo methods would be easier given multiple views, and the baseline would also be increased allowing more accurate depth estimation. We will then consider how to deal with occlusion and visibility issues within a multiple-view framework.

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)

References

1. Beder, C., Bartczak, B., Koch, R.: A combined approach for estimating patchlets from PMD depth images and stereo intensity images. In: Hamprecht, F.A., Schnörr, C., Jähne, B. (eds.) DAGM 2007. LNCS, vol. 4713, pp. 11–20. Springer, Heidelberg (2007)
2. Beder, C., Koch, R.: Calibration of focal length and 3d pose based on the reflectance and depth image of a planar object. *International Journal of Intelligent Systems Technologies and Applications* 5, 285–294 (2008)

3. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23, 1222–1239 (2001)
4. Diebel, J., Thrun, S.: An application of markov random fields to range sensing. In: *Proceedings of Conference on Neural Information Processing Systems (NIPS)*. MIT Press, Cambridge (2005)
5. Gudmundsson, S.A., Aanaes, H., Larsen, R.: Environmental effects on measurement uncertainties of time-of-flight cameras. In: *International Symposium on Signals Circuits and Systems* (2007)
6. 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)
7. 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)
8. 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 the 3DIM 2009* (2009)
9. Kolb, A., Barth, E., Koch, R., Larsen, R.: Time-of-Flight Sensors in Computer Graphics. In: Pauly, M., Greiner, G. (eds.) *Eurographics 2009 - State of the Art Reports Eurographics*, pp. 119–134. (2009)
10. 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)
11. Liu, C., Yuen, J., Torralba, A.: Sift flow: Dense correspondence across scenes and its applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2010)
12. 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)
13. Ogale, A.S., Aloimonos, Y.: Shape and the stereo correspondence problem. *International Journal of Computer Vision* 65, 147–162 (2005)
14. 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*. National Physical Laboratory, pp. 115–117 (2009)
15. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision* 47(1-3), 7–42 (2002)
16. Schuon, S., Theobalt, C., Davis, J., Thrun, S.: Lidarboost: Depth superresolution for tof 3d shape scanning. In: *Proceedings of the IEEE CVPR 2009* (2009)
17. Tola, E., Lepetit, V., Fua, P.: Daisy: an efficient dense descriptor applied to wide baseline stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32(5), 815–830 (2010)
18. 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* (2008)