

# Indoor and Outdoor Depth Imaging of Leaves With Time of Flight and stereo vision Sensors: Analysis and Comparison

Wajahat Kazmi<sup>a</sup>, Sergi Foix<sup>b</sup>, Guillem Alenya<sup>b</sup>, Hans Jørgen Andersen<sup>a</sup>,  
wajahat@create.aau.dk, sfoix@iri.upc.edu, galenya@iri.upc.edu, hja@create.aau.dk

<sup>a</sup>*Department of Architecture, Design and Media Technology, Aalborg University, Sofiendalsvej 11, 9200 Aalborg, Denmark.  
Tel: +45 9940 7156*

<sup>b</sup>*Institut de Robotica i Informatica Industrial, CSIC Universitat Politècnica de Catalunya, Llorens Artigas 4-6, 08028 Barcelona, Spain.  
Tel:+34 9340 54261*

---

## Abstract

In this article we analyze the response of Time of Flight cameras (active sensors) for close range imaging under three different illumination conditions and compare the results with stereo vision (passive) sensors. Time of Flight sensors are sensitive to ambient light and have low resolution but deliver high frame rate accurate depth data under suitable conditions. We introduce some metrics for performance evaluation over a small region of interest. Based on these metrics, we analyze and compare depth imaging of leaf under indoor (room) and outdoor (shadow and sunlight) conditions by varying exposures of the sensors. Performance of three different time of flight cameras (PMD CamBoard, PMD CamCube and SwissRanger SR4000) is compared against selected stereo-correspondence algorithms (local correlation and graph cuts). PMD CamCube has better cancellation of sunlight, followed by CamBoard, while SwissRanger SR4000 performs poorly under sunlight. stereo vision is more robust to ambient illumination and provides high resolution depth data but it is constrained by texture of the object along with computational efficiency. Graph cut based stereo correspondence algorithm can better retrieve the shape of the leaves but is computationally much more expensive as compared to local correlation. Finally, we propose a method to increase the dynamic range of the ToF cameras for a scene involving both shadow and sunlight exposures at the same time using camera flags (PMD) or confidence matrix (SwissRanger).

*Keywords:* leaf imaging, depth, exposure, time of flight, stereo vision, sunlight.

---

## 1. Introduction

In agricultural automation, 2D imaging has addressed a variety of problems, ranging from weed control (Slaughter and Giles, 2008) and disease detection (Garcia et al., 2013) to yield estimation (Nuske et al., 2011), inter plant space sensing (Tang and Tian, 2008) and structural analysis McCarthy (2009), to name a few. But most of these tasks are either large scale analysis or they tend to deal with simpler plant canopies, for example, at early growth stages (Astrand and Baerveldt, 2004). The reason is obvious; when looking inside plant canopy, 2D imaging is not robust to occlusion of plant organs such as overlapping leaves and branches.

To address this problem, 3D imaging has also been applied. Among the most noticeable applications of 3D vision are the construction of dense models for simulation of plant structures (Takizawa et al., 2005) and for estimating 3D properties of plant canopies (Chapron et al., 1993; Preuksakarn and Boudon, 2010; Santos and Oliveira, 2012). If not obvious, then at least an ambiguous difference between the application domains of 2D and 3D imaging in agriculture can be observed. 2D has been successfully applied for outdoor and 3D for indoor applications

and large scale analysis in outdoor scenario such as navigation in the field (Kise and Zhang, 2008).

The reason for this gap is threefold; firstly, plants have complicated free form, non-rigid structures that cannot be approximated by simple geometrical shapes making it necessary to observe minute details and hence placing stringent demands on the quality and the efficiency of 3D imaging technology. Secondly, the huge variations in outdoor illumination (sunny, partially cloudy, overcast, shadow), which can change the perceived shape of objects to a large extent and that even constrains 2D imaging. Thirdly, the technology for 3D data acquisition is largely designed for indoor applications and exporting it to outdoor scenario either limits the scope or makes the system too complex to be practical. For example, (Biskup et al., 2007) used stereo vision for only measuring leaf inclination for outer leaves of plant canopies under outdoor lighting and (Nakarmi and Tang, 2010) used a Time of Flight camera for plant space measurement covering the view from sunlight otherwise the sensor saturates. In general, any such approach for 3D analysis is focused at a particular application and cannot be easily adapted for obtaining slightly different measurements.

Even after all the shortcomings, 3D sensing is vi-

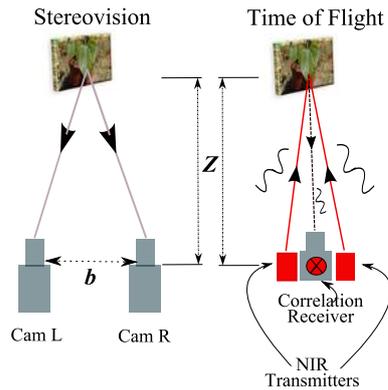


Figure 1: Comparison of Stereo and ToF techniques

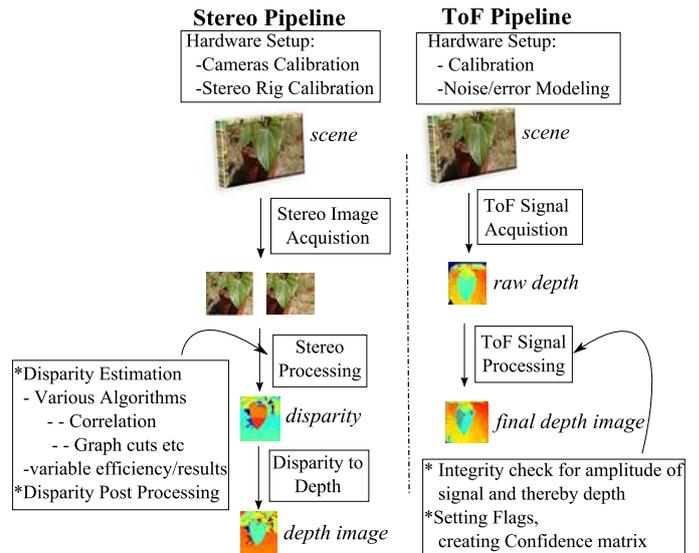


Figure 2: Block Diagrams of Stereo and ToF depth image acquisition pipelines

tal. Plant phenotyping facilities require accurate depth measurements of plant organs (such as, leaf count/angles/areas, plant height or sampling points on specific sections of a plant) either for classification of large varieties of plants produced due to experimental genetic variations (van der Heijden et al., 2012) or robotic manipulation such as for measuring chlorophyll content in leaves (Alenya et al., 2011) or automated fruit picking (Jimenez et al., 2000). In field operations, it has great potential in precision agriculture for reducing the amount of herbicides as 3D data can help in not just improved recognition and localization of weeds by resolving occlusion but also in estimation of volume of the infestation, thereby enabling deployment of optimal amounts of chemicals (Nielsen et al., 2004; Kazmi et al., 2010).

Recently, Fiorani et al. (2012) discussed state-of-the-art technologies in use for biological imaging and pointed out that in depth knowledge is required regarding physics of the sensors and parameters of software/algorithms used in order to benefit optimally. This is a bottleneck in agricultural automation because the objects (plants) pose one of the most demanding tests to image acquisition and computer vision. Systems optimized for man made structured environments are not optimal for a natural setup of agriculture. Limitations of imaging system combined with environmental factors make agricultural imaging a complex puzzle to solve. Therefore, it is important to segregate environmental factors and evaluate the sensor performance w.r.t to each one.

One of the most important factors is light, both indoor and outdoor. Lighting must be diffused to reduce errors. Under outdoor conditions, various shading arrangements have been used to cater for that or else experiments are performed on days with overcast (Frasson and Krajewski, 2010). But the problem arises when introducing a shade makes the system either too complicated, such as, in weed detection (Piron et al., 2011) or sunlight is unavoidable, for example, to understand the effect of lighting variations on the plant canopies (Van der Zande et al., 2010), to track the diurnal/nocturnal movement of the

leaves (Biskup et al., 2007) or with changing positions of the sun (van Henten et al., 2011). In such cases, exposure of the 3D imaging system must be either robust to variation in ambient illumination or at least tangible, somehow. The effect of ambient illumination on the camera response varies with the type of sensor used.

### 1.1. Common 3D Data Acquisition Techniques and Challenges

The most widespread method of acquiring 3D data is stereo vision. But it has a big set of problems. Stereo correspondence and depth accuracy vary with the type of algorithm used. Local correspondence algorithms are efficient but less accurate than global ones which could be, computationally, very expensive. Besides, performance is adversely effected by lack of surface texture of the object and specular highlights.

Among the active sensing technologies, structured light projection and laser range scanners are used for creating accurate and detailed 3D models, but such systems can be expensive and complex. Structured light has interference issues outdoors. New low cost versions of structured light cameras have low resolution and are highly sensitive to outdoor lighting (such as RGBD cameras e.g. Microsoft Kinect<sup>1</sup>). Laser scanners include mobile parts and require longer imaging times.

On the other hand, recent advances in the Time of Flight (ToF) based range sensors have revolutionized the industry and several brands of off-the-shelf 3D cameras are available in the market. They use near infrared (NIR) emitters and generally produce low resolution depth images. However, a gradual increase in sensor resolution has been observed over the last few years. ToF cameras produce high

<sup>1</sup><http://www.microsoft.com/en-us/kinectforwindows>

frame rate (up to 50 fps) depth images and therefore are highly suitable for real-time applications. But the problem of lack of performance under sunlight, still remains i.e. these sensors are guaranteed to work only in indoor environments. Some of the ToF cameras have an on-board background illumination rejection circuitry such as PMD (Möller et al., 2005), but with varying performance under sunlight depending on the operating range and the power of NIR emitters.

The challenge in ToF cameras is to find a suitable integration time (IT: a controllable parameter related to the length of time sensor integrates the returned signal) according to the ambient illumination because a different calibration has to be applied for each IT and the calibration is a costly process. For stereo vision, the challenge is the performance and accuracy of correspondence algorithm and the effects of ambient illumination on the accuracy of disparity map. Fig. 1 show a comparison of working principle and Fig. 2 of data processing pipelines for both stereo vision and ToF technologies.

In our previous work, we have evaluated the performance of one ToF camera for close range leaf imaging (Kazmi et al., 2012). But every ToF camera has different sensor properties and robustness against background illumination. A qualitative comparison of the response of several different ToF cameras with stereo vision under indoor/outdoor illumination conditions, particularly for agricultural purposes, is not available in literature. Such sensor characteristics would be very helpful for analyzing the performance of these sensors and weighting the cost of making a choice.

## 1.2. Objective

In this article, our objective is to estimate and compare the response of ToF and stereo vision sensors for depth imaging of leaves using some of the commonly used cameras. We will first review their current applications in agriculture. Since a lot of literature has addressed resolution and accuracy of stereo vision (e.g. Scharstein and Szeliski, 2002; Kytö et al., 2011) we will only provide a short insight into the precision of ToF cameras.

We will introduce some metrics for qualitative evaluation of depth data. We also propose a method for obtaining the most suitable camera configurations for imaging under different illumination conditions. The method is based on observing the trends in camera precision and detecting the non-linearities in the amplitude. Additionally, we show that for ToF cameras, using this information through pixel flags or confidence matrix, a high dynamic range image can be obtained by combining two different exposure for scenes with both sunlight and shadow present at the same time.

The breakdown of this paper is as follows: In Sec. 2 we will briefly review state-of-the-art applications of stereo vision and ToF in agriculture. Light reflection characteristics for the leaf surface will be presented in Sec. 3. In Sec. 4, precision of ToF cameras will be discussed. Sec. 5

deals with experiments in which Sec. 5.1 will introduce the cameras used in the experiments followed by the experimental setup in Sec. 5.2. Metrics for qualitative evaluation of depth data will be explained in Sec. 5.3. Data analysis for ToF will be done in Secs. 5.4 and 5.5 and for stereo vision in Sec. 5.6. In Sec. 5.7, we will explain how to exploit camera flags to enhance the dynamic range of ToF cameras. This follows a comparative discussion in Sec. 6 along with a brief analysis of the validation data. Sec. 7 concludes the paper.

## 2. 3D Vision In Agriculture

In agriculture, almost all the commonly known technologies for 3D data acquisition have been used, such as stereo vision, ToF sensors as well as structured light projection and laser range scanning. However, stereo vision and ToF sensors are easily deployable and less complicated modes of 3D data acquisition. Therefore, we discuss these two only.

### 2.1. Stereo Vision in Agriculture

Stereo analysis has been successfully used indoors for example Mizuno et al. (2007) used stereo vision for wilt detection in indoor conditions. Going further deep inside the canopy, Chapron et al. (1993) and Takizawa et al. (2005) used stereo vision to construct 3D models of plants in indoor conditions. From the models, they extracted information, such as plant height, leaf area and shapes, which are helpful in plant recognition. Yoon and Thai (2009) combined stereo vision with NDVI index to create a stereo spectral system for plant health characterization in lab conditions.

For in-field operations, it has been successful for imaging at larger scales, for example Kise and Zhang (2008) and Blas (2010) used it for guidance and navigation in the fields. Rovira-Más et al. (2005) used aerial stereo images for growth estimation. Use of stereo vision for corn plant space sensing both indoor and outdoor has been demonstrated by Jin and Tang (2009).

To some extent such structural measurements from stereo based 3D data have been tried outdoor. Ivanov et al. (1995) used top down stereo images of maize plants in the fields to find structural parameters such as leaf orientation, leaf area distribution and leaf position to construct a canopy model. Even after performing destructive analysis of the of the plant to view inner leaves, the 3D model properties were not promising. However, methods and imaging apparatus has improved a lot since then. Biskup et al. (2007) used stereo vision for measuring leaf angle and tracking their diurnal and nocturnal movements for soyabean leaves both for indoor and outdoor. Leaf inclination angle was found by fitting a plane to the reconstructed 3D leaf surface.

Stereo vision performance, however, is poor for close range observation of surfaces, such as, leaf, because of

the homogeneous texture which produces pockets of missing depth information. Global correspondence algorithms, which are computationally more expensive, tend to deal with this problem better. Andersen et al. (2005) conducted experiments on 3D analysis of wheat plants in lab conditions and compared simple correlation based stereo matching (local) to simulated annealing (global) which is a more accurate method and offered better results. Nielsen et al. (2004) used a trinocular stereo system for weed identification against tomato plants but found it difficult to extend it for real-time applications.

Besides texture, sunlight is also an important factor affecting stereo vision performance. In order to avoid sunlight, either a shade is used or else experiments are carried out on days with overcast (Frasson and Krajewski, 2010). Along with the environmental factors, camera settings must be tuned to provide optimal results. Nielsen et al. (2005, 2007) experimented with 3D model creation using stereo with real and simulated plant models. They pointed towards the need for further research in the close range stereo imaging of plants particularly taking into account color distortion in the sensors and exposure control.

As can be observed from this brief but representative literature review, keeping wind factor aside, stereo vision has two major problems in outdoor conditions. First is the strong sunlight and the second is the inherent limitations of the stereo matching process which is not robust to all sorts of surfaces and objects expected in agricultural scenarios. This reduces the efficacy of stereo vision and limits either the scope or the scale of the application.

## 2.2. ToF Imaging in Agriculture

ToF cameras have not been a dominant mode of 3D data acquisition so far. There are two main reasons. Firstly their low resolution and secondly, the cost. The cost of the ToF cameras has recently fallen and resolution has also improved slightly (max 200x200, still not comparable to stereo cameras). Therefore, as compared to stereo vision, fewer applications of ToF in agriculture have appeared. But due to their benefits over conventional 3D systems, (as discussed in Sec. 1.1, they are becoming more popular.

Kraft et al. (2010) and Klose et al. (2009) investigated feasibility of ToF cameras for plant analysis. They found it a good candidate for plant phenotyping but they failed to account for the IT which is a very important parameter and without it, ToF data evaluation becomes somewhat meaningless. Alenya et al. (2011) used ToF camera in indoor environments by combining depth data with RGB images for leaves. Going a step further, Song et al. (2011) used a combination of ToF and stereo images for plant leaf area measurements in green house conditions to increase the resolution for the depth data.

Nakarmi and Tang (2010) used ToF camera in corn fields for inter-plant space measurement. Wind and sunlight were blocked from the view using a shade.

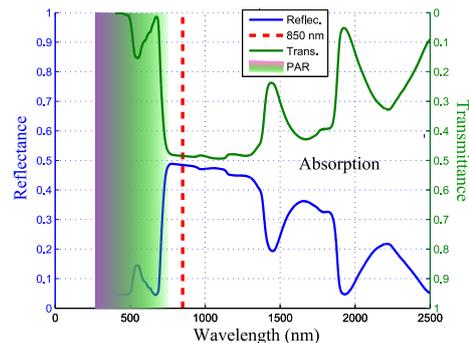


Figure 3: Reflectance-Transmittance characteristics of a green Soybean leaf (Feret et al., 2008)

This brief overview of ToF applications to agriculture imply that there are two major challenges with ToF; low resolution and sensitivity to outdoor illumination.

## 3. Light Reflectance from Leaf Surface

ToF cameras have NIR emitters. Although the NIR light is modulated at 10-400 MHz carrier frequency, its reflection, transmission and absorption depends on the NIR light. A leaf response to light interaction (which includes both the photo-synthetically active radiation (PAR) as well as infrared spectrum) varies with the wavelength of incident light. Since ToF cameras depend on travel time of light for depth estimation, it is therefore, important to carry out a brief survey on optical characteristics of light interaction with plant leaves, especially in the NIR.

### 3.1. Leaf Optical Characteristics

From the surface of the leaf some part of the incident light is reflected, some transmitted and the rest is absorbed. Woolley (1971) found that the reflectance and transmittance for soybean and maize leaves were both high in NIR region as compared to visible spectrum as the plants in general absorb significant amount of incident visible light. Similar results were achieved by reflectance-transmittance model proposed by Jacquemoud and Baret (1990). It showed almost 50% reflectance in the NIR region for green soya been leaves.

Although, this reflectance is low but in their findings, on a wavelength scale from 400 nm to 2500 nm which also includes part of the visible spectrum, the only region having highest reflectance and lowest possible transmittance and absorption is the NIR region between 750 nm and 1000 nm (Fig. 3). ToF cameras operating at 850-870 nm are therefore ideally suited for green leaf imaging but due to high transmission, the transmitted part is also partly reflected from the inner elementary layer of the leaf surface (Jacquemoud and Baret, 1990). Besides, leaf thickness also slightly affects the reflectance-transmittance ratio (Gates et al., 1965) and this along with color, effects the ToF data (Klose et al., 2009; Kraft et al., 2010). To some

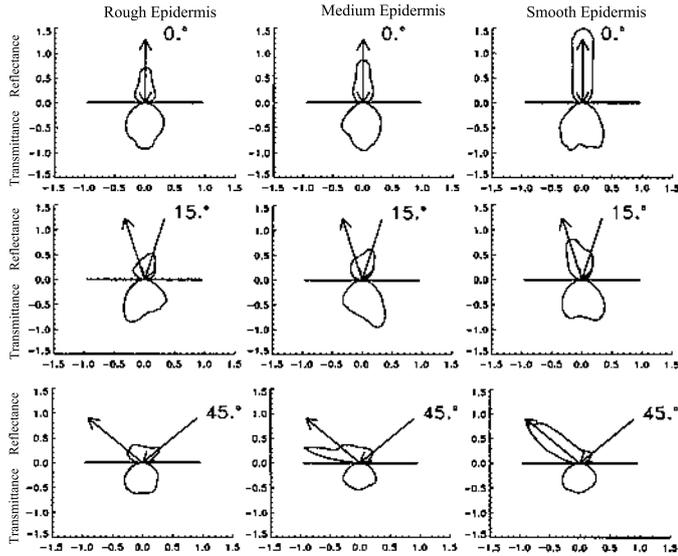


Figure 4: Leaf Reflectance in NIR using Monte Carlo Ray Tracing and 3D model of dicotyledon leaf (Govaerts and Jacquemoud, 1996)

extent, these errors can be taken care of through accurate calibration. But neither of them render ToF useless.

As far as PAR is concerned, much of the green wavelength is absorbed by the leaf (see Fig. 3). But unlike ToF imaging, this is not a problem for stereo vision since the depth estimate does not depend on the travel time of light and instead is derived from the correspondence of binocular vision.

For ToF imaging, ideally, the higher the reflectance, the better. But then comes the problems of sensor saturation due to strong reflections. As the ToF sensors are bi-static in nature, therefore, at short ranges the measured depth varies with the angle of incident illumination. For these reasons, it is also important to have some idea of light scattering properties of leaves in general.

### 3.2. Are Leaves Lambertian Reflectors?

A lambertian surface is the one which reflects isotropically in all directions. It is important to understand the scattering, specially of NIR light from the leaf surface because if the reflected beam is not directed back towards the sensor, estimates may be compromised depending on the amount of scatter. There has been a lot of work done towards understanding of light scattering from the leaf surface. It is measured by bidirectional reflectance factor (BRF), which is the ratio of radiance reflected by the surface to the radiance reflected from a perfectly diffused reflector, sampled over all directions. Brakke (1992) conducted experiments on red maple, red oak and yellow poplar leaves using both visible and NIR light. He found the scattering of both the wavelengths more isotropic for normal and near-normal incidence than higher incidence angles.

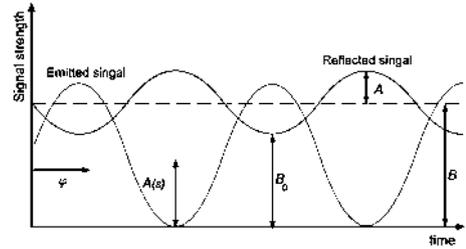


Figure 5: Influence of background illumination on ToF signal, Mufti and Mahony (Mufti and Mahony, 2011)

Chelle (2006) used digitized model of maize in order to verify lambertian approximation for both PAR and NIR in case of dense crop canopies because, in the case of plant canopies, leaf specularities does not dominate. For a single leaf however, the specular reflections play an important role especially for leaves with smooth epidermis which produce strong speckle particularly in the NIR band (due to high reflectivity as compared to PAR). This is also verified by Govaerts and Jacquemoud (1996) who used a 3D model of typical dicotyledon leaf and used Monte Carlo Ray Tracing to simulate reflectance characteristics of the leaves for varying surface roughness in NIR. As evident in Fig. 4, for leaves with smooth reflecting surfaces, a near normal incidence reflects most of the light back towards the same direction.

Since the primary leaf in our experiment has fairly smooth surface (Sec. 5.2), for ToF cameras, we keep the incidence angle of NIR light as well as the camera (receiver) orthonormal to the leaf surface. This may lead to early saturation and in case of visible spectrum, the orthonormal orientation of the stereo camera may find a speckle due to strong reflection which can produce mismatch in correspondence as it will change the position from left to right camera. This is the challenge which can be controlled through exposure tuning, as will be discussed in Sec. 5.6.

## 4. Precision of ToF Data

ToF cameras typically return registered depth, amplitude and intensity images of the same size (detailed description in Sec. 5.1.1). As discussed by May et al. (2006), IT affects the amplitude and intensity. Here, we will discuss the effects of amplitude and IT on depth precision.

Precision of depth data is directly related to the amplitude as given by (Lange and Seitz, 2001):

$$\Delta L = \frac{L}{2\pi} \cdot \Delta\varphi = \frac{c}{4\pi f_{mod}} \frac{\sqrt{B}}{\sqrt{2A}} \quad (1)$$

in this equation,  $\Delta L$  is the depth precision due to photon-shot noise (quantum noise),  $B$  (intensity) is the offset to account for background illumination  $B_0$  and mean amplitude  $A$  of the returned active illumination of the camera

(Mufti and Mahony, 2011): (see Fig. 5 for a graphical depiction)

$$B = B_0 + A \quad (2)$$

and  $L$ , the maximum depth, is calculated by the following relation:

$$L = \frac{c}{2f_{mod}} \quad (3)$$

where  $f_{mod}$  is the modulation frequency (Table 1) and  $c$  is the speed of light. For an ambitious reader, the details of the derivation can be found in (Lange, 2000, chap. 4). ToF signal to noise ratio (SNR) is given by (Mufti and Mahony, 2011):

$$SNR = \frac{\sqrt{2}A}{\sqrt{B}} \quad (4)$$

and by substitution, Eq. 1 can be reduced to:

$$\Delta L = \frac{c}{4\pi f_{mod}} \frac{1}{SNR} = \sigma_L \quad (5)$$

The Poisson distribution of the process of arrival of photons at the sensor represents the photon-shot noise. It can be approximated by a Gaussian distribution in case of very large number of photons, as in ToF cameras, which is the standard deviation  $\sigma_L$  of the range measurement (Mufti and Mahony, 2011). From Eq. 1, low amplitude ( $A$ ) decreases signal to noise ratio and makes depth invalid. On the other hand, a very high amplitude, saturates the sensor. And amplitude is directly controlled through IT for a given working distance, due to which IT plays a key role in precise depth estimation. At higher ITs, even before reaching saturation amplitude, sensor may receive sufficient background reflections to relate one pixel to several depths rendering the data invalid. Strong background illumination  $B_0$  (such as sunlight) increases  $B$  and in order to reduce its effect, IT must be reduced which in turn decreases  $A$ . From Eq. 2,  $B$  also includes  $A$ , but due to square root dependence of  $\sigma_L$  on  $B$ , an increase in  $A$  results in an overall increase of precision (Mutto et al., 2012).

If a working setup of ToF camera is moved from indoor to outdoor under sun,  $B$  increases without an increase in  $A$  due to strong sunlight, and therefore, for the same IT, precision drops. Therefore, no single IT can satisfy different ambient illumination settings. An optimal IT, hence, is a best compromise among  $\sigma_L$ ,  $A$  and  $B$  (precision, amplitude and background illumination). This fact makes use of the ToF cameras more complicated as for precise measurements, ToF cameras must be calibrated for a specific IT otherwise it will have integration time related errors (Foix et al., 2011).

## 5. Experiments

Experiments were performed using both ToF and stereo sensors with different correspondence algorithms. In this

section we will explain in detail all the equipment, experimental setup, data acquisition and the subsequent analysis.

### 5.1. Material and Methods

Four cameras were used in this experiment (see Fig. 4):

- PMD CamBoard (ToF)
- PMD CamCube (ToF)
- SwissRanger SR4000 (ToF)
- Point Grey Bumblebee XB3 (Stereo)

In the following, we will explain their specification and types of output data (see Table 1).

#### 5.1.1. ToF: Specifications and Output Data

ToF cameras work on the standard Lock-in Time of flight principle (Foix et al., 2011). NIR light is modulated at a carrier frequency. Continuous Wave (CW) modulation is used and phase difference is calculated to estimate depth rather than directly measuring the turn around time of the signal. Almost all ToF cameras deliver same type of data but have different sensor sizes. Maximum resolutions in Table 1 were used in this experiment at 20-30 fps.

*Depth.* Depth image returned by ToF cameras contains the  $Z$  coordinate of the scene in meters. Integrity of the depth values is judged by amplitude and flags (PMD) or confidence matrix (SR4000).

*Amplitude.* The amplitude has no specific units. It represents the amount of light reflected from the scene. Higher amplitude means more confidence in the measurement. But a very high returned signal strength leads to saturation which is an indication that no more photons can be accommodated by the pixels thus producing unreliable depth.

*Intensity.* Every ToF camera returns a typical 8 or 16 bit gray scale intensity image of the same resolution as depth image.

*Flags or Confidence Matrix.* PMD cameras produce a flag matrix to comment on the integrity of the every pixel. Each flag is a 32 bit value. If set, the reason could be one or more of the flags in Table 2. Sometimes, with Invalid flag, other bits provide information for the possible reason of invalidity. In case of SR4000, a confidence matrix does a similar job. Higher confidence of a pixel imply higher integrity of the corresponding depth data. On saturation, the confidence falls.

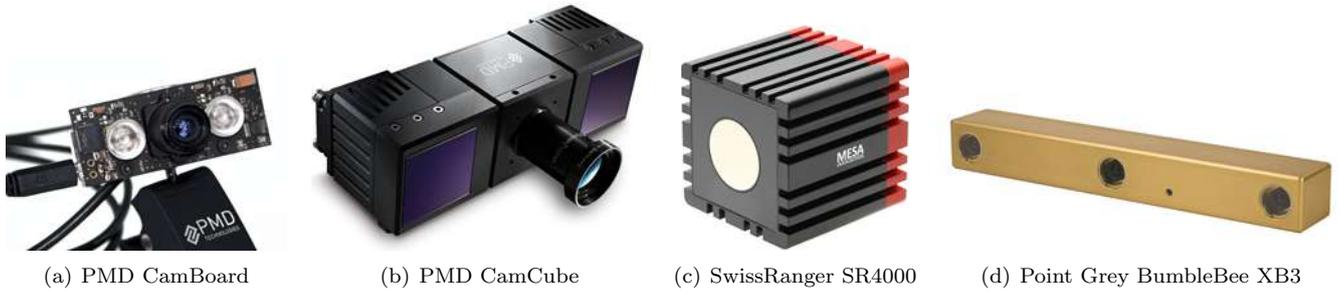


Figure 6: Cameras used in this experiment

Table 1: Camera Specifications

Name	Type	Mod. Freq. (MHz)	NIR (nm)	Max. Res.	Max. Range (m)	Max. FPS	Max. Exposure Time [step]( $\mu$ s)	Field of View	Precision (mm) $1\sigma$	Accuracy (mm) $1\sigma$	Output Images
PMD CamBoard	ToF	10-40	850	200x200	7	60	1-14000 [1]	40°(h) 40°(v)	10 <sup>†</sup>	—	Depth, Amplitude, Intensity, Flag
PMD CamCube	ToF	10-40	870	200x200	7	40	1-50000 [1]	40°(h) 40°(v)	<3 <sup>‡</sup>	—	Depth, Amplitude, Intensity, Flag
SwissRanger SR4000	ToF	29-31	850	176x144	5	50	300-25800 [100]	43°(h) 34°(v)	4 <sup>*</sup>	—	Depth, Amplitude, Intensity, Confidence
Point Grey Bumblebee XB3	Stereo (12,24)cm baseline	—	—	1280x960	3-4	15	10-66630 [10]	66°(h)	—	( $\Delta x = \Delta y$ , $\pm 0.166$ ), ( $\Delta z$ , $\pm 1.1$ ) <sup>‡</sup>	Color, Greyscale, Disparity

Manufacturer specifications for central pixels at [Range, Object Reflectivity, Modulation Frequency]:

<sup>†</sup>[1 m, 80%, 20 Mhz], <sup>‡</sup>[4 m, 75%, 20 Mhz], <sup>\*</sup>[2 m, 99%, 30 Mhz]

<sup>‡</sup>at 80 cm, 480x640 resolution (480 pixels focal length), baseline 12 cm, calibration error 0.1 RMS pixels and correlation error 0.1 pixels (Point Grey, 2012)

### 5.1.2. Stereo Vision Specifications

Point Grey Bumblebee XB3 at 12 cm short baseline was used in this experiment. Its specifications are listed in the Table 1. This camera has lenses locked in a fixed assembly and is provided with company calibration for both stereo rig and lens distortion which is quite accurate so that the disparities lie only in the horizontal direction. In this experiment, camera resolution was set at 480x640 with 3.75 fps. Images were obtained by varying Shutter Time (ST).

The camera comes with Triclops package which generates disparity images using a local correlation algorithm based on Sum of Squared Differences (SSD) <sup>2</sup>. It is optimized for efficiency and has a number of validation steps to prune disparity. Stereo accuracy reported in Table 1 is for Triclops by Point Grey Research. As mentioned in (Point

Grey, 2012), the actual values may vary a lot depending on the surface texture and correspondence algorithm. In order to compare quality of Triclops, we have also used a non-optimized implementation of a local correlation algorithm and Graph Cut (GC) based global stereo matching (Kolmogorov and Zabih, 2002). GC based algorithms apply global correspondence, are slower than local correlation, but perform better in terms of disparity accuracy and shape retrieval (Scharstein and Szeliski, 2002). The implementation shared by Vladimir Kolmogorov <sup>3</sup> were used for both. Triclops matching window size of 11x11 was used (*tclps.m11*). For non-optimized local correlation matching, windows sizes of 9x9 (*corr.m9*), 11x11 (*corr.m11*) and 13x13 (*corr.m13*) were used. For GC, default values of the parameters ( $\lambda_2 = \lambda$ ,  $\lambda_1 = 3\lambda$ ,  $K = 5\lambda$ , Intensity Threshold=5, data cost=L2) were used since they

<sup>2</sup><http://www.ptgrey.com/support/kb/index.asp?a=4&q=48&ST=triclops>

<sup>3</sup><http://pub.ist.ac.at/~vnk/software.html>

produced the best results with  $\lambda$  set to *AUTO*. Disparities were calculated with interaction radius 1 (*GC\_rad1*) and 2 (*GC\_rad2*). For all the three algorithms, sub-pixel precision was used for disparity computation (Haller and Nedeveschi, 2012). Processing times in Table 4 were noted on Intel Core-i5 (quad core) 2.40 GHz processor with 4 GB RAM.

## 5.2. Imaging Setup

The primary object in this experiment was a plant leaf, *Anthurium Andraeanum* (Fig. 7 (a)). Plants were grown in pots and the camera-to-leaf distance was between 25-40 cm for ToF and 75-85 cm for stereo. The distance of ToF was kept so low in order to get a high resolution view of the leaf since the ToF cameras are already very low in resolution. Image acquisition was done under indoor (room) and outdoor (shadow, sunlight) lighting conditions by varying exposures (IT for ToF and ST for stereo camera). The absolute distance between the lens of the camera and the leaf could be slightly different for each setting because the plant and camera mount was displaced every time. The cameras were mounted on a tripod looking down. Orientation of the cameras was kept roughly orthonormal (image plane roughly perpendicular to leaf surface normal) in order to keep maximum reflected light directed towards the lens. This is a realistic constraint in the current experimental scenario (Alenya et al., 2013) as argued in Sec. 3.1. During the test, company provided calibration was used for all the cameras.

Specifications in the manufacturer data sheet were at 20 Mhz  $f_{mod}$  for PMD Cameras and 30 Mhz for SR4000 (as reported in Table 1), so we used the same frequencies for the corresponding cameras to maintain consistency. Although, as from Eq.1, absolute value of precision depends on both  $A$  and  $f_{mod}$ , but in our case, we observe the changes in precision instead w.r.t to IT at a fixed  $f_{mod}$ .

In case of ToF, the operation range may be below the calibrated range but it only affects the absolute distance values and not the integrity or precision of the data. The only drawback would be that the data is not corrected for lens distortion and focal length and therefore the depth will not be accurate in meters. This is irrelevant as long as we focus on the validity of the depth data w.r.t to amplitude and pixel flags or confidence. After all, the purpose of the test is to find an optimal exposure so to calibrate the camera at that IT for higher accuracy for each of the illumination conditions individually.

Note that light conditions are crucial in the experimental setup, and in the case of sunlight, background illumination filtering can help to obtain better images. PMD camera architecture includes filtering of the infrared bandwidth (Kraft et al., 2004) and SR4000 uses a glass substrate bandpass filter<sup>4</sup>. For stereo, if over-exposition occurs, an appropriate bandwidth filter would help to acquire

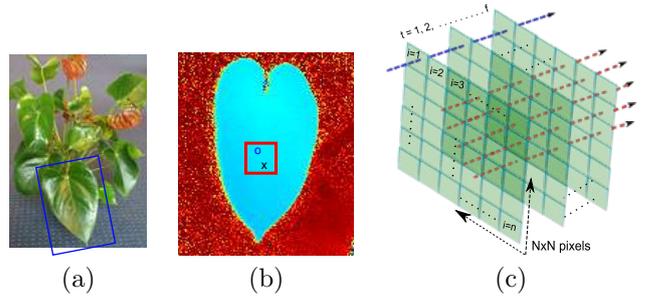


Figure 7: (a) One of the plants of *Anthurium Andraeanum* used in the experiment (b) 20x20 ROI on a leaf depth image used for analysis (c) Scheme for Mean and Std.Dev. across frames inside an  $N \times N$  ROI

better images. But in our experiments, we do not use any external filters and instead use them out of the box. This analysis in fact can help in selection of suitable filters.

## 5.3. Data Interpretation and Metrics for Evaluation

ToF data is high frame rate (up to 50 fps) so we analyzed ToF data by aggregating it across several frames using various statistical measures. This approach provides a fairly good estimate of camera response as described by Nayar and Mitsunaga (2000). On the border of the image, amplitude is usually lower than in the middle (Kahlmann and Remondino, 2006), therefore, focusing on a small window near the center of the image reduces errors. Hence a 20x20 Region Of Interest (ROI) was chosen on the leaf image. Data was analyzed for individual pixels as well as the entire ROI.

### Standard Deviation Across Frames

In order to empirically estimate the precision (Eq. 5), we first find the mean value of depth for each pixel  $i$  across  $f$  frames:

$$\mu_i = \frac{\sum_{t=1}^f L_{i,t}}{f} \quad (6)$$

where  $L_{i,t}$  is the depth value of pixel  $i$  of  $t$  frame. The scheme of data accumulation is shown in Fig. 7 (c). The standard deviation across frames is:

$$\sigma_{L_i} = \sqrt{\frac{\sum_{t=1}^f (L_{i,t} - \mu_i)^2}{f - 1}} \quad (7)$$

In order to extract a 1D metric to compare precision over IT variation, we find mean of  $\sigma_{L_i}$ :

$$\mu_{\sigma_L} = \frac{\sum_{i=1}^n (\sigma_{L_i})}{n} \quad (8)$$

where  $n = N \times N$  is the total number of pixels in the region of interest (ROI).

In case of stereo vision, only a single frame disparity was generated from a pair of stereo images. Mean depth and standard deviation were calculated from the ROI on a that disparity image only.

<sup>4</sup>[http://www.mesa-imaging.ch/dlm.php?fname=pdf/SR4000\\_Data\\_Sheet.pdf](http://www.mesa-imaging.ch/dlm.php?fname=pdf/SR4000_Data_Sheet.pdf)

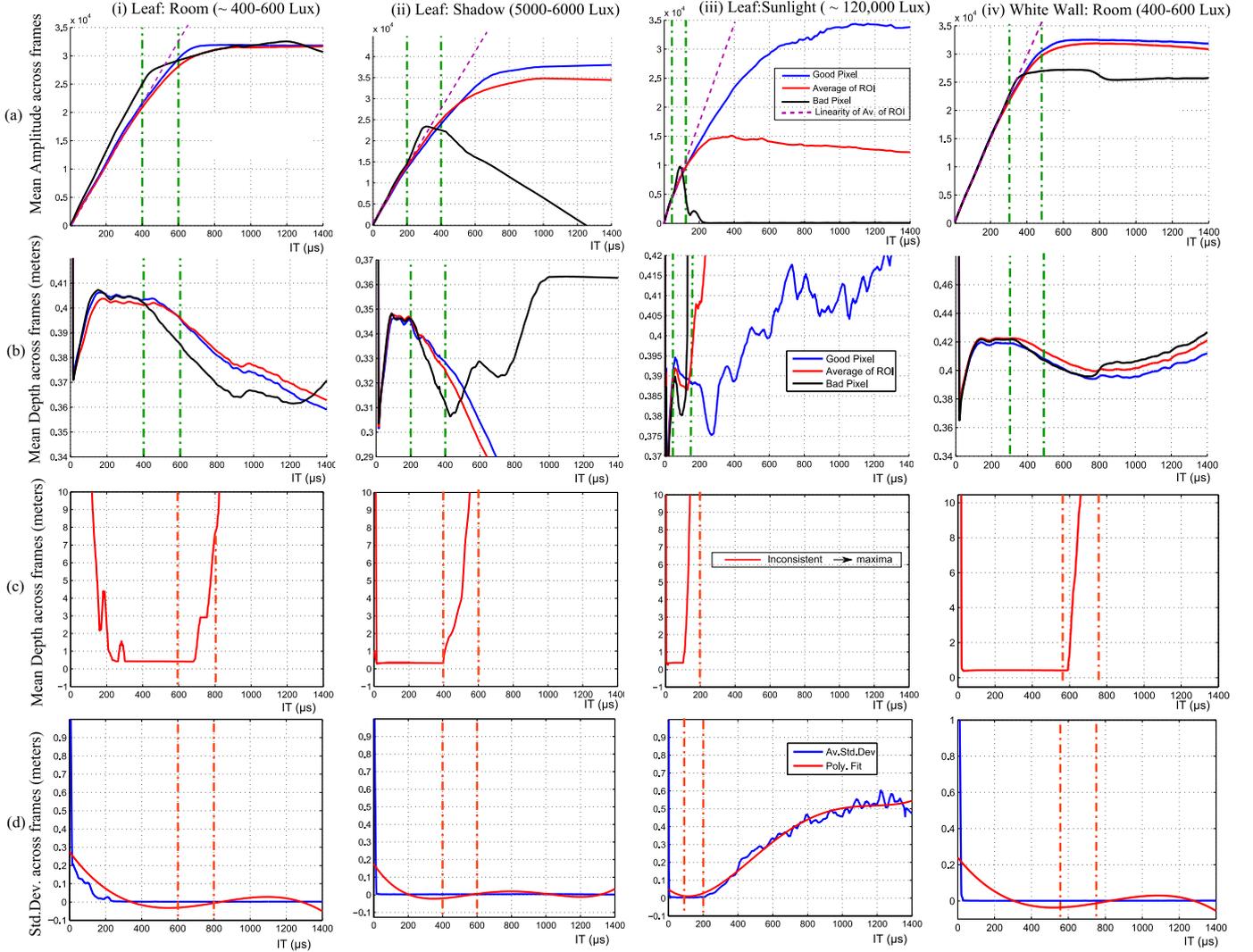


Figure 8: CamBoard response in amplitude, depth and precision between 25-40 cm aggregated in a 20x20 (ROI) under room, full shadow and sunlight conditions for one leaf in view and a white wall. Good and Bad pixels are among the last and first ones to be flagged inconsistent, respectively (Table 2). Both pixels are inside the ROI (c) *Inconsistent*  $\rightarrow$  *Maxima* shows the curve when inconsistent pixels are set to maximum depth manually (d) Poly. Fit is the trend line for the mean Std.Dev. curve (blue line) of ROI (a) A quick glance shows CamBoard performance in descending order as (iv) white wall: best to (i) room leaf: better, (ii) shadow leaf: satisfactory and (iii) sunlight leaf: worst.

#### 5.4. Camera Response Analysis: PMD

Maximum value of amplitude in PMD cameras is above  $2.5 \times 10^4$  units. Fig. 8, 9 row (a) show mean amplitude for CamBoard and CamCube, respectively. The graph shows characteristics for a single pixel and a 20x20 ROI on the leaf for room, shadow and sunlight conditions. In these graphs, good pixels are those that reach amplitude saturation and before that, their amplitude deviates from linearity in all the three conditions. Bad pixels on the other hand, exhibit a very unpredictable behavior and do not necessarily reach saturation amplitude at all. Due to these bad pixels, the mean amplitude deviates from linearity much earlier.

Assuming the same material properties for all the pixels on the sensor of the camera, it appears that non-linearity is not related to amplitude maxima only. It may occur

even before. Fig. 8, 9 row (b) show the corresponding depth values of pixels which start getting out of synchronization as soon as the corresponding amplitudes deviate from linearity. The IT range where this behavior occurs is shown between dotted green bars, we call it the *green zone*.

According to May et al. (2006) and also by our observation, this deviation depends on the distance between the object and the camera. For closer objects, it occurs earlier (lesser IT) than for distant objects. Similarly, it occurs earlier for more reflective objects at a given distance, for instance, the white wall in column (iv).

It can also be seen that ambient illumination affects the behavior of the amplitude curve. Deviation from linearity occurs earlier under sunlight (strong background illumination) than in shadow and room conditions. If the IT is

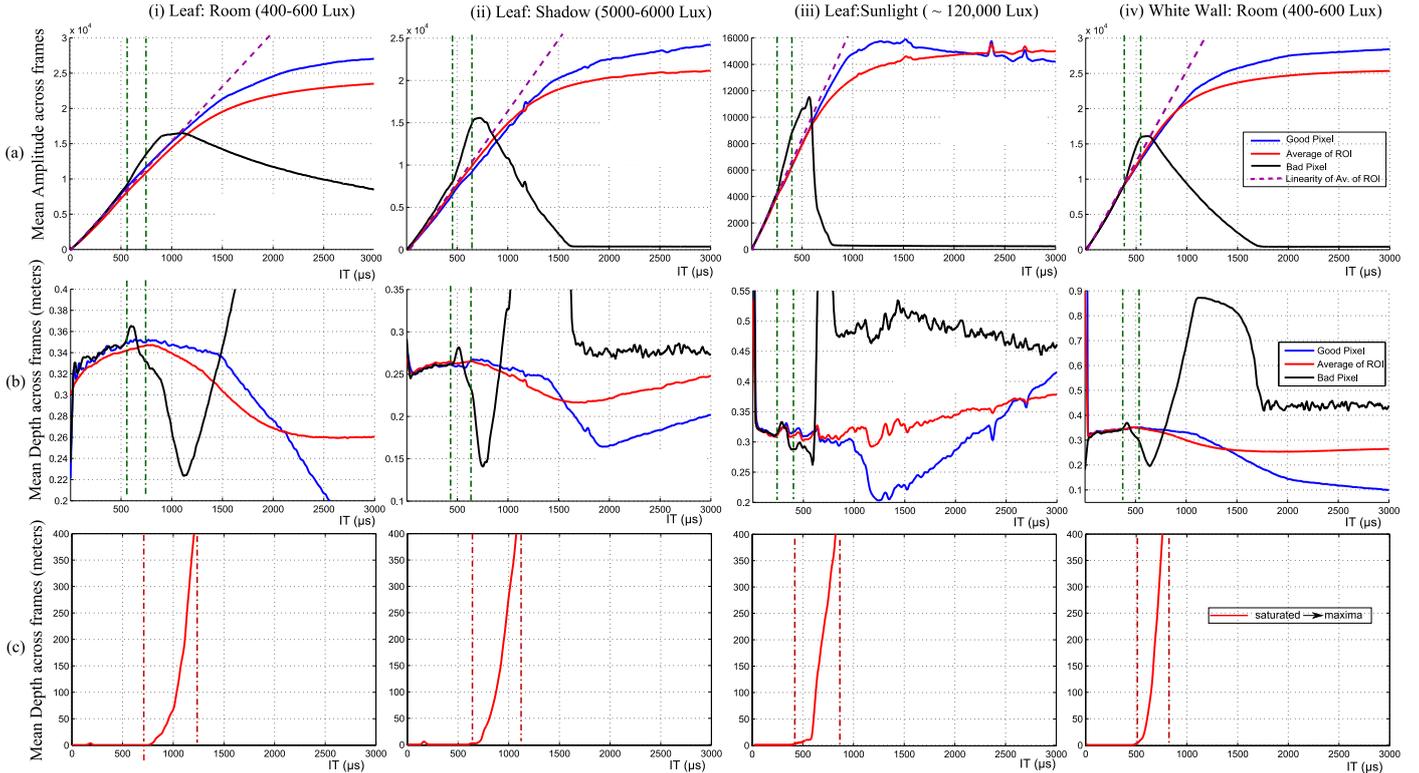


Figure 9: CamCube response in amplitude and depth between 25-40 cm in a 20x20 (ROI) under room, full shadow and sunlight conditions for one leaf in view and a white wall. Good and Bad pixels are among the last and first ones to be flagged saturated, respectively (Table 2). Both pixels are inside the ROI (c) *Saturated*  $\rightarrow$  *Maxima* shows the curve when saturated pixels are set to maximum depth manually (a) A quick glance shows CamCube performance in descending order as (iv) white wall: best to (i) room leaf: better, (ii) shadow leaf: satisfactory and (iii) sunlight leaf: worst.

Table 2: PMD Flags

Flag	Identifier	Description
Invalid	01(hex)	Depth unreliable due to very high or low amplitude
Saturated	02(hex)	Amplitude saturation reached (CamCube)
Inconsistent	04(hex)	Raw data inconsistent (CamBoard)
Low Signal	08(hex)	Not enough signal strength
SBI Active	10(hex)	Suppression of background light active (CamBoard)

increased further than the green zone, the deviation continues and worsens to a point that the PMD on-board system sets inconsistent flag for the CamBoard and saturation flag for CamCube, which in-turn also sets the invalid flag in both (Table 2). Inconsistency, according to CamBoard documentation, means multiple depth values pertaining to one pixel or incompatible amplitude which can also be caused by saturation. In rows (a) and (b), *good* and *bad* pixels are among the last and first ones to be flagged inconsistent/saturated, respectively.

As discussed in Sec. 4, depth precision is directly related to the amplitude of the received signal therefore amplitude must be high enough to enable correct depth measurement, still below the saturation level. Amplitude increases with IT. In order to find the highest possible IT suitable for a given setting, let us consider the precision or standard deviation (Eq. 8) in Figure 8 row (d). This standard deviation is the mean precision of the ROI (Eq. 8). Since both PMD cameras show similar trends in precision w.r.t mean amplitude and depth, so only CamBoard precision curve is used for analysis.

From the figure, in case of room conditions, the precision is quite low at very low IT. It improves with increasing IT. The most important part is the first valley of the standard deviation or its trend line i.e. when precision is the highest for the first time. This indicates that there is a consensus among the values of one pixel across all the frames and for the whole ROI. Any second valley will not be important because it will be due to saturation which would still bring consensus among frames. The rise of the trend line after the first valley indicates fall of precision which is indicated between red dotted bars, we call it the *red zone*. Number of bad pixels start increasing somewhere in the red zone. But the depth or amplitude values may not significantly change at this stage. They are just different from frame

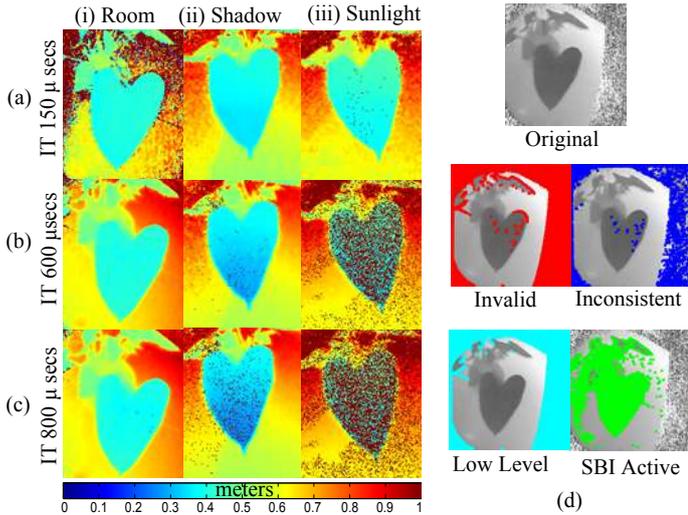


Figure 10: CamBoard: (a,b,c ) Depth images of plant leaves under the three different ambient illumination conditions. Inconsistent pixels appear as dark spots. (d) Flags set at 900 IT  $\mu$ s under room conditions

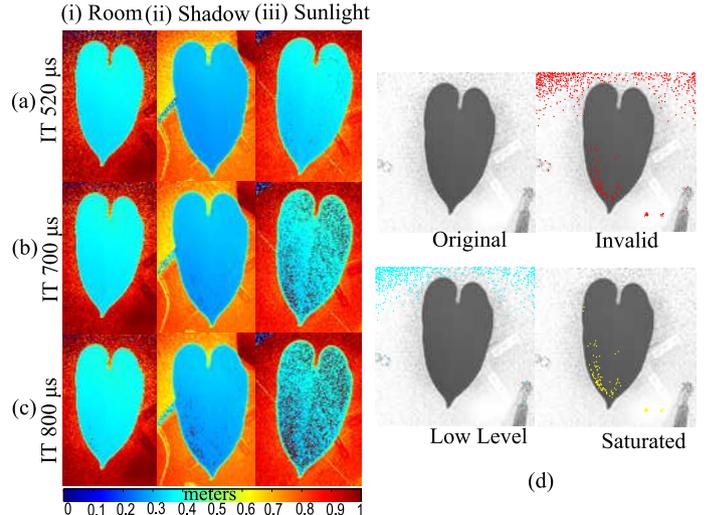


Figure 11: CamCube: (a,b,c) Depth images of plant leaves under the three different ambient illumination conditions. Saturated pixels appear as dark spots. (d) Flags set at IT 900  $\mu$ s under room

to frame, due to which precision drops.

In order to see the corresponding change in the mean depth curve, we set the depth of the inconsistent or saturated pixels to a very high value as soon as the flags are set. The result is displayed (Fig. 8, 9 (c)). The sharp rise in the curve indicates the IT at which the pixels are turning bad. This provides us the upper threshold of the IT for each ambient condition. The ITs at which the number of bad pixels become significant is shown in Table 3. In our general observation, the point where the green zone starts usually lies at IT values 20-30 % below these values.

In other words, the appearance of inconsistent or saturated flag is an early warning of non-linearity of amplitude and hence increasing IT any further will only worsen the credibility of data.

In all the three conditions, the point with highest precision lies in the green zone. Shadow and room illumination conditions can widely vary which will shift the green and the red zones slightly. Still the flags serve as an indication. Fig. 10 and Fig. 11 shows some images of leaves in the three illumination condition. The three IT values are those at which bad pixels are noticeable in one of the illumination conditions. Fig. 10 (d) and Fig. 11 (d) show flags set at IT 900  $\mu$ s under room conditions for CamBoard and CamCube, respectively.

#### Comparison of Leaf with White Wall

Fig. 8, 9 column (iv) show the response of white wall at approximately 35-40 cm from the PMD cameras, aggregated in a 20x20 ROI, under room conditions. White wall is chosen in order to benchmark the ToF imaging for leaves as white wall is highly if not perfectly reflecting planar surface, is fairly lambertian and such surfaces are used for measuring the ToF imaging quality (Kahlmann and Re-

Table 3: IT for Saturation Thresholds

Sensor	Environment	IT ( $\mu$ s)
CamCube	Room	750
-do-	Shadow	600
-do-	Sunlight	500
CamBoard	Room	700
-do-	Shadow	400
-do-	Sunlight	100
SR4000	Room	800
-do-	Shadow (5KLux)	500
-do-	Sunlight	less than 300

mondino, 2006). We compare it to the best characteristics of the leaf images which is under room conditions (column (i)). In PMD cameras, wall characteristics are better than leaf as the amplitude (a) and (b) are more synchronized over the entire range of IT than the leaf. As already discussed in Sec. 3.1 leaves have high transmittance in NIR, this could possibly be the cause of a slight late saturation meanwhile the wall has stronger reflectance, therefore, an earlier saturation point. Other than that, the overall performance of the leaf under room conditions is quiet comparable to that of the white wall. This validates further the use of ToF imaging for leaf analysis.

#### 5.5. Camera Response Analysis: SwissRanger

SwissRanger cameras produce a confidence matrix instead of flags, which serves a similar purpose. High confidence for each pixel is desired in order to guarantee reliable data. The confidence is related to the precision.

Fig. 12 row (d) show the precision curve. Due to the high density of the data in PMD cameras (small step size:

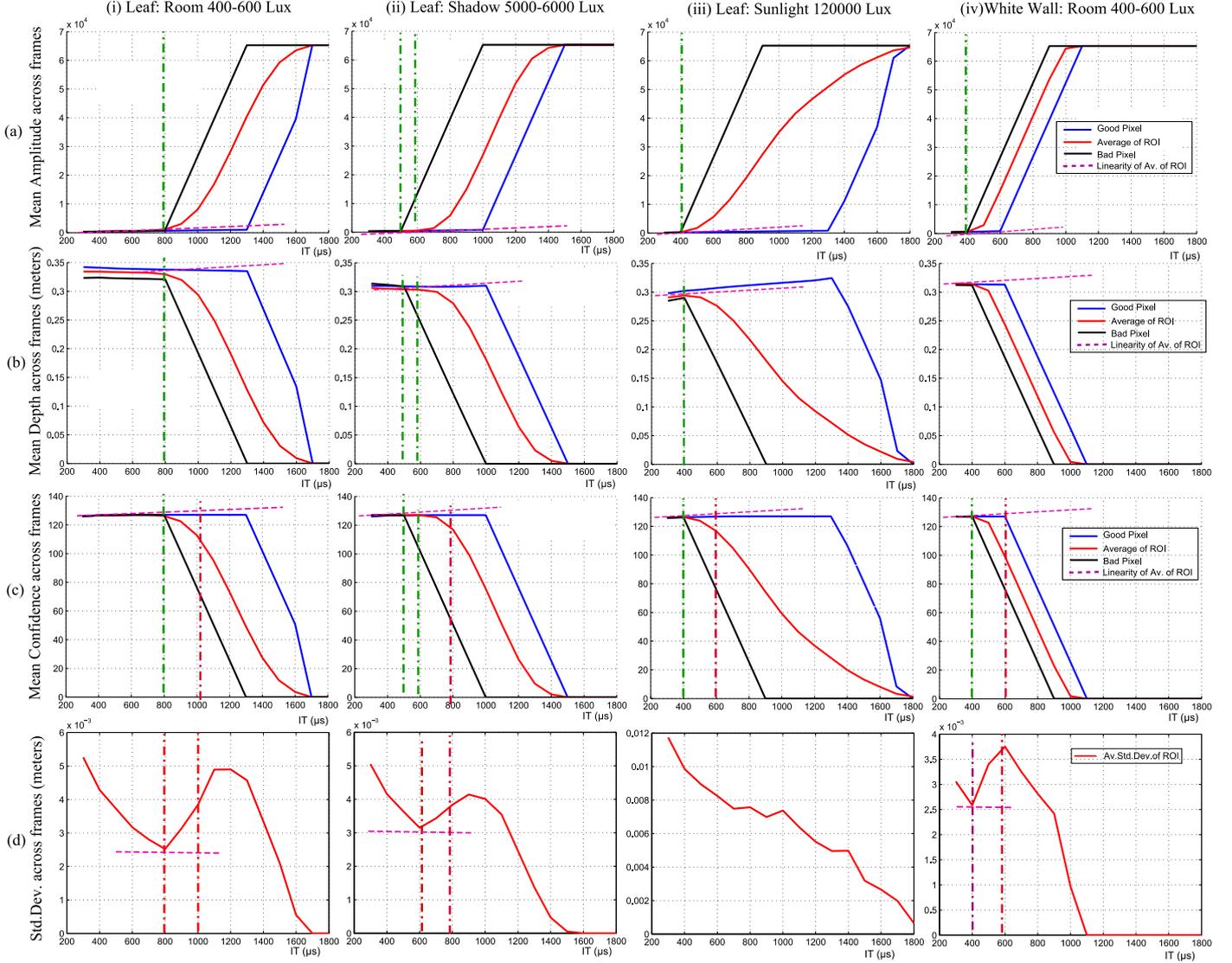


Figure 12: SR4000 response in amplitude, depth and precision at 25-40 cm in a 20x20 (ROI) under room, full shadow and sunlight conditions for one leaf in view and a white wall. Good and Bad pixels are among the last and first ones to show low confidence, respectively. Both pixels are inside the ROI. (d,iii) Observe the distinct trend in precision under sunlight due to rapidly increasing number of saturated pixels.

Table 1), a polynomial fit is used to display trends of mean standard deviation while for the SR4000, sparsity of the data produces abrupt and vivid changes. These abrupt changes make SR4000 data interpretation a lot easier than PMD, yet it is a drawback in terms of fine tuning of the camera parameters.

As can be seen from Fig. 12 (d), the first valley of mean standard deviation (precision) is quite obvious (dashed horizontal red line). The steep rise and subsequent sharp fall in precision reduces the green zone to almost a single IT in columns (i),(iii) and (iv) or a very narrow band as in column (ii) unlike a range of IT values in PMD cameras.

Matching curves in row (d) to row (a),(b) show that the fall in precision is a consequence of appearance of bad pixels. As soon as pixels' amplitude deviate from linearity, depth values start getting out of synchronization, increas-

ing the mean standard deviation of the ROI. The change in the pattern of the data is reflected in the corresponding confidence values (row (c)). The confidence stays high, as long as the mean amplitude maintains linearity. Bad pixels carry low confidence. Therefore, as soon as the amplitude linearity is disturbed, the mean confidence drops. The limiting value of confidence, in this case, is 120 units. It may be the case in general. Data corresponding to the confidence below this mark is unreliable under any illumination condition.

Fig. 12 column (iii) shows the performance of SR4000 under sunlight. Since the lowest possible IT in SR4000 is  $300 \mu s$  (see Table 1) which is still high for sunlight operation, the appearance of bad pixels is quite early and (d,iii) shows a very different trend in precision which is due to increasing number of pixels reaching saturation. Fig. 13

Table 4: Stereo Processing Times

Algorithm	Time (secs)
Point Grey Triclops	< 1
Correlation	3
Graph Cut	250 for 7 iterations

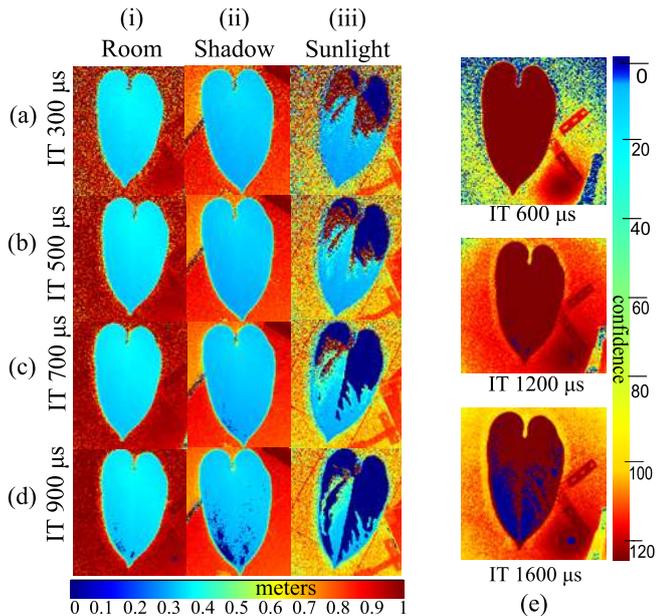


Figure 13: SR4000: (a,b,c,d) Depth images of plant leaves under the three different ambient illumination conditions. Pixels with low confidence appear darker. (e) Confidence images with varying IT under room conditions

shows SR4000 depth image under the three different ambient illuminations at four different ITs. The IT values are chosen so that low confidence in data is noticeable under one of the illumination conditions. Fig. 13 column (iii) shows that there is no IT value under sunlight at which a reliable depth image can be obtained with SR4000. Hence, this camera is not suitable for operation under sunlight; a fact clearly mentioned in the manufacturer’s documentation. Fig. 13 (e) shows confidence images of SR4000 at different IT’s.

#### Comparison of Leaf with White Wall

Fig. 12 column (iv) show the response of white wall under room conditions, at approximately 35 cm from the SR4000 camera, aggregated in a 20x20 ROI. Again, as discussed in Sec. 5.5, unlike PMD cameras, the sparsity of the data does not provide enough information of the gradual degradation of the data except the point of saturation or low confidence which appears earlier than the leaf under room conditions. But like PMD cameras, the overall performance of the leaf under room conditions and also under shadow is quiet comparable to that of the white wall. The early saturation is due to the wall being more reflective than the leaf.

#### 5.6. Stereo Depth Analysis

Stereo data acquisition is only varied by one parameter, i.e. exposure or Shutter Time (ST). Selected stereo algorithms have several parameters to tweak for best performance. Showing variation in disparity w.r.t each one is beyond the scope of this paper. Therefore, we chose

the best parameters as mentioned in Sec. 5.1.2 and varied the support region only for calculating matching cost due to which varying disparities were produced. Depth values were calculated by  $Z = fb/d$  where  $Z$  is the depth in meters,  $f$  is the focal length of the camera in pixels,  $b$  is the baseline (distance between the centers of the two cameras in meters) and  $d$  is the disparity measured in pixels.

Fig. 14 rows (a) and (b) show the mean depth and standard deviation, respectively, in a 20x20 pixels ROI on the depth image of the leaf. Here, the standard deviation represents error in the estimation of depth from stereo matching.

The leaf was at 80 cm. The retrieval of depth information seems similar in all the global and local algorithms. For instance, in case of room (a,i), all the curves stay around 80 cm throughout the ST range which means that the sensor did not saturate. Although, their accuracy varied. As Triclops is optimized for a particular hardware and real-time applications, it can only produce good results for a short range of ST between 20 ms to 100 ms. Due to in-built validation steps, it produced high standard deviation for higher values of ST which is due to holes appearing in the disparity image (Fig. 12(b,i)). This is because the data was being discarded as it did not satisfy the validation conditions. Non-Optimized version of local correlation performed better for the entire ST range with more accurate results for increasing support regions. This is also the case for GC based algorithm. But to complement this analysis, we must either obtain ground truth, which in case of plants is a difficult task (Nielsen et al., 2007) or else compare them by visual inspection. Fig. 18 shows depth images under room of all the algorithms used, for few selected STs and support regions sizes. The ST selection is based on patterns in the standard deviation curves from best to worst indicated in Fig. 14 row (b) with red markers on the ST axis.

These images reveal more about the quality of the stereo matching. Even though, correlation based algorithms retrieved depth with less noise (low std. dev.), specially for bigger support regions, the shape characteristics are much poorer. This is because the support regions act as averaging filter and with bigger windows, noise tends to reduce in the ROI, but at the same time, edges are blurred producing poor shape definition. GC on the other hand, appears to produce reasonably well shapes of leaves over a wide range of ST and with fairly accurate depth estimates.

In case of outdoor conditions, saturation does occur.

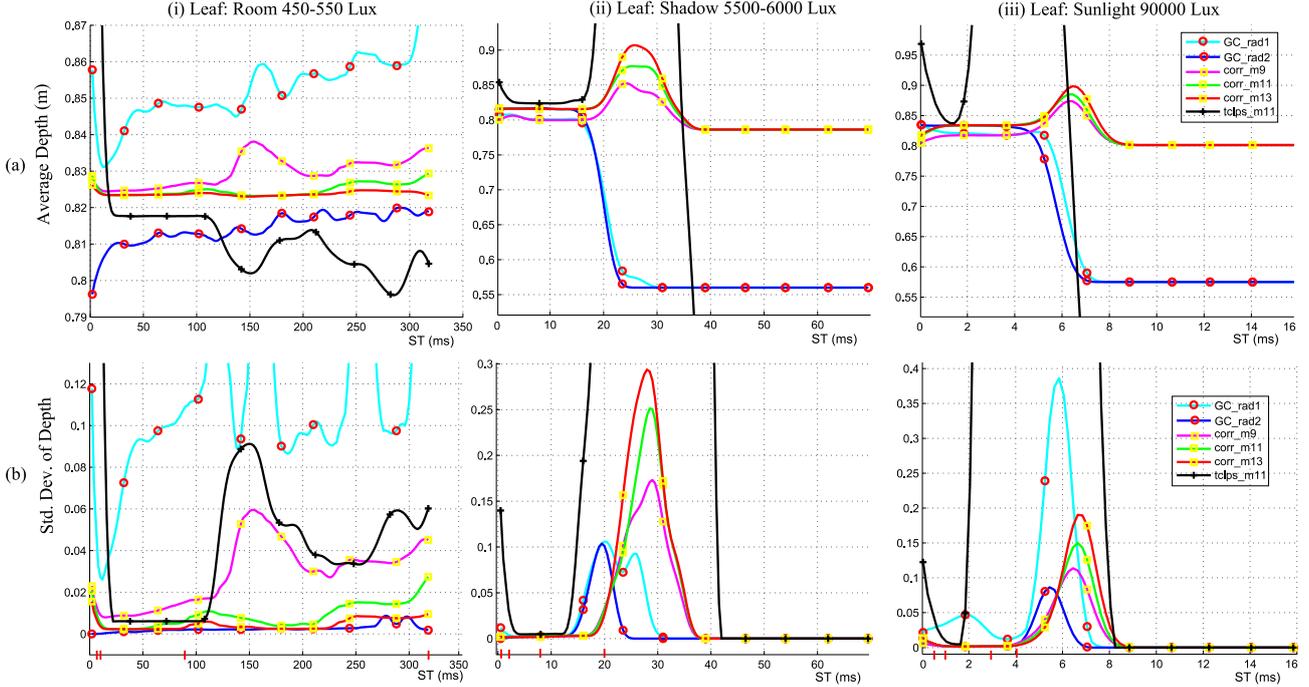


Figure 14: Stereo vision: Mean depth and Standard Deviation at 30-35 cm in a 20x20 (ROI) under room, full shadow and sunlight conditions for one leaf in view (Fig. 7 (a))

Looking at Fig. 14 (b,ii) and (b,iii), the ST values beyond which stereo matching is not able to retrieve any reliable depth information ends near 20 ms for shadow and 4 ms under sunlight. If exposure is increased beyond this, the sharp rise in standard deviation implies a quick loss of data which is due to sensor saturation. Selection of shutter times is hence important in different illumination environments. Fig. 19 and Fig. 20 show depth images for selected STs under shadow and sunlight.

Even below the saturation threshold, there is a slight variation in the performance as saturation is approached. Since the best depth image results w.r.t to shape are obtained with *GC* based stereo matching algorithm, so based on that, Fig. 18 (d), 19 (d) and 20 (d) show that the best ST values are 90 ms, 2.5 ms and 1 ms respectively. The penalty of operating beyond these values changes with the environment and is maximum in case of sunlight.

### 5.7. Imaging Under Mixed Illumination Conditions

In outdoor applications, it is highly likely to encounter a situation in which the leaf is partly under sunlight and partly under a shadow. The shadow in such a case is brighter than a complete shadow due to possible diffraction of sunlight which means that the saturation will occur sooner.

Fig. 15 and Fig. 16 row (a) show depth images of leaves under sunlight with some part under shadow of the camera tripod, for CamBoard and CamCube, respectively. Images at two different exposures are taken, one suitable for

shadow and other for sunlight. Leaf surface can be observed at different depths for lower ITs and partly saturated for higher ITs in both cases. This difference is more obvious in the corresponding intensity images (row (b)) especially with CamCube because it has more NIR emitters than CamBoard. Images at higher ITs cannot be used, for obvious reasons of saturation, but lower IT is not feasible either.

In order to increase the dynamic range of the ToF cameras, we exploit cameras flags and replace the depth values of all the pixels with Inconsistent/Saturated flags set at higher ITs with the corresponding values from lower ITs. The results are shown in Fig. 15 and Fig. 16 (a,iii) and (b,iii). The net effect of combining the exposures, both in depth and intensity is better than the original. This approach will require calibration for only two IT values and as discussed in Sec. 5.4 and we operate in the corresponding green zones (high precision) of the two ITs. Otherwise, we are either lower than green zone or well into the red zone of the two different ambient illumination settings.

For SR4000, confidence matrix can be used in place of flags for increasing the dynamic range. But as it is evident from Fig. 13 column (iii), SR4000 does not work under sunlight.

In case of stereo vision, although the best depth images w.r.t shape in Sec. 5.6 were obtained at ST 2.5 ms under shadow and 1 ms under sunlight, but the shadow in case of mixed lighting (as discussed earlier) is brighter than a complete shadow. Fig. 17 shows color and corresponding depth images obtained from *GC.rad2* at four different

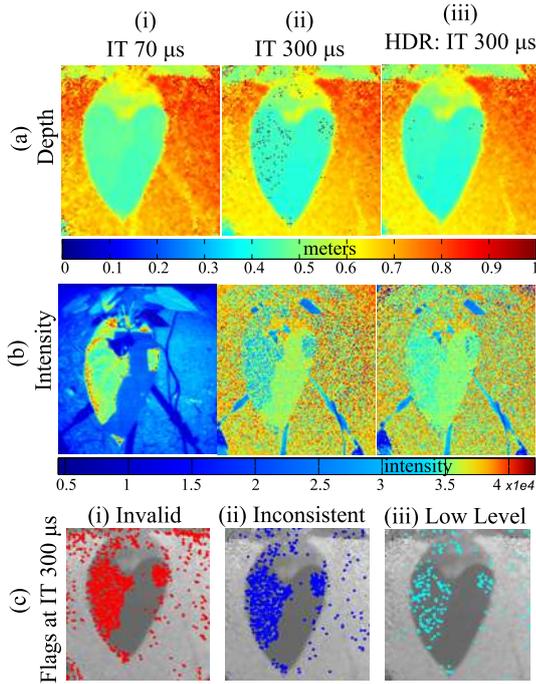


Figure 15: CamBoard images including both shadow and sunlight exposures. (a,iii) and (b,iii) show High Dynamic Range (HDR) images after combining two different exposures (c) Flags set at IT 300  $\mu$ s

exposures. Owing to this brighter shadow, a good compromise is reached at ST 1.45 ms. With higher STs, edges of the leaf are being erased due to high specular reflections on the section under sunlight (column (iv)).

Even though the dynamic range of stereo vision cameras seems sufficient for the task at hand, still, higher dynamic range can be obtained by preprocessing both the left and right images using standard methods such as combining images with varying exposures (Mann and Picard, 1995) before applying stereo matching. Of course, this will further add to the efficiency constraints of the stereo vision.

## 6. Discussion

Analysis done in Secs. 5.4, 5.5 and 5.6 addressed the response of ToF and stereo vision independently. For stereo vision, Fig. 14 show that, except for subtle variations, retrieval of depth information is almost same for any algorithm, whether local or global, unless further validation steps are involved such as in Triclops. Stereo vision can work on a wide range of exposures while ToF has a very limited exposure under outdoor conditions or may not even work at all, such as SR4000. Among the ToF cameras used in this work, PMD CamCube has better cancellation of sunlight (Table 3) followed by Camboard while the SNR of SR4000 renders it useless under sun. But in any case, ToF sensors are generally more sensitive to sunlight than stereo vision.

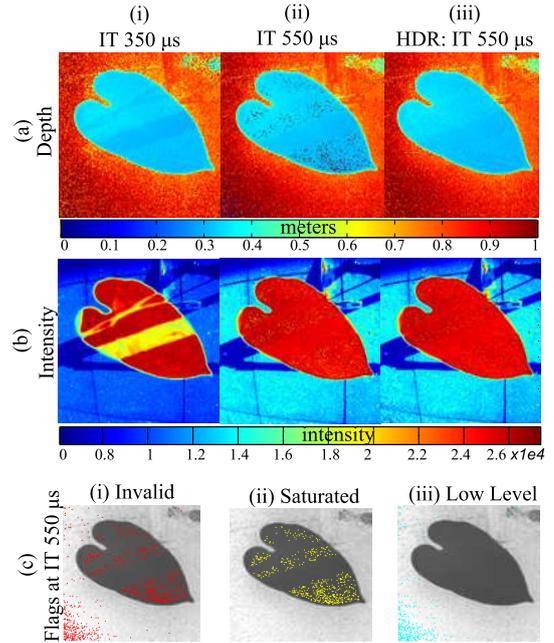


Figure 16: CamCube images including both shadow and sunlight exposures. (a,iii) and (b,iii) show High Dynamic Range (HDR) images after combining two different exposures. (c) Flags set at IT 550  $\mu$ s

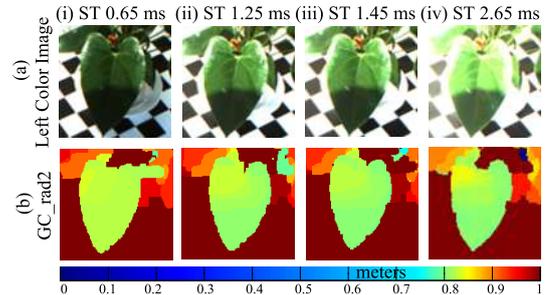


Figure 17: Depth images obtained through stereo correspondence for scene involving both shadow and sunlight exposures (GC Only)

Stereo vision, on the other hand, has its own draw backs. Visual inspection of the depth images from both the Time of Flight (ToF) and the stereo vision show that better shape estimation can only be achieved with global matching algorithms such as GC which still is not as good as ToF depth images. Algorithms using local correlation may provide sufficient depth information for tasks such as leaf angle estimation and leaf motion tracking (Biskup et al., 2007), but accurate shape retrieval which is crucial for numerous other applications (such as 3D model creation or plant specie recognition using leaf shapes) the local algorithms are not suitable. Global stereo matching such as GC, are usually very time consuming (Table 4). Over the years, several other stereo matching algorithms have appeared and older ones have evolved both in terms of accu-

Table 5: Leaf surface characteristics of the validation flower plants (based on inspection)

	Thickness	Roughness
<b>Low</b>	Pelargonium	Anthurium
<b>Medium</b>	Hydrangea	Hydrangea Cyclamen Orchidaceae
<b>High</b>	Orchidaceae Cyclamen Anthurium	Pelargonium

racy and efficiency<sup>5</sup>. Still, an accurate algorithm with an efficient implementation over a specialized hardware cannot make it comparable to ToF (30-50 fps).

### Analysis of Validation Data

In order to further validate these findings, we tested four more flower families, namely Cyclamen, Hydrangea, Orchidaceae and Pelargonium. Together, they represent a wide range of leaf characteristics varying in thickness, shape, surface roughness and texture as shown in Table 5. As for the ToF cameras, we have already shown that the performance characteristics are related to amplitude non-linearity (Secs. 5.4 and 5.5), so we only evaluated the amplitude characteristics under room conditions at 30-40 cm range for this validation set (Fig. 21). The amplitude curves clearly show that the given set of variations in leaf characteristics have a minimal effect on the response of ToF sensors. The subtle variations are contributed by changes in relative camera to leaf distance and leaf thickness. PMD CamCube appears to be least affected by such variations.

Fig. 22 shows color and depth images retrieved from both ToF and stereo vision sensors. Stereo data was obtained under room conditions at 80 ms and ToF images at IT 400  $\mu$ s. The exposures are chosen to be the ones with best performance under room conditions following the analysis from Secs. 5.4, 5.5 and 5.6. Again, the results validate the findings since the best shape retrieval can be observed among ToF cameras with PMD CamCube and SR4000 providing the best results followed by PMD CamBoard. Among the stereo matching algorithms, again GC has better shape retrieval but still not as good as ToF sensors. The local correlation algorithms do not perform well in this regard as well.

## 7. Conclusion

We have performed a detailed analysis and comparison of the response of ToF and stereo vision for close range leaf imaging under room, shadow and sunlight conditions. In case of ToF, the comparison with flat white wall indicates the suitability of using ToF imaging for agricultural applications. Since ToF cameras are sensitive to ambient light,

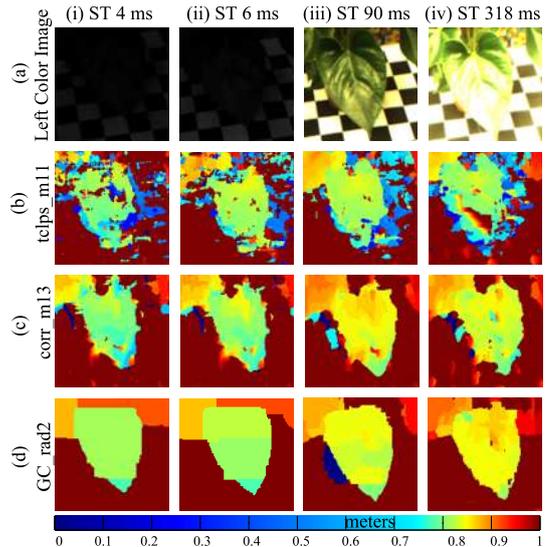


Figure 18: Depth images acquired through stereo vision under room conditions. ST values are chosen from the trends in the std.dev. (see Fig. 14 (b,i) red markers on horizontal axis)

we have proposed a method to detect suitable IT for the three commonly faced conditions using the appearance of inconsistent/saturation flags or confidence matrices. This scheme can be extended to any ambient illumination setting and object in general. Choosing a specific IT for a given condition allows higher accuracy through optimal calibration for that IT which adds to the value of imaging in close range tasks. Stereo vision, although, is relatively more robust to outdoor illumination than ToF, suffers due to the correspondence problems along with the efficiency bottlenecks which make ToF more preferable, even under outdoor conditions.

The advantage of stereo vision is its high resolution output which is unmatched by any ToF camera so far (ToF max. resolution 200x200). This is the low ToF resolution which demands such close range observation of plant organs (less than half meter), otherwise, SNR of ToF signal deteriorates. Given the rapidly evolving technology, it seems more likely that ToF sensors will bridge the gap and come up with more robust background illumination rejection and higher resolution than stereo vision research producing highly accurate and efficient algorithms.

Generally, all stereo cameras and some ToF cameras (e.g. SR4000) have an option for selecting the exposure automatically. But such automatic selection can compromise precision. The analysis presented in this article was established on one type of plant and validated on four other plant families. So it can be safely concluded that ToF imaging is suited for leaf imaging in general. While the camera settings may change slightly, this article provides a method to find optimal camera parameters for any plant (object in general), ambient illumination and distance.

<sup>5</sup><http://vision.middlebury.edu/stereo/eval/>

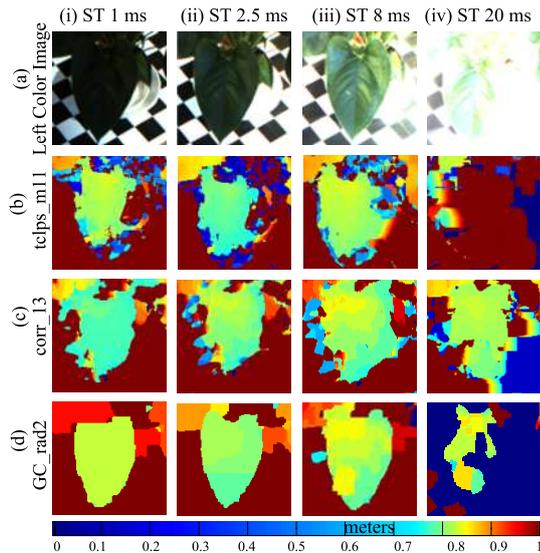


Figure 19: Depth images acquired through stereo vision under shadow conditions. ST values are chosen from the trends in the std.dev. (Fig. 14 (b,ii) red markers on horizontal axis)

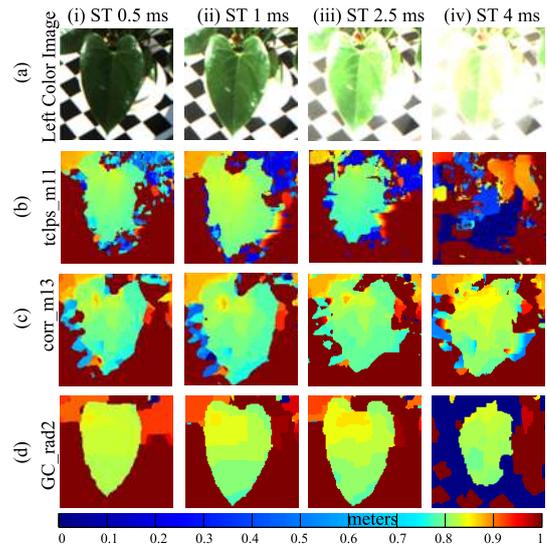


Figure 20: Depth images acquired through stereo vision under sunlight conditions. ST values are chosen from the trends in the std.dev. (Fig. 14 (b,iii) red markers on horizontal axis)

## Acknowledgments

This work is supported by the Danish Council for Strategic Research under project ASETA (www.aseta.dk) grant no. 09-067027, the Spanish Ministry of Science and Innovation under projects PAU + DPI2011-27510, the EU project GARNICS (FP7-247947) and the Catalan Research Commission (SGR-00155). Wajahat Kazmi is funded by ASETA and Sergi Foix is supported by PhD fellowship from CSIC's JAE program. Authors want to thank PMDTec and Ilona Schwengber for their support and insightful information about CamBoard and Cam-Cube.

## References

Alenya, G., Dellen, B., Foix, S., Torras, C., 2013. Robotized Plant Probing: Leaf Segmentation Utilizing Time-of-Flight Data. *Robotics Automation Magazine, IEEE* 20 (3), 50–59.

Alenya, G., Dellen, B., Torras, C., 2011. 3D modelling of leaves from color and ToF data for robotized plant measuring. In: *IEEE International Conference on Robotics and Automation*. Shanghai, China, pp. 3408–3414.

Andersen, H. J., Reng, L., Kirk, K., Nov. 2005. Geometric plant properties by relaxed stereo vision using simulated annealing. *Computers and Electronics in Agriculture* 49 (2), 219–232.

Astrand, B., Baerveldt, A., 2004. Plant recognition and localization using context information. In: *Proc. of the IEEE Conference Mechatronics and Robotics*. Luoyang, pp. 13–15.

Biskup, B., Scharr, H., Schurr, U., Rascher, U., Oct. 2007. A stereo imaging system for measuring structural parameters of plant canopies. *Plant, cell & environment* 30 (10), 1299–308.

Blas, M. R., 2010. Fault-Tolerant Vision for Vehicle Guidance in Agriculture Fault-Tolerant Vision for Vehicle Guidance in Agriculture. Ph.D. thesis, Denmark Technical University, Lyngby.

Brakke, T. W., 1992. Goniometric measurements of light scattered. In: *12th Annual International Geoscience and Remote Sensing Symposium (IGARSS)*. pp. 508–510.

Chapron, M., Ivanov, N., Boissard, P., Valery, P., 1993. Visualization of corn acquired from stereovision. In: *Proceedings of IEEE Systems Man and Cybernetics Conference*. Vol. 5. pp. 334–338.

Chelle, M., Sep. 2006. Could plant leaves be treated as Lambertian surfaces in dense crop canopies to estimate light absorption? *Ecological Modelling* 198 (1-2), 219–228.

Feret, J.-B., François, C., Asner, G. P., Gitelson, A. a., Martin, R. E., Bidet, L. P., Ustin, S. L., le Maire, G., Jacquemoud, S., Jun. 2008. PROSPECT-4 and 5: Advances in the leaf optical properties model separating photosynthetic pigments. *Remote Sensing of Environment* 112 (6), 3030–3043.

Fiorani, F., Rascher, U., Jahnke, S., Schurr, U., Apr. 2012. Imaging plants dynamics in heterogenic environments. *Current opinion in biotechnology* 23 (2), 227–35.

Foix, S., Alenya, G., Torras, C., 2011. Lock-in time-of-flight (ToF) cameras: a survey. *Sensors Journal, IEEE* 11 (9), 1917–1926.

Frasson, R. P. D. M., Krajewski, W. F., 2010. Three-dimensional digital model of a maize plant. *Agricultural and Forest Meteorology* 150 (3), 478–488.

Garcia, R., Francisco, J., Sankaran, S., Maja, J., Feb. 2013. Comparison of two aerial imaging platforms for identification of Huanglongbing-infected citrus trees. *Computers and Electronics in Agriculture* 91, 106–115.

Gates, D., Keegan, H., Schleiter, J., Jan. 1965. Spectral properties of plants. *Applied optics* 4 (1), 11.

Govaerts, Y., Jacquemoud, S., Nov. 1996. Three-dimensional radiation transfer modeling in a dicotyledon leaf. *Optical Society of America* 35 (33), 6585–6598.

Haller, I., Nedeveschi, S., Feb. 2012. Design of Interpolation Functions for Subpixel-Accuracy Stereo-Vision Systems. *IEEE Transactions on Image Processing* 21 (2), 889–98.

Ivanov, N., Boissard, P., Chapron, M., Andrieu, B., 1995. Computer stereo plotting for 3-D reconstruction of a maize canopy. *Agricultural and Forest Meteorology* 75 (13), 85–102.

Jacquemoud, S., Baret, F., Nov. 1990. PROSPECT: A model of leaf optical properties spectra. *Remote sensing of environment* 34 (2), 75–91.

Jimenez, A., Ceres, R., Pons, J., 2000. A survey of computer vision methods for locating fruit on trees. *Transactions of the ASAE-American Society of Agricultural Engineers* 43 (6), 1911–1920.

Jin, J., Tang, L., 2009. Corn plant sensing using real-time stereo vision. *Journal of Field Robotics* 26 (6-7), 591–608.

Kahlmann, T., Remondino, F., 2006. Calibration for increased accu-

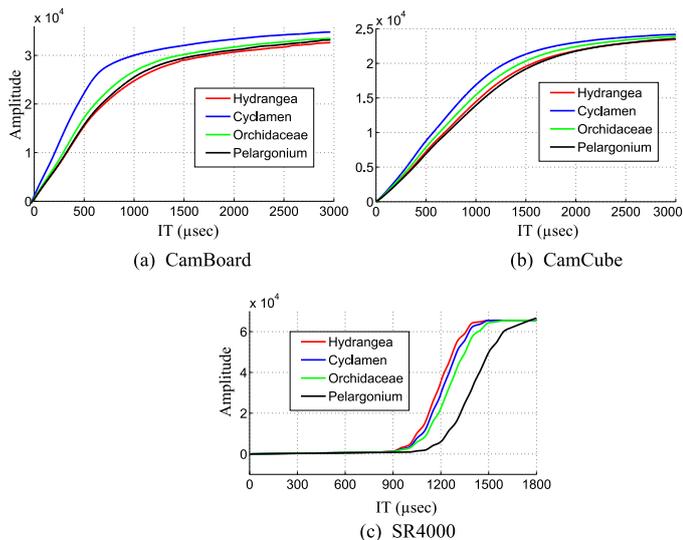


Figure 21: TOF amplitude characteristics of the four flower families comprising the validation set

racy of the range imaging camera SwissRanger. In: ISPRS Commission V Symposium. No. 4. Dresden, Germany, pp. 136–141.

Kazmi, W., Bisgaard, M., Garcia-Ruiz, F., Hansen, K., la Cour-Harbo, A., 2010. Adaptive Surveying and Early Treatment of Crops with a Team of Autonomous Vehicles. In: European Conference on Mobile Robots. Orebro, Sweden, pp. 253–258.

Kazmi, W., Foix, S., Alenya, G., Nov. 2012. Plant leaf imaging using time of flight camera under sunlight, shadow and room conditions. In: Proceedings of International Symposium on Robotic and Sensors Environments. IEEE, Magdeburg, Germany, pp. 192–197.

Kise, M., Zhang, Q., Oct. 2008. Development of a stereovision sensing system for 3D crop row structure mapping and tractor guidance. *Biosystems Engineering* 101 (2), 191–198.

Klose, R., Penlington, J., Ruckelshausen, A., 2009. Usability study of 3d time-of-flight cameras for automatic plant phenotyping. *Bornimer Agrartechnische Berichte* 69, 93–105.

Kolmogorov, V., Zabih, R., 2002. Multi-camera Scene Reconstruction via Graph Cuts. In: Proceedings of the 7th European Conference on Computer Vision-Part III. ECCV '02. Springer-Verlag, London, UK, pp. 82–96.

Kraft, H., Frey, J., Moeller, T., Albrecht, M., Grothof, M., Schink, B., Hess, H., Buxbaum, B., 2004. Elimination Based on Improved PMD (Photonic Mixer Device)-Technologies. In: OPTO, AMA Fachverband. Nürnberg.

Kraft, M., Regina, N., Freitas, S. a. D., Munack, A., 2010. Test of a 3D Time of Flight Camera for Shape Measurements of Plants. In: CIGR Workshop on Image Analysis in Agriculture. No. August. Budapest.

Kytö, M., Nuutinen, M., Oittinen, P., Jan. 2011. Method for measuring stereo camera depth accuracy based on stereoscopic vision. *SPIE/IS&T Electronic Imaging: Three-Dimensional Imaging, Interaction and Measurement* 7864, 1–9.

Lange, R., 2000. 3D time-of-flight distance measurement with custom solid-state image sensors in CMOS/CCD-technology. Ph.d. dissertation, University of Siegen.

Lange, R., Seitz, P., 2001. Solid-state time-of-flight range camera. *Quantum Electronics, IEEE Journal of* 37 (3), 390–397.

Mann, S., Picard, R. W., 1995. On being 'undigital' with digital cameras : Extending Dynamic Range by Combining Differently Exposed Pictures. *Science* (323), 422–428.

May, S., Werner, B., Surmann, H., 2006. 3D time-of-flight cameras for mobile robotics. In: IEEE International Conference on Intelligent Robots and Systems. Beijing, China, pp. 790–795.

McCarthy, C., 2009. Automatic non-destructive dimensional mea-

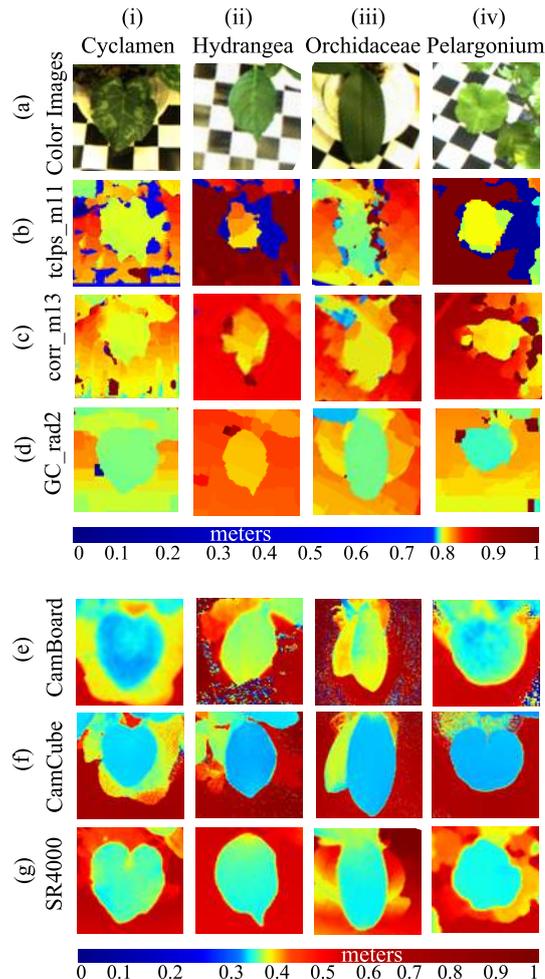


Figure 22: ToF (IT 400  $\mu$ s) and Stereo (ST 80 ms) depth images of the four flower families comprising the validation set

surement of cotton plants in real-time by machine vision. Ph.D. thesis, University of Southern Queensland.

Mizuno, S., Noda, K., Ezaki, N., Takizawa, H., Yamamoto, S., 2007. Detection of Wilt by Analyzing Color and Stereo Vision Data of Plant. In: Gagalowicz, A., Philips, W. (Eds.), *Computer Vision/Computer Graphics Collaboration Techniques*. Vol. 4418 of Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 400–411.

Möller, T., Kraft, H., Frey, J., Albrecht, M., Lange, R., 2005. Robust 3d measurement with pmd sensors. In: Hilmar, I., Timo, K. (Eds.), *Robust 3D Measurement with PMD Sensors, Proceedings of the 1st Range Imaging Research Day at ETH Zurich*. No. Section 5. Zurich, Switzerland.

Mufti, F., Mahony, R., Sep. 2011. Statistical analysis of signal measurement in time-of-flight cameras. *ISPRS Journal of Photogrammetry and Remote Sensing* 66 (5), 720–731.

Mutto, C. D., Guido, P. Z., Cortelazzo, M., Mar. 2012. Time-of-Flight Cameras and Microsoft Kinect. In: *SpringerBriefs in Electrical and Computer Engineering*. Springer, Ch. 2, p. 21.

Nakarmi, A. D., Tang, L., 2010. Inter-plant Spacing Sensing at Early Growth Stages Using a Time-of-Flight of Light Based 3D Vision Sensor. In: ASABE Paper No. 1009216. St. Joseph, Michigan, pp. 1–15.

Nayar, S., Mitsunaga, T., 2000. High dynamic range imaging: spatially varying pixel exposures. In: *Proceedings of Conference on Computer Vision and Pattern Recognition*. Vol. 1. IEEE, pp. 472–

- Nielsen, M., Andersen, H., Granum, E., 2005. Comparative study of disparity estimations with multi-camera configurations in relation to descriptive parameters of complex biological objects. In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*. Beijing, China.
- Nielsen, M., Andersen, H. J., Slaughter, D. C., Granum, E., Jan. 2007. Ground truth evaluation of computer vision based 3D reconstruction of synthesized and real plant images. *Precision Agriculture* 8 (1-2), 49–62.
- Nielsen, M., Andersen, H. J. r., Slaughter, D. C., Giles, D. K., 2004. Detecting leaf features for automatic weed control using trinocular stereo vision. Minneapolis, USA.
- Nuske, S., Achar, S., Bates, T., Narasimhan, S., Singh, S., Sep. 2011. Yield estimation in vineyards by visual grape detection. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems*. pp. 2352–2358.
- Piron, a., Heijden, F. V. D., Destain, M., Nov. 2011. Weed detection in 3D images. *Journal of Precision agriculture* 12 (5), 607–622.
- Point Grey, R., 2012. Stereo Accuracy and Error Modeling.
- Preuksakarn, C., Boudon, F., 2010. Reconstructing plant architecture from 3D laser scanner data. In: *6th International Workshop on Functional-Structural Plant Models*. pp. 1999–2001.
- Rovira-Más, F., Zhang, Q., Reid, J. F., 2005. Creation of Three-dimensional Crop Maps based on Aerial Stereoimages. *Biosystems Engineering* 90 (3), 251–259.
- Santos, T. T., Oliveira, A. A., Aug. 2012. Image-based 3D digitizing for plant architecture analysis and phenotyping. In: *Saúde, A. V., Guimarães, S. J. F. (Eds.), Workshop on Industry Applications (WGARI) in SIBGRAPI 2012 (XXV Conference on Graphics, Patterns and Images)*. Ouro Preto, MG, Brazil.
- Scharstein, D., Szeliski, R., Apr. 2002. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. *International Journal of Computer Vision* 47 (1-3), 7–42.
- Slaughter, D., Giles, D., 2008. Autonomous robotic weed control systems: a review. *Computers and Electronics in Agriculture* 1 (issue: 2002), 63–78.
- Song, Y., Glasbey, C., van der Heijden, G., 2011. Combining stereo and Time-of-Flight images with application to automatic plant phenotyping. In: *Proceedings of the 17th Scandinavian Conference on Image Analysis (SCIA 2011)*. Vol. 1. Lecture Notes in Computer Science, Springer Verlag, Berlin, pp. 467–478.
- Takizawa, H., Ezaki, N., Mizuno, S., Yamamoto, S., 2005. Plant Recognition by Integrating Color and Range Data Obtained Through Stereo Vision. *Journal of Advanced Computational Intelligence and Intelligent Informatics* 9 (6), 630–636.
- Tang, L., Tian, L. F. L., 2008. Real-time crop row image reconstruction for automatic emerged corn plant spacing measurement. *Transactions of The ASABE* 51 (3), 1079–1087.
- van der Heijden, G., Song, Y., Horgan, G., Polder, G., Dieleman, A., Bink, M., Palloix, A., van Eeuwijk, F., Glasbey, C., 2012. SPICY: towards automated phenotyping of large pepper plants in the greenhouse. *Functional Plant Biology* 39 (11), 870–877.
- Van der Zande, D., Stuckens, J., Verstraeten, W. W., Muys, B., Coppin, P., Jun. 2010. Assessment of Light Environment Variability in Broadleaved Forest Canopies Using Terrestrial Laser Scanning. *Remote Sensing* 2 (6), 1564–1574.
- van Henten, E., Marx, G., Hofstee, J., Hemming, J., Sarlikioti, 2011. Measuring Leaf Motion of Tomato by Machine Vision. In: *International Symposium on Advanced Technologies and Management Towards Sustainable Greenhouse Ecosystems: Greensys, ISHS Acta Horticulturæ* 952.
- Woolley, J. T., May 1971. Reflectance and transmittance of light by leaves. *Plant physiology* 47 (5), 656–62.
- Yoon, S., Thai, C., 2009. Stereo Spectral Imaging System for Plant Health Characterization. In: *ASAE, Annual international meeting*.