Design and calibration of multi-camera systems for 3D computer vision: lessons learnt from two case studies

Tom Botterill[1,2], Matthew Signal[2], Steven Mills[3], Richard Green[1]

[1]Department of Computer Science, University of Canterbury, Christchurch, NZ
[2]Tiro Lifesciences, Christchurch, NZ
[3]Department of Computer Science, University of Otago, Dunedin, NZ

Abstract. This paper examines how the design of imaging hardware for multi-view 3D reconstruction affects the performance and complexity of the computer vision system as a whole. We examine two such systems: a grape vine pruning robot (a 4.5 year/20 man-year project), and a breast cancer screening device (a 10 year/25 man-year project). In both cases, mistakes in the initial imaging hardware design greatly increased the overall development time and cost by making the computer vision unnecessarily challenging, and by requiring the hardware to be redesigned and rebuilt. In this paper we analyse the mistakes made, and the successes experienced on subsequent hardware iterations. We summarise the lessons learned about platform design, camera setup, lighting, and calibration, so that this knowledge can help subsequent projects to succeed.

Keywords: Multi-view reconstruction, 3D reconstruction, camera hardware, camera calibration, lighting

1 Introduction

Computer vision forms an integral part of ever more complex systems, including robot systems, medical imaging systems, smart vehicle systems, and 3D motion capture systems. Each system requires imaging hardware including cameras, lights, and often enclosures to control imaging conditions. This imaging hardware is frequently assembled specifically for the application [1, 19, 23, 11]. In this paper we argue that careful design of the imaging hardware greatly reduces the overall development effort required, hence increasing the likelihood of success and improving overall performance. Computer vision systems chain together many different processes, from low level segmentation and feature extraction, through to high level model fitting, with these models ultimately used to make decisions, e.g. on robot controls, diagnoses. Errors in the imaging process, or limitations of the imaging process, propagate through the different computer vision algorithms and affect the performance and accuracy of the system as a whole. In our experience, much development effort is spent on compensating for mistakes made

when designing the systems and collecting data, and more careful design would simplify and speed up the development process, while improving performance overall.

This paper focusses on the effects of hardware choices on multi-camera systems for 3D reconstruction. Multi-camera systems are popular for applications requiring high-accuracy, high resolution 3D reconstructions. Many design considerations are just as relevant for systems using depth cameras, where similar challenges with lighting[1], image resolution and image quality [8] exist.

The paper is organised as follows: Section 2 summarises how the image acquisition process affects the performance of the computer vision system, and Section 3 describes two case studies that we use to illustrate these effects. Section 4 reviews the choice of cameras and lenses, camera positioning, illumination, enclosure design and calibration procedures. The impact of each decision, trade-offs required, and lessons learnt from the two case studies are discussed. Recommendations are also summarised in the checklist that is provided as supplementary material, and from `http://hilandtom.com/PSIVT2015-Checklist.pdf`.

2 Effects of imaging quality on computer vision

The imaging hardware affects a 3D computer vision system's performance in four ways: accuracy, robustness, development cost, and efficiency.

To most straightforward effect of image quality on computer vision systems is on the accuracy of measurements taken from images, e.g. localisation accuracy or the accuracy of a 3D measurement. This is the case for many computer vision methods, including those formulated as a data-plus-spatial energy minimisation (e.g. active shape/contour models, dense stereo, dense optical flow [22]). Improving resolution, focus, pixel signal-to-noise ratios, etc. are straightforward ways to minimise different kinds of image noise [10, 19], which enables more weight to be given to the data terms, hence increasing accuracy. By averaging across many pixels, computer vision systems often achieve subpixel or sub-greylevel accuracy [19, 2].

A much more challenging class of errors in computer vision are the large discrete errors known as gross errors. These errors are generally too large to average away, and can cause partial or total system failure. Gross errors that create challenges for multi-view 3D computer vision systems include incorrectly detecting and/or matching features, segmentation errors, and errors at depth discontinuities. These errors are prevalent when objects are occluded, partly outside the camera's field of view, or when their appearance is affected by variable lighting, reflections, and shadows. A system's susceptibility to these errors is referred to as its robustness.

An important effect of imaging conditions on computer vision is on development time. When imaging conditions are poor, considerable development time must be spent on modelling shadows and lighting effects [12], and handling the

[1] `https://support.xbox.com/en-GB/xbox-360/kinect/lighting`

matching ambiguities and increased outlier rates that result (see, for example, the vast literature on making RANSAC-based robust matching frameworks perform well [9]). In addition, if imaging hardware is redesigned, time consuming code changes may be required throughout the entire system [19], as the errors present and the visual effects to model change.

The imaging process also affects computational efficiency: pixel-level algorithms (e.g. segmentation, dense optical flow or dense stereo) may be slower for higher resolution images, however as soon as a higher-level representation is obtained (e.g. features are extracted) then the resolution no longer affects computation times, and higher quality images may improve performance, e.g. by increasing feature localisation accuracy or by reducing matching ambiguities. Even pixel-level algorithms may be no slower for higher resolution images if fewer iterations are required. Later-stage 3D algorithms can be considerably more efficient when there are fewer outliers: when outlier rates are low, fewer iterations of RANSAC are needed [9] and efficient non-robust quadratic cost functions can be used in bundle adjustment [3]. In our experience, and as is often the case in software [13, Section 25.2], the biggest increases in efficiency come from having more development time available for optimising and parametrising algorithms once the rest of the system is working, and once critical loops are identified. Improving imaging hardware improves computational efficiency by making the development process more efficient.

3 Case studies

In this paper we use two case studies to illustrate the effects of hardware design on system performance and development. The first is a grape vine pruning robot, and the second is a prototype breast cancer screening system. Both use synchronised cameras to image their subject, and use customised imaging enclosures and artificial lighting to control imaging conditions. Both imaging hardware systems have been completely redesigned and rebuilt, at considerable expense, as the importance of the hardware design has become apparent.

3.1 Grape vine pruning robot

The first case study is a grape vine pruning robot [5]. Grape vines are pruned by selectively cutting canes on each plant. The robot system, shown in Figure 1, consists of a mobile platform which straddles a row of vines, and images them with a trinocular stereo camera rig as the platform moves. A computer vision system builds a 3D model of the vines, an AI system decides which canes to prune, and a six degree-of-freedom robot arm makes the required cuts. The main challenge for the computer vision is building a sufficiently complete and structurally correct 3D model of the vines that the AI can make good decisions about where to cut, and so that a path planner can plan a collision-free path for the robot arm to make the required cuts. The project started in 2010, and has employed up-to five full time researchers and developers (including graduate students). 27 people have worked on the project in total.

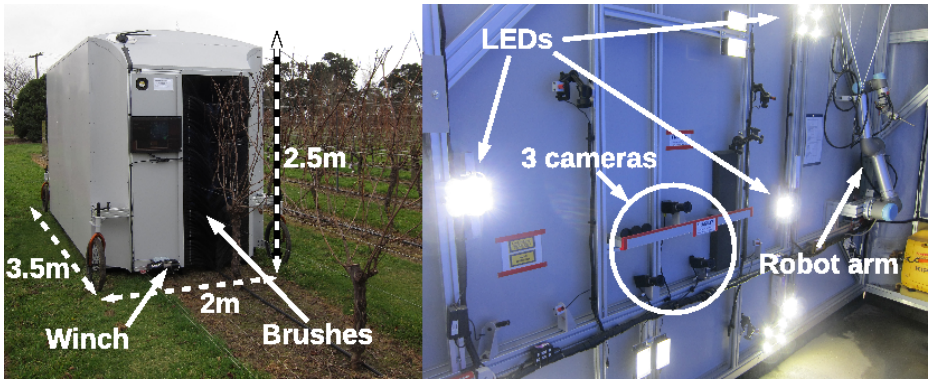


Fig. 1. The pruner robot's mobile platform completely covers the vines, blocking out sunlight (left). Inside are high-powered LEDs, three cameras, a robot arm, a generator and the desktop PC that runs all of the software (right).

3.2 DIET machine

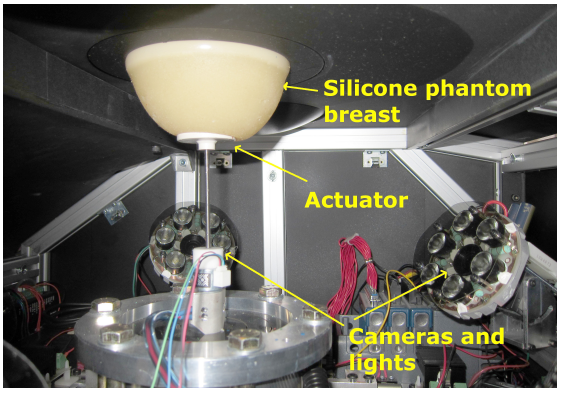


Fig. 2. The DIET machine, showing two of the five cameras, the actuator, and a silicone phantom breast. The machine measures 80cm by 71cm by 38cm.

The second case study is the Digital Image-based Elasto-Tomography (DIET) system, a prototype breast cancer screening system [2]. A breast is imaged by five cameras while being vibrated, the computer vision system estimates the 3D surface motion, and the observed surface motion is used to infer the internal stiffness of the breast, hence identifying tumours. The computer vision system first identifies the profile of the breast in each image, and reconstructs a 3D surface model from these profiles. The surface motion is measured using dense optical flow, then the 3D surface motion is given by fusing the optical flow with

the reconstructed surfaces. The current DIET machine is shown in Figure 2. The project started in 2005 and has employed up-to five full time researchers and developers.

4 Designing a multi-camera system for 3D imaging

This section examines each design decision made when building a multi-camera system for 3D reconstruction. The case studies are used to illustrate how each decision affects the performance of the system as a whole.

4.1 Imaging enclosure design

Imaging enclosures allow lighting levels to be controlled, and provide a uniform background. If lighting levels vary too much, the cameras cannot simultaneously image the brightest parts of the scene (where the sensor is saturated) and the darkest parts (where details are lost in sensor noise). Uniform-coloured backgrounds aid the foreground/background segmentation—the greater the difference between the distributions of colours on the foreground and background, the simpler, and hence more robust, the segmentation will be.

Designing imaging enclosures for 3D imaging is hard because 3D objects cast shadows, different parts are at different distances and angles to different lights (which affects their appearance from different viewpoints), and because multiple overlapping images from different viewpoints are required to give a complete 3D reconstruction.

Canopies are often used by agricultural robots to shade direct sunlight, or to completely control illumination [20, 18]. For outdoor applications where it is hard to control lighting, many robots operate only at night, to avoid interference from sunlight, and where active illumination ensures that only subjects close to the light source are illuminated [15, 21, 8].

The pruner robot's canopy consists of a rigid MayTec[2] aluminium frame with sheet aluminium cladding. Sunlight is excluded with brushes. The inside is lined with corflute corrugated plastic, then covered with photo studio non-reflective chroma-key blue backdrop paper[3]. The background provides a seamless matte blue background behind the vines. The problem with this design is that the background is not sufficiently rigid: sagging causes dark shadows which are detected as vines (especially if background subtraction-based methods are used for segmentation), and wrinkles in the cardboard have similar scale and appearance to wires. These artefacts increase the number of incorrectly detected vines, and increase the levels of robustness required throughout the computer vision software. In addition, the background is fragile and prone to damage, rendering the entire system is unusable. A more robust design would use a rigid backing.

[2] http://www.maytec.org/

[3] `http://savageuniversal.com/products/seamless-paper/studio-blue-seamless-paper`

The first pruner robot did not use brushes, and was unusable during daylight, as sunlight shone on both the vines and background. The current design can be used in all weather, however small shafts of sunlight get through gaps in the brushes (Figure 7). These saturated regions are detected and masked out before foreground/background segmentation.

The first DIET machine used a shiny black perspex background. The segmentation was unreliable, as the measured colour of specular reflections off the background was often the same as the breast, because the boundary between the machine and breast was in shadow, and because seams close to the breast edge were detected instead of the breast edge. The current machine uses a matte black perspex background and adds a marker to the actuator (a black and white circle). Together with lighting improvements, (Section 4.3) this makes the segmentation far simpler and more reliable (Figure 6).

4.2 Camera positioning and lens selection

Lenses should be selected and cameras should be positioned so that enough of the subject is visible, and so that stereo baselines are sufficient to achieve the required accuracy. This can be a challenging trade-off, as longer baselines give greater accuracy only if feature matching and localisation errors do not also increase (e.g. because of appearance changes).

To design the pruner robot, we built a software model of the canopy, and tested lenses and camera positions within this model so that the entire height of the vine was visible, from the highest canes down to the middle of the trunk, with the blue background behind the vines. Vine dimensions were provided by vineyard managers. Unfortunately vines are often lower than the system was designed for, and some rows cannot be modelled because the vine's head regions are outside the camera's field of view. Moving cameras is not simple, because positions are restricted by the frame, lenses and background position. Even when the vines can be modelled, reconstruction is more likely to fail for important low canes that are only partially visible. This introduces structural errors into the reconstruction, and affects pruning decisions. The next iteration of the pruner robot will be designed based on field measurements.

The first DIET machine also missed data because of poor camera positioning—data from one-in-three patients from an early clinical trial were unusable because the breast was partly outside the camera's field of view.

4.3 Lighting design

Lighting must be setup so that scenes are evenly illuminated, so that objects' appearances do not change depending on their position [12]. This is challenging when imaging 3D scenes where the camera's field of view or depth of field are large, where shadows and occlusions are common, or where object's are shiny and show specular reflections—these make the same object appear differently in different cameras, and may saturate the sensor. Even when objects are pure Lambertian reflectors (their colour appears the same from any viewpoint), obtaining

even illumination is challenging, because light intensity drops quadratically with distance from the light source, and because many light sources' (including "wide angle" machine vision lights) intensities drop as the angle from the light's centre increases. [4] used a computer model of the pruner robot and the DIET machine to design more effective lighting configurations, which give more even lighting throughout the scene. For the pruner robot, the optimal configuration of 14 light sources provides illumination levels that vary by 48% across the vines, whereas a simpler configuration (a regular grid) give 70% variation, and a single point source gives 96% variation. The most effective configurations position lights in a 2m wide ring around the cameras. Having more light sources provides more robustness to shadows, and positioning most light sources further away from cameras mitigates the effect of light intensity decreasing with depth.

For the DIET machine, the optimal configuration of five light sources is a large circle just above the cameras. The existing machine was modified to obtain this configuration: lighting variation between the top and bottom of the breast fell from 40% to 30% when 5 of 30 LEDs were masked out.

4.4 Light sources

Machine vision lights and strobes are widely available, however current commercial solutions don't have the wide-angle and high intensity that the pruner robot and DIET machine require [4]. It is straightforward to build suitable lights from high-power wide angle (or "unlensed") LEDs, heat sinks, and commercially-available "constant current" LED power supplies, however obtaining constant light levels is challenging, due to artefacts remaining from the mains AC power input [12], and because power supplies modulate the voltage to keep the current constant while the LED's resistance changes with temperature. On the pruner robot, an additional capacitor on each power supply smooths out high frequency flickering[4]. On the DIET machine, the amount of light is not proportional to the strobe duration, and varies with the LED's temperature. Updates to the strobe duration are damped to prevent large lighting fluctuations.

4.5 Camera data acquisition

Camera manufacturers provide APIs and example programs for grabbing images from cameras onto a PC, and these example programs are easy to adapt for particular applications. The challenges in data acquisition are synchronising cameras and getting large amounts of image data onto the computer and saved to disk. Camera APIs also provide control of colour balance, shutter time, etc. (see supplementary material[5] for a summary of trade-offs required). Auto-exposure and auto white balance cause image changes that make registering views more challenging. The pruner robot has controlled lighting, so these settings can be

[4] `http://www.red.com/learn/red-101/flicker-free-video-tutorial`
[5] Also available from `http://hilandtom.com/PSIVT2015-Checklist.pdf`.

fixed. The DIET machine also fixes these settings, but controls brightness and saturation by changing the strobe duration.

When imaging moving objects, cameras are usually triggered simultaneously so that images are captured at the same moment. Images may also need to be synchronised with strobes, or other events. The most common synchronisation method is to use the camera's external trigger input. Alternatively, several Firewire (1394a/b) cameras using the same card, or multiple cards[6], can be synchronised in software. External triggering is used by commercially-available multi-camera systems[7]. The synchronisation methods used in the pruner robot and DIET machine are summarised in Figure 3. Note that modern computer hardware allows uncompressed or losslessly-compressed high resolution (1.3 megapixel) images from three cameras to be saved to disk at over 30 frames per second, without the need for specialised hardware or video capture cards. Although the pruner robot's computer vision system only requires 2.5 frames per second, the high framerates provide data that is useful for evaluating the effects of different robot speeds [5].

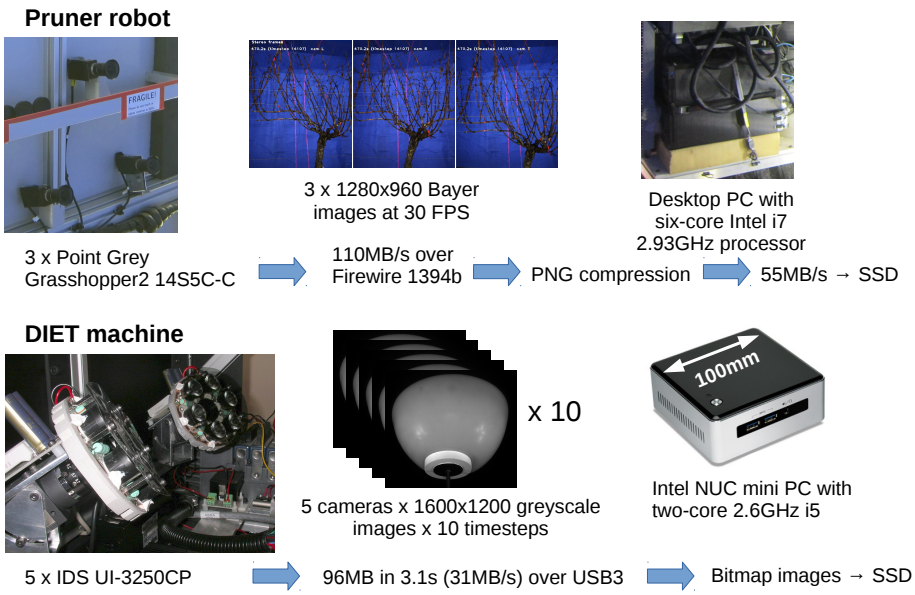


Fig. 3. Examples of image acquisition hardware setups. Modern USB3 and 1394b cameras allow high data-rate uncompressed imaging without specialised hardware (i.e. capture cards).

[6] `https://www.ptgrey.com/KB/10574`
[7] e.g. `http://www.4dviews.com/`

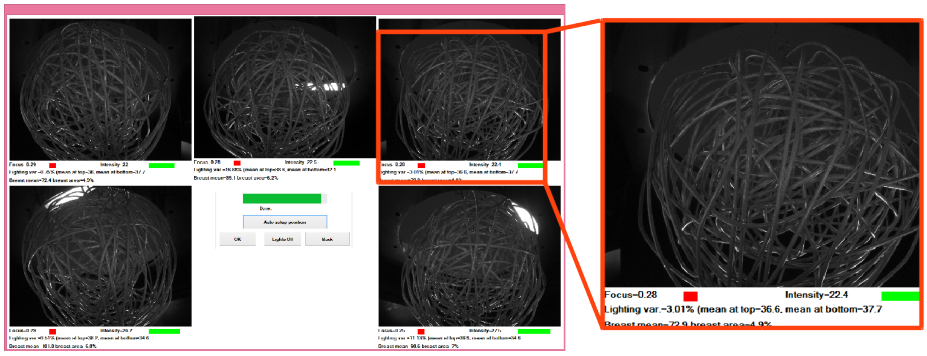


Fig. 4. The DIET machine's GUI for setting camera focus. The wire ball has sharp edges at a range of depths. The focus and aperture are set to maximise the sum of squared differences between neighbouring pixels.

4.6 Lens focus

Important properties of machine vision lenses are their focal length, or zoom (which is usually fixed), and their aperture and focus (which are either fixed, have manual control, or can be controlled automatically). The aperture controls how much light the sensor receives, and the focus setting controls the range of depths for which the image is in focus, for a given aperture. The wider the aperture, the more light is received, but the narrower the range of depths for which the subject is in focus. Setting up lenses so that objects are in focus wherever they appear is challenging: if one part of the scene is in focus, others might not be, and manually inspecting an entire high-resolution image for focus is hard (inspecting edges to verify they aren't blurred requires zooming-in). A slight loss of focus might be acceptable for some applications (although spatially-varying focus is generally undesirable) as many computer vision methods, e.g. optical flow, invariant features, use Gaussian blurring to reduce the effects of noise and quantisation.

Contrast detection autofocus [14] is commonly used in consumer cameras. The camera scans across a range of focus settings, and selects the setting that maximises a measure of focus. The effect of a lens being out of focus is to blur the image, and hence an image that is out of focus has smaller differences between neighbouring pixels. For a fixed aperture, a simple and effective [14] measurement of focus is the sum of squared differences between neighbouring pixels. If this focus measure is displayed as a camera captures images, the lens's focus can be adjusted until this measure is maximised (Figure 4).

The challenge in using fixed-focus cameras for 3D machine vision is that the range for which they are in focus must include the full range of depths where the object might be visible, across the entire image. A sum-of-squared-differences focus measure is only valid for a fixed scene, so this scene should contain textured objects at a suitable range of depths, across the entire image.

The pruner robot's cameras were setup by imaging vines (which have many sharp edges at the required range of depths), using both a GUI that automatically measures focus, and manual inspection of the images. There are still regions of the images which are not well focussed throughout the required depth range, and the wire detector performs poorly in these regions, impacting the completeness of the 3D reconstruction of the scene.

The first DIET machine was focussed manually. Images from an early clinical trial are out of focus in places, which probably impacts the accuracy of skin motion tracking. The current DIET machine uses a GUI to automatically measures focus. The focus is set while imaging a wire ball (Figure 4), because the wire ball has sharp edges at the entire range of depths where the breast surface could be. The optimal parametrisation for optical flow estimation has a larger kernel size (blur size) for computing derivatives for the current machine than the old machine, because images are now sharper.

Autofocus cannot generally be used for 3D computer vision because changing the focus may change the camera calibration. Objects move around the image when the focus on the lenses on the pruner robot is adjusted (10MP, $\frac{2}{3}''$, 5mm Goyo C-mount lenses).

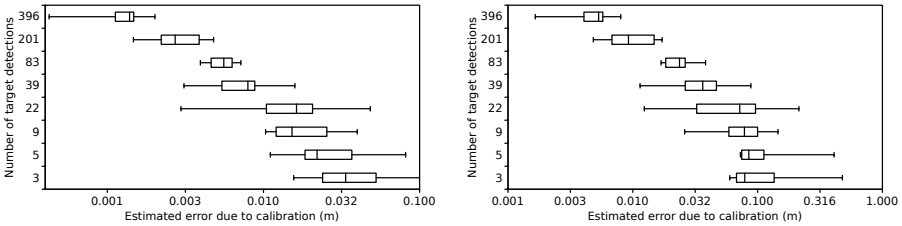
4.7 Camera calibration

Camera calibration is the process of finding a transform (a camera matrix, and possibly distortion parameters) that maps the position of objects in the world to their coordinates in the image. Zhang's [24] widely-used method for camera calibration involves imaging a calibration target of known dimensions (often a checkerboard pattern), locating the calibration target in each image, then optimising calibration parameters and estimated target positions to minimise the image distance between projected and measured target positions (e.g. using Levenberg-Marquardt optimisation). The calibration models the effects of the lens and sensor size (intrinsic parameters), and the position and orientation of each camera (extrinsic parameters).

OpenCV [16] has routines for detecting calibration targets, and for calibrating pairs of stereo cameras. Both the pruner robot and the DIET machine use OpenCV's target detection routines, then use Zhang's method to estimate the intrinsic and extrinsic parameters for all of the cameras jointly, avoiding any loss of accuracy from combining multiple pairwise calibrations.

OpenCV's calibration pattern detector is most reliable when calibration targets have a large white border[8]. A problem with checkerboard and dot-pattern targets is that their orientation is ambiguous—the target can be detected in different orientations (180 degrees out) in images from two cameras. If undetected, this will cause the calibration to fail. Either marking the calibration target (e.g. adding a coloured mark to one corner, which can be detected and used to resolve the ambiguity), or using a non-symmetric pattern (e.g. [17]) prevents this

[8] `http://docs.opencv.org/modules/calib3d/doc/camera_calibration_and_3d_reconstruction.html`



(a) Mean error across scene. (b) Maximum error anywhere in scene.

Fig. 5. Effect of the number of targets detected in both images of a stereo pair on the accuracy of reconstructed 3D points. Box plots show the range, quartiles, and median errors, from 15 calibrations for each number of targets. The scene is 2m wide, with a depth of field of 0.5m to 1.75m.

problem. Some printers scale the width and height of a target differently; this is another potential point-of-failure to check.

Capturing calibration target images in a range of poses throughout the region where objects are imaged is important for obtaining accurate calibrations [17]. Robot Operating System (ROS) has guidelines for capturing target images for stereo calibration[9]. Figure 5 shows the effect of the number of target images on the accuracy of the calibration of one pair of cameras on the pruner robot. For ground truth, we assume that a calibration with 802 targets detected in both images is accurate (so the error estimates are actually lower bounds). We then estimated calibrations from randomly-selected subsets of the detected targets. The accuracy of the subset calibrations is measured by sampling 3D points throughout the region of interest, projecting to 2D using the accurate calibration, reconstructing with the subset calibration, and comparing to the original 3D points. Higher accuracy is obtained when more images are used: capturing less than fifty images gives average errors of more than 1cm in the 3D reconstruction due to calibration alone. The ROS guidelines, and [17], note that common practice is to capture dozens of target images, to give a suitable distribution of pattern positions for every pair of cameras. For multi-camera systems, we recommend capturing up-to a thousand images of a pattern, so that hundreds of targets are detected for every pair of cameras. In our experiment, 1352 stereo images of the target were captured, and the target was detected in both images 802 times. For the DIET machine, over 600 images are needed to ensure there are at least 40 detections for every pair of cameras. Capturing this many images is an inexpensive way of reducing the errors in the 3D reconstruction.

Obtaining an accurate camera calibration is challenging. [17] write that "Reliable and accurate camera calibration usually requires an expert intuition to reliably constrain all of the parameters in the camera model", and conduct human trials to show that accurate calibrations are rarely obtained by novices. They propose using a software tool to guide users through the calibration pro-

[9] `http://wiki.ros.org/camera_calibration/Tutorials/StereoCalibration`

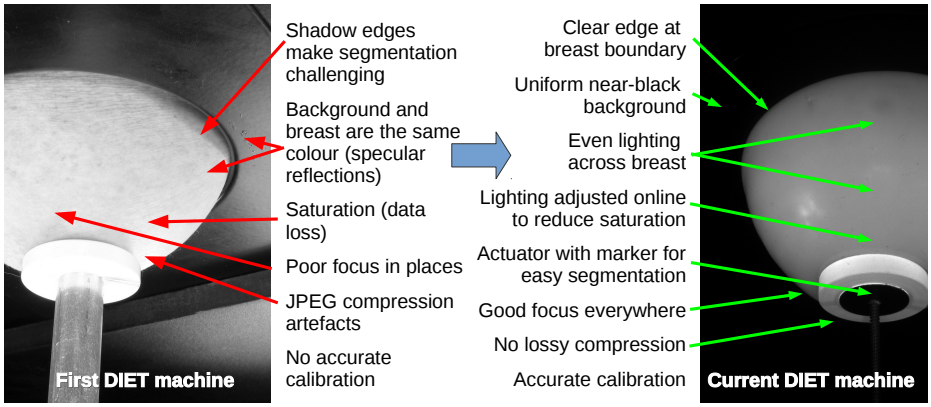


Fig. 6. Hardware changes have improved several different aspects of the DIET machine, so that the images can now be used to track skin texture rather than requiring markers.

cess for a single camera. Much early research on the DIET project was on camera calibration with various calibration objects ([7], Chapters 4 and 5), however the methods were ill-conditioned (often using only two faces of a single object), and an accurate calibration was rarely obtained [6]. We recommend capturing images of standard calibration targets whenever data is collected, both to validate that the calibration is unchanged, and so that cameras can be calibrated retrospectively if necessary.

For multi-camera systems for 3D reconstruction, an extra consideration is the choice of the origin. Often one camera is chosen to be at the origin (e.g. in OpenCV's routines), but it may be more appropriate to choose an origin which aligns the 3D model with the machine. The pruner robot selects the origin so that the volume in which the vines move is an axis-aligned box. This simplifies the application of constraints on the 3D reconstruction, which come from the physical dimensions of the machine, and means that if one camera moves and loses calibration, other calibrations (the robot arm position, and a laser line structured light scanner) do not change. The DIET machine's origin is in the machine's centre, and is aligned with the patient, so that tumour positions can be matched between the patient and the 3D model.

4.8 Camera mounting

Calibrating cameras is time consuming, and undetected calibration changes are a potential source of error. It is important to attach cameras securely so that the calibration does not change. Most machine vision cameras are mounted with either a single $\frac{1}{4}''$ tripod screw, or four small bolts. The tripod mount screws are prone to loosening if cameras are knocked, and are easy to over-tighten, so we recommend using the four small bolts. The pruner robot has a metal guard protruding beyond the lens (Figure 1) to prevent people or vines from accidentally knocking the lenses.

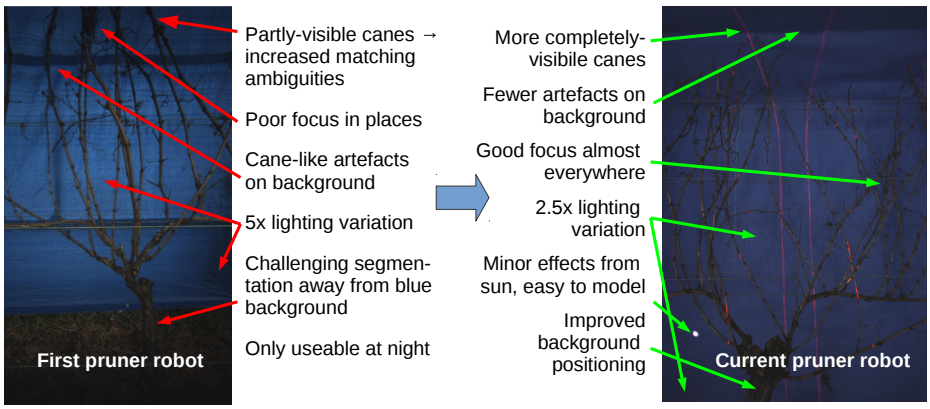


Fig. 7. Hardware changes have improved several different aspects of the vine images from the first pruner robot to the current system.

5 Conclusion

Designing camera systems for 3D computer vision is challenging because of the many factors affecting image quality. Carefully designed hardware systems result in simpler and more robust computer vision systems, and shorter development times. In two case studies, two different teams of engineers made similar mistakes when setting up multi-camera systems for 3D reconstruction, and these mistakes have unnecessarily delayed both projects. In this paper we have listed and analysed many of the design considerations that must be taken into account when designing multi-camera systems, so that future projects don't make the same mistakes. These recommendations are summarised in a checklist and a list of trade-offs, which is provided as supplementary material[10].

References

1. C. W. Bac, E. J. van Henten, J. Hemming, and Y. Edan. Harvesting robots for high-value crops: State-of-the-art review and challenges ahead. *Journal of Field Robotics*, page 888911, 2014.
2. T. Botterill, T. Lotz, A. Kashif, and G. Chase. Reconstructing 3D skin surface motion for the DIET breast cancer screening system. *IEEE Transactions on Medical Imaging*, 33(5):1109–1118, 2014.
3. T. Botterill, S. Mills, and R. Green. Refining essential matrix estimates from RANSAC. In *Proc. Image and Vision Computing New Zealand*, 2011.
4. T. Botterill, S. Mills, R. Green, and T. Lotz. Optimising light source positions to minimise illumination variation for 3D vision. In *3DIMPVT*, pages 1–8, 2012.
5. T. Botterill, S. Paulin, R. Green, S. Williams, J. Lin, V. Saxton, S. Mills, X. Chen, and S. Corbett-Davies. A robot system for pruning grape vines. *Pre-print under review*, 2015. Available online at http://hilandtom.com/tombotterill/pruner-preprint.pdf.

[10] Also available at `http://hilandtom.com/PSIVT2015-Checklist.pdf`.

6. R. Brown, J. Chase, and C. Hann. A pointwise smooth surface stereo reconstruction algorithm without correspondences. *Image and Vision Computing*, 2012.
7. R. G. Brown. *Three-dimensional motion capture for the DIET breast cancer imaging system*. PhD thesis, Department of Mechanical Engineering, University of Canterbury, 2008.
8. Y. Chéné, D. Rousseau, P. Lucidarme, J. Bertheloot, V. Caffier, P. Morel, É. Belin, and F. Chapeau-Blondeau. On the use of depth camera for 3D phenotyping of entire plants. *Computers and Electronics in Agriculture*, 82:122–127, 2012.
9. O. Chum. *Two-View Geometry Estimation by Random Sample and Consensus*. PhD thesis, Czech Technical University in Prague, 2005.
10. C. Czeranowsky and M. Schwr. How do you assess image quality? Technical report, Basler, 2015.
11. A. Erol, G. Bebis, M. Nicolescu, R. D. Boyle, and X. Twombly. Vision-based hand pose estimation: A review. *CVIU*, 108(1):52–73, 2007.
12. I. Jahr. *Handbook of Machine Vision*, chapter Lighting in Machine Vision, pages 73–203. 2006.
13. S. McConnell. *Code Complete*. Microsoft Press, 2nd edition, 2004.
14. H. Mir, P. Xu, and P. van Beek. An extensive empirical evaluation of focus measures for digital photography. In *IS&T/SPIE Electronic Imaging*, pages 90230I–90230I. International Society for Optics and Photonics, 2014.
15. S. Nuske, K. Wilshusen, S. Achar, L. Yoder, S. Narasimhan, and S. Singh. Automated visual yield estimation in vineyards. *Journal of Field Robotics*, 31(5):837–860, 2014.
16. OpenCV Computer Vision Library, n.d. http://opencv.org/.
17. A. Richardson, J.-P. Strom, and E. Olson. Aprilcal: Assisted and repeatable camera calibration. In *Intelligent Robots and Systems (IROS)*, pages 1814–1821, 2013.
18. A. Silwal, A. Gongal, and M. Karkee. Apple identification in field environment with over the row machine vision system. *Agricultural Engineering International: CIGR Journal*, 16(4):66–75, 2014.
19. A. Telljohann. *Handbook of Machine Vision*, chapter Introduction to building a machine vision inspection, pages 35–71. 2006.
20. Vision Robotics Corporation. http://visionrobotics.com/. online, 2015. Retrieved July 2015.
21. Q. Wang, S. Nuske, M. Bergerman, and S. Singh. Automated crop yield estimation for apple orchards. In *Experimental Robotics*, pages 745–758, 2013.
22. A. Wedel, T. Pock, C. Zach, H. Bischof, and D. Cremers. An improved algorithm for TV-L1 optical flow. In *Statistical and Geometrical Approaches to Visual Motion Analysis*, pages 23–45. 2009.
23. B. Zhang, W. Huang, J. Li, C. Zhao, S. Fan, J. Wu, and C. Liu. Principles, developments and applications of computer vision for external quality inspection of fruits and vegetables: A review. *Food Research International*, 62:326–343, 2014.
24. Z. Zhang. A flexible new technique for camera calibration. *T. Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, 2000.