Removing occlusions from light field data
Research project proposal

Ashley Stewart

October 28, 2016

Contents

1 Introduction

Computer vision is developing rapidly, empowering industrial robots, autonomous vehicles, surveillance cameras and other intelligent agents. In many applications, objects can obstruct a camera's view, preventing ideal function. This often occurs in surveillance and navigation during rainfall, hail or snow, among other scenarios. It also occurs in the surgical theatre for both external training videos and internal views such as those gathered by an endoscope. Occlusion removal enables the reconstruction of occluded segments to effectively 'see behind' such obstacles. This project proposes to implement a novel occlusion removal method using light field cameras.

Occlusion removal has already been demonstrated across a range of capture mediums (see section 2.3). However, most methods require either an elaborate, expensive, or stationary camera setup, only work in controlled environments, cannot be achieved quickly, automatically, or in real-time, or are designed to work only for images rather than video.

Light field cameras are a relatively new technology enabling the capture of light fields emanating from scenes. Little work has been done on occlusion removal specifically for light field cameras, with only one paper that we know of directly touching on the topic; one which is also inflexible for general applications [7]. A method to remove occlusions using light field cameras would be convenient for a number of reasons. Firstly, because light field cameras capture and produce a model of a light field, artificial views have a high level of metric accuracy, making them appropriate for robotic vision or applications where accuracy is important. The use of light field cameras also provides a layer of abstraction; an occlusion removal method that can be applied to one light field camera could potentially be applied to any light field camera. This means it does not necessarily require a custom or elaborate camera setup. Finally, light field video is an area that has not yet seen a great deal of research. It is possible that the properties of light field video present benefits for occlusion removal.

As well as contributing to research in occlusion removal and light field video, this project may build upon methods for object tracking, background segmentation (or *foreground detection*) and synthetic aperture focusing.

1.1 Aim

The aim of this project is to explore the properties of light field video signals and how they can be applied to achieve occlusion removal. Specifically, the project aims to answer the following research questions:

- How can the properties of light field video be exploited for occlusion removal?
- How can occlusion removal be applied to light fields?
- How can occlusion removal be applied to light field video?
- How can occlusion removal be achieved in real-time via a light field video stream?
- Can existing occlusion removal methods be made more robust for light field cameras?
- Could such occlusion removal methods be integrated into real-world applications using light field cameras?

A series of demonstrations have been proposed to illustrate a novel occlusion removal implementation for light field video, the results of which may lead to answers to some of the above research questions (see section 1.4.1). The effectiveness of the implementation can be tested via the proposed demonstrations using the following criteria:

- Robustness to occluders of varying shapes and sizes
- Robustness to occluders at varying distances
- Robustness to occluders moving at various speeds (video only)
- Robustness to occluded objects moving at various speeds (video only)
- Continued effectiveness during camera motion (video only)
- Computation time (video only)

Finally, in light of project findings, the project aims to make justified comments about the occlusion removal potential for several applications of interest (outlined in section 1.3).

1.2 Scope

Although occlusion removal techniques are useful across a wide range of applications, the project will only consider a small selection (see section 1.3). A series of demonstrations have been proposed which illustrate properties relevant to those applications.

Additionally, demonstrations are limited by the available resources. They will only be tested via the available Raspberry Pi Light Field Array Camera, which will be calibrated via Xu et al's procedure for mobile light field cameras [20]. Demonstration of occlusion removal in conditions such as rainfall, hail and snow or via an endoscope will also not be possible given the available time and resources. Instead, the demonstration design will illustrate key properties of those applications (e.g. small, fast-moving occluders to illustrate effectiveness during rainfall).

1.3 Applications

Occlusion removal has many applications, including surveillance, navigation, robotic vision, as well as surgical applications, among others. As previously stated, demonstrations have been designed to illustrate the possible effectiveness in a limited range of specific applications. Three main applications of interest have been described.

Scenarios where camera vision is hindered, such as during rainfall or in hail leads to applications of interest. If occlusion removal can be demonstrated in such scenarios, the implications would be significant. This would provide great benefits to automated and piloted navigation systems as well as surveillance technology.

Other applications of interest are found from scenarios where solid objects are obstructing an area of interest. There is particular interest for such technology in the surgical theatre. In a knee arthroscopy, for example, the surgeon must insert an arthroscope and endoscope into incisions on either side of the patient's knee. The area within the incision is then filled with fluid, and vision through the endoscope is obscured by floating debris, bubbles, and the arthroscope. An endoscope with occlusion removal capabilities would allow the surgeon to have increased awareness of the inside of the knee, and provide useful images for analysis and training.

The final application of interest involves the use of a light field camera in the surgical theatre, as a replacement for the traditional top-down camera. This would allow larger obstructions such as medical equipment to be removed from view for training purposes.

1.4 Outcomes

The proposed outcomes take the form of a scientific research article along with a series of demonstrations.

The research article will document the implementation, demonstration results, along with any findings related to the project aims. Comments will also be made about findings that shed light on the potential for occlusion removal in applications of interest, along with an outline for possible future work. The article may be published at a conference such as the Australasian Conference on Robotics and Automation. The implementation may also be made publicly available by integrating it into Dansereau's *Light Field Toolbox* for MATLAB [2] to be used for a range of applications.

1.4.1 Demonstrations

A series of practical demonstrations have been outlined, from which useful findings and results will be recorded. Additionally, justified comments will be made about the demonstration results and potential effectiveness for applications of interest. The demonstrations have been designed so that they can be achieved in the early and late stages of the project lifetime, and each are closely tied to major project milestones. These demonstrations will act on light fields, light field video, and a live light field stream.

The first demonstration will illustrate the effectiveness of calibration via digital refocusing. This demonstration can be achieved easily in the early stages of the project using the light field toolbox. The demonstration scene will be indoors and will contain objects of varying distances to show changing clarity as the focal plane is adjusted. Digital refocusing is a key feature of synthetic aperture focusing, a method which can be used to passively remove occlusions [7].

The second demonstration will occur later in the project, once an initial implementation of an occlusion removal algorithm has been developed. Ideally this will be Xu et al's camera selection algorithm [20], but if this is found to be impractical, it may need to be a variation of a background modelling or synthetic aperture focusing method. The demonstration can be assessed against some of the relevant criteria specified in the project aims (see section 1.1).

If the results of the second demonstration were successful, then the third demonstration will be pursued. This demonstration will work with light field video, and will thus require modifications to be made to the occlusion removal algorithm from the second demonstration (see section 4.3). This will allow the testing of full effectiveness criteria specified in the project aims (see section 1.1). There are risks associated with this demonstration that should be considered (see section 5).

If the results of the third demonstration were successful, a fourth and final demonstration will be pursued. This demonstration will work with a light field video stream, and thus may require optimisations to be made to the previous implementation. The same criteria as the previous demonstration will be used to test effectiveness, with a stricter cut-off for computation time. As with the previous demonstration, associated risks should be considered (see section 5).

2 Preliminary Literature Review

2.1 Introduction

Implementing occlusion removal for a light field camera is a complicated problem. In order to effectively implement and test a method, suitable camera hardware must be calibrated to capture accurate light fields. Several calibration methods exist, each with their own advantages and disadvantages. These must be examined in order to sensibly select an appropriate method. Additionally, existing occlusion removal literature must be analysed in order to identify possible starting points for the project. The following is an abridged extract from the full literature review available in appendix 7.3.

2.2 Camera calibration

2.2.1 Procedures for single camera calibration

The classic photogrammetry approach [11] solves the calibration problem for a single camera using a two-step procedure. The first step is to estimate the intrinsic and extrinsic camera parameters linearly via a closed-form solution. The second step uses nonlinear minimisation to obtain the final values, generally via the Levenberg-Marquardt algorithm [9].

Significant research has been achieved which further develops this two-step approach in different ways. Zhang, Heikkilä, and Tsai have made such developments, and their methods and source code are publicly available [6,15,23]. Zhang's method is particularly flexible, as a good estimate of the camera parameters can be made by capturing images of a planar pattern from at least two orientations (usually the pattern is a checkerboard - the corners of each square act as convenient feature points). Sun and Cooperstock provide an overview and empirical evaluation of the accuracy of these well-known methods [13].

2.2.2 Procedures for light field camera calibration

Single camera calibration approaches such as Zhang's are on their own unsuitable for light field cameras. Rigid transformations between pairs of viewpoints become inconsistent when the cameras are calibrated independently [20]. This inconsistency causes inaccurate estimations of the distances between each camera. Specialised calibration methods are therefore required.

Ueshiba and Tomita describe an extension to Zhang's method for multi-camera systems, which recover the rigid displacements between cameras, as well as the intrinsic parameters [16]. The handiness and flexibility of Zhang's method is maintained, as only two captures of a known planar pattern at different orientations are needed. The algorithm presents a homography matrix to act between the camera image and planar pattern. This leads to a measurement matrix with an unknown scale. This matrix can then be factorised to find camera and plane parameters. However, lens distortion is not considered in this method.

Svoboda et al. present a convenient method to calibrate multi-camera arrays, which also uses this factorisation approach [14]. However, instead of using a planar reference pattern, the calibration object is a freely moving bright spot, such as one generated by a laser pointer. This method was designed for virtual environment applications, and deals with a fixed volume and static camera system, and therefore is inflexible to a more dynamic camera system such as the one in our project.

Xu et. al. present a method to calibrate a mobile camera array in which the working volume and viewpoints need not be fixed [20], which also performs in Zhang's style by moving a checkerboard pattern. The method is flexible enough to allow the user to assign the number of viewpoints, and global optimisation of the intrinsic parameters is optional. The method also models radial distortion and achieves accurate results.

Dansereau et al. describe a method to calibrate a lenslet-based camera [1]. The light field camera's initial pose is estimated by taking the mean or the median of each image's pose estimate by following a conventional single camera approach [6,15,23]. Following this, the camera's intrinsic parameters are estimated through a closed-form solution for the camera's intrinsic matrix. The estimates are then refined through an optimisation such as those used in conventional approaches. Finally, distortion parameters are introduced and a full optimisation takes place. This method also introduces a practical

4D intrinsic matrix and distortion model which relate the indices of pixels to corresponding spatial rays. The source code for this method is publicly available from Dansereau's *Light Field Toolbox* for MATLAB.

Vaish et al. present a method that uses a plane plus parallax framework to calibrate large camera arrays [17]. Assuming all cameras lie on a plane parallel to the reference plane, camera positions can be recovered (such as in Ueshiba and Tomiba's approach [16]). This is achieved by measuring the parallax of a single scene point that is not on the reference plane. The light field can then be parameterised as a light slab (Levoy's two-plane parameterisation) [8]. Since the method assumes that all cameras are on precisely the same plane, and only calculates projection to a reference plane in advance of calibration, the accuracy is somewhat diminished for certain setups. This approach is therefore suitable for applications such as synthetic aperture photography, where planar cameras are commonly used.

2.3 Occlusion Removal

2.3.1 Applied to video from a single view

Occlusion removal can be achieved on still video (e.g. on surveillance cameras) by exploiting the temporal axis to perform *background subtraction* (also called *foreground detection*). A common background subtraction method for video data involves thresholding the error between estimates of images with and without occlusions. This generally involves computing a confidence level for each pixel with past and future frames. This allows a background model, and therefore an occlusion layer to be built. The numerous approaches to this problem differ in the type of background model used, and the procedure used to update the model [5,10,12,19]. For example, Stauffer and Grimson's method models each background pixel as a mixture of Gaussians, updating them via an on-line approximation [12].

2.3.2 Applied to light fields

An occlusion removal method specific to light fields involves exploiting focusing techniques via a synthetic aperture. Given enough views and a sufficiently wide synthetic aperture, focussing on a region of interest can effectively blur out occluders in the reconstructed image to the point that they disappear. Vaish et al's plane plus parrallax calibration technique shows improved occlusion removal results via synthetic aperture focussing, compared to results from metric calibration techniques [17].

Vaish et al. have also explored occlusion removal through 3D reconstruction, using cost functions that are robust to occluders. This has been shown to improve the occlusion removal quality of synthetic aperture focusing [18]. Although this technique improves on the results of ordinary synthetic aperture focusing, its performance drops significantly in the case of complicated or severe occlusion.

To overcome these issues with severe occlusion, Yang et al. consider occluded object imaging a problem of light ray selection from optimal camera views [21]. An optimal camera selection algorithm and greedy optimisation is used to to propagate visible ray information from depth focus planes. This approach leads to a much clearer reconstruction of occluded objects.

2.3.3 Applied to camera array video

For the case where an array of cameras is used (though not necessarily a light field camera or cameras even on the same plane), synthetic aperture focusing can be combined with object detection and tracking algorithms. Joshi et al. uses a straightforward approach built from existing single camera tracking algorithms, which tracks moving objects through severe occlusions [7]. The method tracks objects with up to 70% occlusion on all cameras via detection aggregation across views.

Similarly, Yang et al. describe a method to remove occlusions from video via object tracking and synthetic aperture focusing [22]. A synthetic aperture imaging system is used to model precise locations of objects in a controlled scene (see Figure 1). Yang's method also enables the seamless interaction among detection, imaging and tracking modules via a hybrid framework, and introduces an improved synthetic aperture focusing method. However, its use is limited to controlled scenes such as in their setup.

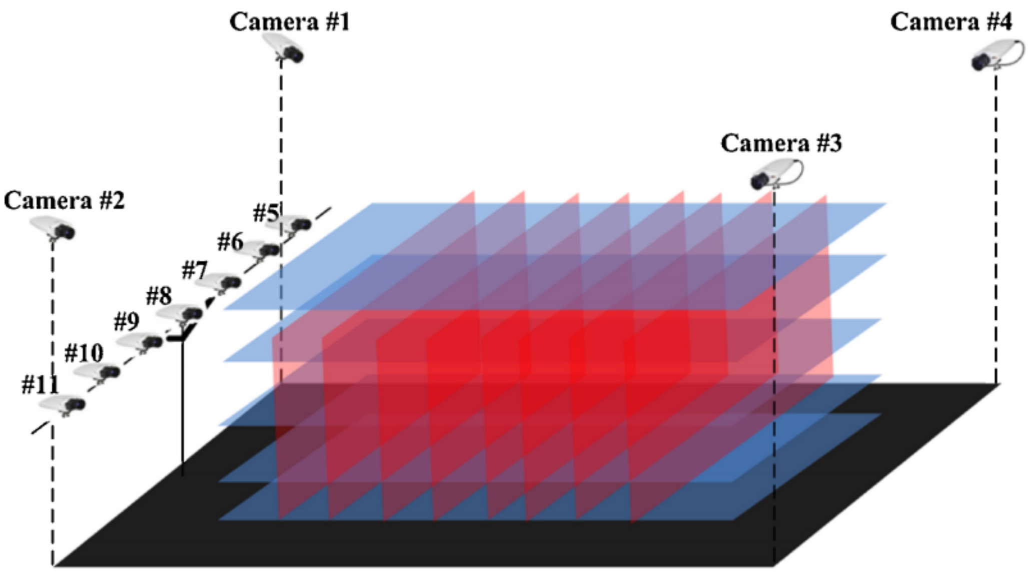


Figure 1: Hybrid synthetic aperture imaging system. Adapted from Tao Yang, Yanning Zhang, Xiaomin Tong, Xiaoqiang Zhang, and Rui Yu. A new hybrid synthetic aperture imaging model for tracking and seeing people through occlusion. *Circuits and Systems for Video Technology, IEEE Transactions on*, 23(9):1461–1475, 2013.

2.3.4 Applied to light field video

Research on occlusion removal on light field video is limited. However, it has been shown that some occlusion removal can be achieved through depth-velocity filtering. Edussooriya et al. have analysed the spectrum of a light field video corresponding to a Lambertian object with constant depth and velocity. They have showed that the object can be enhanced and modelled when occluded based on its depth and velocity via a 5D depth-velocity filter [4]. This method is applicable to a small set of applications, as it only works when objects are modelled at constant velocity and depth.

2.4 Review Conclusion

An overview of camera calibration literature and methods have been presented, which is necessary in order to sensibly choose a calibration procedure for the available light field camera. A similar overview for occlusion removal procedures has also been given, some of the methods of which will provide the foundations for our implementation.

It is clear from the review that current research in light field video is limited. The existing methods described are also relatively inflexible to more dynamic scenes. However, plenty of work has been done on occlusion removal in other contexts, and their concepts may be applied to produce a novel increment or new method. There is a great opportunity for a valuable addition to this area.

3 Resources required

In order to complete the project, several important physical and human resources are required. Fortunately, all such resources have been made available for the duration of the project.

A light field camera will be needed to perform the demonstrations and execute the proposed methodology. As such, a light field camera built as a final year undergraduate project has been made available (see figure 2). A summarised version of the full technical report [3] for the device has been provided in appendix 7.1. The device uses a 4x4 array of Raspberry Pi camera modules each connected to a Raspberry Pi 2 (RPi). The RPis are networked together, allowing them to act as essentially one device. Captured images can be sent through the network via an Ethernet cable to a server machine for further processing.

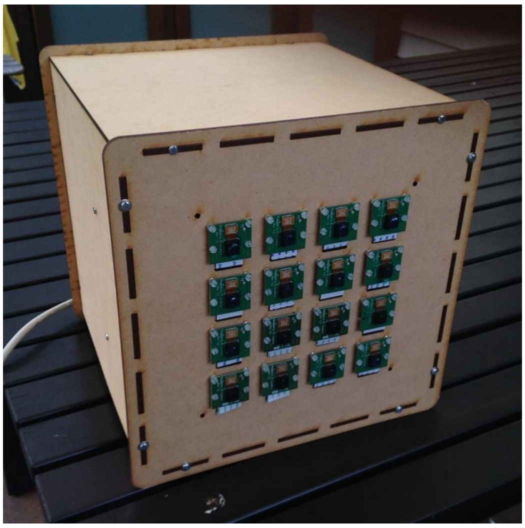


Figure 2: Raspberry Pi Light Field Array Camera

In addition to the light field camera, a workstation will be needed to interact with the camera and develop the necessary implementation. Workstation details are available in the MAPS risk assessment associated with the project.

For effective completion of the project, several human resources will be needed. The supervisory team will provide useful guidance and advice about the project's direction and problems encountered. A technical team is also needed in case any hardware related issues occur. Their details are also provided from the MAPS risk assessment page.

4 General methodology

This section details the proposed methodology that the project will follow. It is important to note that the methodology ties very closely with the proposed demonstrations outlined in section 1.4.1.

4.1 Camera setup

The casing of the Raspberry Pi Light Field Array Camera is made from Medium Density Fibreboard (MDF), which is quite malleable. Unfortunately, this means that small forces applied to the frame can seriously affect calibration accuracy. To overcome this, a sturdier aluminium front plate has been designed to replace the MDF plate which will maintain calibration accuracy. The replacement plate can be considered the first milestone in the project which then enables the completion of camera calibration. The design for the new plate is available in appendix 7.2.

4.2 Calibration Procedure

Calibration is a necessary step in computer vision which ensures that captured images are consistent with the actual geometry of the scene. Many algorithms have been developed to calibrate conventional cameras and light field cameras, and are discussed in the literature review (see section 2). It is important to note that calibration algorithms each have their advantages and disadvantages. As such, the chosen method should suit the available light field camera, as well as any requirements necessary for occlusion removal.

The main objective of calibration in terms of enabling occlusion removal is metric accuracy. Accurate geometric information about the scene must be recovered if occlusions are to be actively removed. Passive removal of occlusions via synthetic aperture focusing should also be possible. Other objectives include simplicity and flexibility of the method due to time and budget limitations. Additionally, the calibration method must be suitable for a mobile light field camera array, since that is the available camera's classification (rather than a lenslet-based or fixed camera setup).

Vaish et al. describe a method to calibrate a light field camera via a plane plus parallax framework [17]. This method yields clearer results for synthetic aperture focusing, but loses the metric information and camera parameters calculated via other methods such as Zhang's. This may limit the possible applications and occlusion removal features. In their paper, they indicate that metric calibration techniques are better suited to general camera configurations and in applications like metric reconstruction. This may include active occlusion removal, as it ideally involves accurate metric reconstruction of occluded segments.

The chosen procedure is a variation of Zhang's metric calibration method, adapted to suit mobile camera arrays with the goal of light field acquisition. It was proposed and tested by Xu et al. [20]. The method considers one of the cameras in the array a reference camera, and the others relative to the reference, rather than independent viewpoints. However, each camera is initially considered independent in order to obtain initial estimates of intrinsic and extrinsic parameters. Since the relative camera positions are theoretically constant between camera array poses, a full optimisation of camera parameters takes place in terms of each camera's median relative pose, rather than independent poses. This allows the rigid transformations between viewpoints to remain consistent.

An implementation of Dansereau et al's calibration method for lenslet-based plenoptic cameras [1] is available for the project via the light field toolbox. Xu et al's method will be implemented as an adjustment to this method. It will then be used to calibrate the camera array.

Calibration will be achieved by first capturing synchronised images from several camera poses with a planar pattern in view via the CompoundPi software package. CompoundPi provides a means for controlling cameras attached to Raspberry Pis in the same subnet, as recommended in the Raspberry Pi Array Camera documentation [3]. The images will then be downloaded to the workstation machine so that the intrinsic and extrinsic parameters can be calculated via Xu et al's method. Accuracy can then be tested by calculating the re-projection error in calibrated light field shots. Effectiveness can also be illustrated through digital refocusing, a simple and popular application of light field imaging, already handled by the light field toolbox.

4.3 Occlusion Removal

Since one of the aims of this project to *explore* the properties of light field video signals in the context of occlusion removal, the implementation of occlusion removal will likely change across the project lifetime as new findings are made. Therefore, a starting point based on existing research will be identified, along with possible alternatives and other features for possible experimentation.

Yang et al propose and demonstrate a method to remove occlusions via an optimal camera selection algorithm [21]. This is different from the more traditional background model [5, 10, 12, 19] and synthetic aperture focusing methods [7, 17, 18, 22]. It considers occlusion removal a problem of optimal light ray selection in order to produce a 'mosaic' of occluded segments using light rays from ideal viewpoints. This produces results that are much clearer than those obtained through synthetic aperture focusing (see figure 3). It has also not been demonstrated with light field video, nor optimised to take advantage of the properties of light field video signals. It therefore lends itself very well as a starting point for this project.

Figure 3: Occluded view from camera array (left), synthetic aperture focusing (middle) vs. light ray mosaic (right). Adapted from Tao Yang, Yanning Zhang, Xiaomin Tong, Wenguang Ma, and Rui Yu. High performance imaging through occlusion via energy minimization-based optimal camera selection. *International Journal of Advanced Robotic Systems*, 10, 2013.

Yang et al's method will serve as a starting point for demonstrating occlusion removal with the RPi light field camera. The first occlusion removal milestone will be to demonstrate occlusion removal on a previously captured light field using an implementation of the algorithm on MATLAB. Following this, the algorithm will be adapted to function with light field video and finally live light field video. See section 1.4.1 for more details about the proposed demonstrations. If possible and practical, the demonstrations may be achieved in combination with synthetic aperture focusing, background modelling or object tracking features to demonstrate and document their effect on occlusion removal.

5 Risks

There are a number of potential risks associated with the project that relate to the implementation and available resources.

The CompoundPi software package is used to capture synchronised images and videos from the Raspberry Pi camera modules. Although a hardware level synchronisation is preferable, it is unfortunately not possible because the RPi board is not open-source. The synchronisation disparity using the software alternative has been measured to be between 20 and 20 milliseconds [3]. This is a significant disparity which could cause light fields to have blurred movement if objects in the scene are not completely still. It may also mean that recorded light field videos need to be manually sliced so that they have precisely the the same start and end points. A simple way to do this would be to have a running stopwatch visible in all scenes. Recorded video could then be sliced post-capture, ensuring the stopwatch is synchronised between all views.

The synchronisation disparity may also make live occlusion removal difficult. Obviously, the stopwatch solution cannot be applied live. If the synchronisation disparity is significant enough, the live demonstrations may have to be limited further. A possible alternative would be to attempt live occlusion removal with very slow movement, or completely still scenes. Small movements in still scenes may then be made before the scene becomes still again, so that the occlusion removal can act live on the updated scene. Another potential risk with live occlusion removal is related to the high network bandwidth necessary for image transfer, as well as as the computation time required for occlusion removal. In light of these risks, if the live occlusion removal demonstration is unsuccessful, it may be considered an area for future work. However, implementation details and proposed algorithms for live occlusion removal may still be provided.

6 Timeline

The figure below (figure 4) illustrates the proposed timeline for the project. The timeline is closely tied to the methodology (section 4) and demonstration outcomes (section 1.4.1). Each delivery is dependent on the previous delivery (therefore a Gantt style chart would be superfluous). The start date is the 13th of July, 2016, and the final thesis due date is the 28th of October, 2016. Exact dates of other project milestones are not included, as there is a degree of uncertainty. Instead, approximate delivery times have been marked.

Figure 4: Proposed timeline

References

[1] D. G. Dansereau, O. Pizarro, and S. B. Williams. Decoding, calibration and rectification for lenselet-based plenoptic cameras. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 1027–1034, June 2013.

[2] Donald Dansereau. *Light Field Toolbox v0.4*, 2015 (accessed June 10, 2016). `http://www.mathworks.com/matlabcentral/fileexchange/49683-light-field-toolbox-v0-4`.

[3] Rafe Denham. Light field array camera. Technical report, Queensland University of Technology, School for Electrical Engineering and Computer Science, 06 2015.

[4] Chamira US Edussooriya, Donald G Dansereau, Len T Bruton, and Panajotis Agathoklis. Five-dimensional depth-velocity filtering for enhancing moving objects in light field videos. *Signal Processing, IEEE Transactions on*, 63(8):2151–2163, 2015.

[5] Nir Friedman and Stuart Russell. Image segmentation in video sequences: A probabilistic approach. In *Proceedings of the Thirteenth conference on Uncertainty in artificial intelligence*, pages 175–181. Morgan Kaufmann Publishers Inc., 1997.

[6] Janne Heikkila and Olli Silvén. A four-step camera calibration procedure with implicit image correction. In *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, pages 1106–1112. IEEE, 1997.

[7] Neel Joshi, Shai Avidan, Wojciech Matusik, and David J Kriegman. Synthetic aperture tracking: tracking through occlusions. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8. IEEE, 2007.

[8] Marc Levoy and Pat Hanrahan. Light field rendering. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 31–42. ACM, 1996.

[9] Jorge J Moré. The levenberg-marquardt algorithm: implementation and theory. In *Numerical analysis*, pages 105–116. Springer, 1978.

[10] Christof Ridder, Olaf Munkelt, and Harald Kirchner. Adaptive background estimation and foreground detection using kalman-filtering. In *Proceedings of International Conference on recent Advances in Mechatronics*, pages 193–199. Citeseer, 1995.

[11] Chester C. Slama, Charles Theurer, Soren W. Henriksen, and American Society of Photogrammetry. *Manual of photogrammetry*. American Society of Photogrammetry, Falls Church, Va, 4th edition, 1980.

[12] Chris Stauffer and W Eric L Grimson. Adaptive background mixture models for real-time tracking. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, volume 2. IEEE, 1999.

[13] Wei Sun and Jeremy R. Cooperstock. An empirical evaluation of factors influencing camera calibration accuracy using three publicly available techniques. *Machine Vision and Applications*, 17(1):51–67, 2006.

[14] T. Svoboda, D. Martinec, and T. Pajdla. A convenient multicamera self-calibration for virtual environments. *Presence*, 14(4):407–422, Aug 2005.

[15] R. Tsai. A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Journal on Robotics and Automation*, 3(4):323–344, 1987.

[16] Toshio Ueshiba and Fumiaki Tomita. Plane-based calibration algorithm for multi-camera systems via factorization of homography matrices. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 966–973. IEEE, 2003.

[17] V. Vaish, B. Wilburn, N. Joshi, and M. Levoy. Using plane + parallax for calibrating dense camera arrays. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 1, pages I–2–I–9 Vol.1, June 2004.

[18] Vaibhav Vaish, Marc Levoy, Richard Szeliski, C Lawrence Zitnick, and Sing Bing Kang. Reconstructing occluded surfaces using synthetic apertures: Stereo, focus and robust measures. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 2331–2338. IEEE, 2006.

[19] Christopher Richard Wren, Ali Azarbayejani, Trevor Darrell, and Alex Paul Pentland. Pfinder: Real-time tracking of the human body.

Pattern Analysis and Machine Intelligence, IEEE Transactions on, 19(7):780–785, 1997.

[20] Yichao Xu, Kazuki Maeno, Hajime Nagahara, and Rin-ichiro Taniguchi. Mobile camera array calibration for light field acquisition. *arXiv preprint arXiv:1407.4206*, 2014.

[21] Tao Yang, Yanning Zhang, Xiaomin Tong, Wenguang Ma, and Rui Yu. High performance imaging through occlusion via energy minimization-based optimal camera selection. *International Journal of Advanced Robotic Systems*, 10, 2013.

[22] Tao Yang, Yanning Zhang, Xiaomin Tong, Xiaoqiang Zhang, and Rui Yu. A new hybrid synthetic aperture imaging model for tracking and seeing people through occlusion. *Circuits and Systems for Video Technology, IEEE Transactions on*, 23(9):1461–1475, 2013.

[23] Zhengyou Zhang. A flexible new technique for camera calibration. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(11):1330–1334, 2000.

7 Appendix

The attached documents provide information about the Light Field Array Camera and the proposed replacement front plate.

7.1 Light Field Array Camera Documentation

Light Field Array Camera

Rafe Denham

***Abstract*—The aim of this project was to design and build a parallel camera array that can capture a light field. This was completed through the use of Raspberry Pi cameras and Raspberry Pis, an Ethernet network to facilitate communication and a frame to hold it together. The final camera array lacked good calibration, but outputted clear, high resolution, images of still scenes. The resulting camera array facilitates further research into light field arrays and also opens the possibility of parallel processing though the use of the Raspberry Pi cluster.**

I. Background

Since the inception of robotics, autonomous imitations of human senses, actions and behaviours have proven considerably more difficult to implement than perceived by most. One of these enigma is computer vision, which is being rapidly broached in modern robotics and is being applied to automation in a wide range of fields. However, these advances in computer vision analysis and processing are consistently hindered by the need for things that humans do subconsciously.

An array of parallel cameras can retrieve a light field array which can be used to remove occlusions, remove shadows, 3D image, perceive depth and post-capture focus - to name a few applications among many [1] [2] [3] [4]. Applying this technique to robotics allows for better image filtering, object recognition and overall computer vision accuracy.

A. Scope

The project is restricted to building a scalable array of cameras with synchronised shutters which, upon request, send their individual image data streams to embedded computers for initial image processing and compiling. This synchronised data then needs to be sent to an external server for further processing. A frame will be required to facilitate the cameras, embedded computers and associated cable management.

B. Related Work

1) Previous planar camera arrays: Three prominent large monocular camera arrays feature in current literature. The Standford 100 camera array [5], the Massachusetts Institute of Technology (MIT) 64 camera array [6] and the Delaware 9 camera array [7].

The Stanford and Delaware arrays dealt with the high volume data streams associated with capturing several images simultaneously through brute force, while the MIT array used dynamic user input to select regions of interest and only captured the computed relevant regions of the images.

The synchronicity of the images depended upon whether explicit hardware synchronisation was used or not. The Stanford and Delware arrays used explicit hardware synchronisation resulting in a disparity of only 100-200µs [5] [7]. Alternatively, the MIT array used implicit synchronisation though software which caused a larger disparity of 5-10ms [6].

The robustness of these existing arrays is poor as they were designed for scientific use in a controlled environment. All frames consisted of an open steel design, allowing for easy access to manual adjustments and to the cabling. However this also exposed the cameras for incidental bumping requiring each to be manually adjusted and re-calibrated after being relocated.

2) Calibration: In a calibration method decribed by Zhang (2000) [8], each camera is treated individually with one being a reference point for the rest. A number of images are captured at different depths and orientations with reference points compared between them. Zhang describes a camera view through two extrinsic, rotation (R) and translation (t), and five intrinsic parameters, the coordinates of the principle point(u_0,v_0), axis scale factors (α, β) and skew (γ).

An alternative method is described by Vaish et al. [9]. This method is more simple and is only applicable to planar camera arrays. It only adjusts the camera positions by a translation, identifying the rest of the parameters described by Zhang [8] as redundant when applied to a planar array.

3) Light field photography techniques: The captured, decoded and calibrated light field can be used to compute new views of the scene. Through calculating the projected light rays, a synthetic reference plane can be moved, simulating a change in focus of the rendered image. This is visually evident in Figure 1, by varying *D*, while keeping the other parameters constant, the focal length is changed - resulting in a change in focus of the subject.

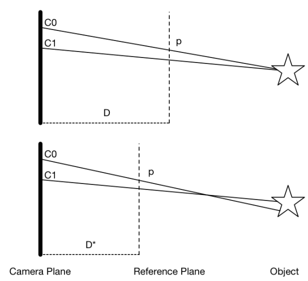
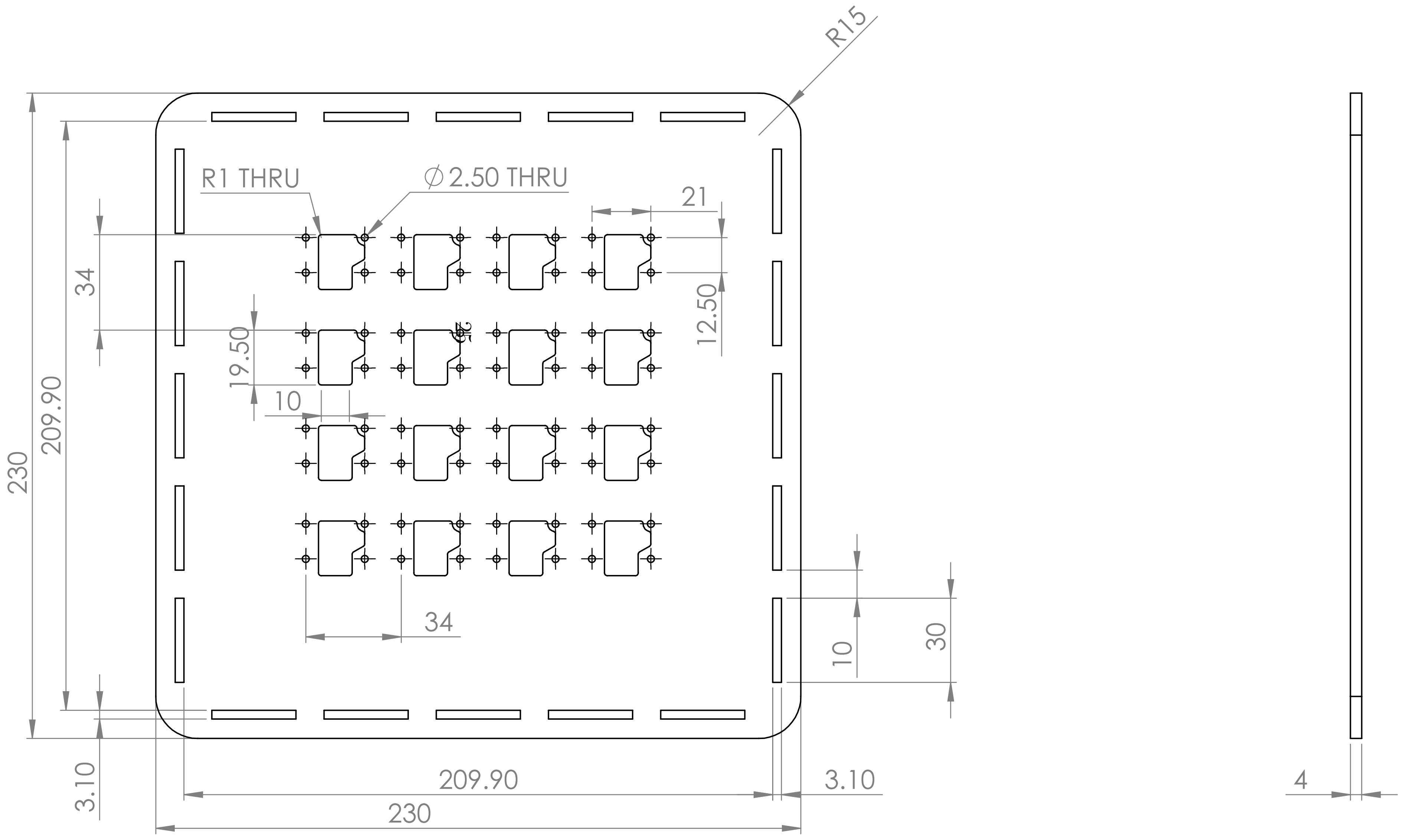


Fig. 1. Diagram of the simulated focal length change.

7.2 Replacement front plate



UNLESS OTHERWISE SPECIFIED: DIMENSIONS ARE IN MILLIMETERS SURFACE FINISH: TOLERANCES: LINEAR: ANGULAR:			FINISH:		DEBURR AND BREAK SHARP EDGES	DO NOT SCALE DRAWING	REVISION
	NAME	SIGNATURE	DATE			TITLE: RaspPi Camera Array	
DRAWN	Steven Martin		11/04/2016				
CHK'D							
APPV'D							
MFG							
Q.A				MATERIAL: Aluminium		DWG NO. FrontPlate2	A3
				WEIGHT:		SCALE:1:2	SHEET 1 OF 1

Removing occlusions from light field data
Literature Review

Ashley Stewart

June 11, 2016

Contents

1 Introduction

Computer vision is developing rapidly, empowering industrial robots, autonomous vehicles, surveillance cameras, and other intelligent agents. Light field camera technology facilitates the capture of light fields emanating from scenes. A light field, as opposed to a conventional pin-hole image, has the additional property of describing the amount of light flowing in every direction through every point on the camera plane. This information can provide advanced spacial awareness to agents, as light fields provide metric information about the scene. Features such as post-capture refocusing, depth perception, glare reduction, ray re-projection, and occlusion removal, among others, can be exploited using light field cameras.

The objective of the project connected to this literature review is to develop and demonstrate occlusion removal on light field data, and in particular on light field video data. The applications of the project outcomes are very broad, and may include surveillance, navigation, robotic vision and medicine, to name a few areas.

Developing an occlusion removal method to act upon light field data is a problem which requires knowledge and application of practices developed across several research areas. Broadly, these areas are: light fields and light field cameras; camera calibration (especially light field camera calibration); and occlusion identification and removal. Important concepts, as well as research in each of these areas will be discussed and reviewed, providing important foundations for the completion of the project, as well as the identification of research gaps.

2 Light fields and light field cameras

2.1 The plenoptic function

The plenoptic function relates a number of parameters central to computer vision. Adelson defines the plenoptic function as the function describing the observable radiance L at any point in space and time [1]. It is given by:

$$L(\boldsymbol{V}, \boldsymbol{E}, t, \lambda) \tag{1}$$

where

$\boldsymbol{V}$ = The centre of projection or viewpoint (V_x, V_y, V_z)
$\boldsymbol{E}$ = The viewing direction (θ, ϕ) parallel with V_z for simplicity
t = Time
λ = Wavelength

A more common version of the function, referred to as the 5D plenoptic function, excludes the wavelength and time parameters.

The plenoptic function is never computed in practice, though familiarity is useful when exploring other concepts in computer vision.

2.2 The 4D light field

Marc Levoy defines a light field as the radiance of a point in a given direction. This is equivalent to the output of the plenoptic function [10]. Levoy also describes a simplification of the 5D plenoptic function, which reduces it to four dimensions. The resulting parameterisation is dubbed the *light slab* representation (see Figure 1). This representation parametrises rays by projecting lines between two arbitrarily placed parallel planes. By convention, points on the first plane are given by (u, v), while points on the second plane are given by (s, t).

In the context of camera arrays where each camera lies on a plane, coordinates on the (u, v) plane identify cameras. Assuming a camera on the (u, v) plane has been identified, coordinates on the (s, t) plane would then identify rays from the focal plane of the scene to the camera.

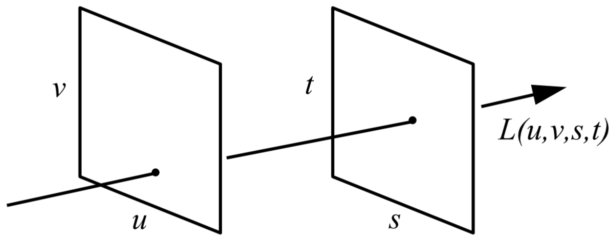


Figure 1: Levoy's light slab parameterisation. Adapted from Marc Levoy and Pat Hanrahan. Light field rendering. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 31–42. ACM, 1996.

2.3 Pinhole camera model

To understand camera calibration algorithms, it is important to discuss the pinhole camera model [15]. Most camera calibration literature will refer to this model, which is the basic camera model used in computer vision. The model performs perspective projection from 2D pixel coordinates to 3D world coordinates (see Figure 2).

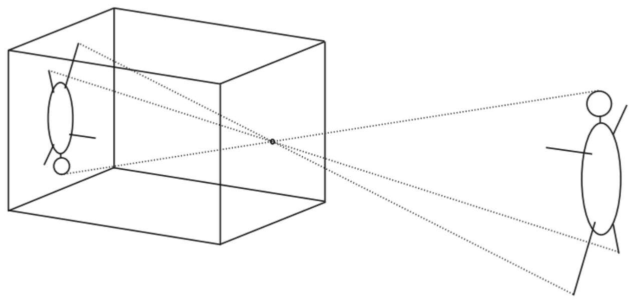


Figure 2: Pinhole Camera Model sketch. Adapted from Peter Sturm. *Computer Vision: A Reference Guide*, chapter Pinhole Camera Model, pages 610–613. Springer US, Boston, MA, 2014.

The pin-hole camera model leads to the following relationship between pixel coordinates and world coordinates:

$$\begin{bmatrix} x_p \\ y_p \\ 1 \end{bmatrix} = \boldsymbol{K} \begin{bmatrix} \boldsymbol{R} & \boldsymbol{t} \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \tag{2}$$

with

$$\boldsymbol{K} = \begin{bmatrix} \alpha_x & \gamma & u_0 \\ 0 & \alpha_y & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$

where

$\boldsymbol{K}$	= Calibration matrix containing intrinsic parameters $(\alpha_x, \alpha_y), \gamma$, and (u_0, v_0)
$\boldsymbol{R}$	= Rotation matrix representing the camera's orientation
$\boldsymbol{t}$	= Coordinates of the centre of projection in world coordinates
(α_x, α_y)	= Focal length in number of pixels
(u_0, v_0)	= Principal point in pixel coordinates
γ	= The skew coefficient between the x and y axes

Some effects are not included in the pinhole camera model, such as radial lens distortion and other geometric distortions. In order to handle such distortions, the camera model must be extended (see section 3).

3 Camera calibration

Camera calibration enables the extraction of accurate metric information from images. A calibrated light field camera is able to produce consistent light fields that measure real-world phenomena. Significant work has been achieved on camera calibration in photogrammetry and computer vision.

Camera calibration is an important concept to explore in order to work on occlusion removal. Occlusion removal deals with reconstructing occluded objects by projecting the occluded rays onto an artificial view. Not only is calibration an important step that must be completed before the project can begin, but it too deals with projecting rays and reconstructing calibrated views.

3.1 The effects of radial lens distortion

The aim of camera calibration is to model the extrinsic and intrinsic camera parameters so that rays can be mapped accurately to pixels. Consider a single camera that can be modelled as a perfect pinhole. Calibration in this case will effectively find the unknowns of the intrinsic matrix $\boldsymbol{K}$ (from the pinhole camera model). However, such a camera is not affected by radial lens distortion, which in reality can be significant (see Figure 3). Many calibration algorithms therefore take radial lens distortion into account by modelling ideal undistorted image coordinates against actual image coordinates.

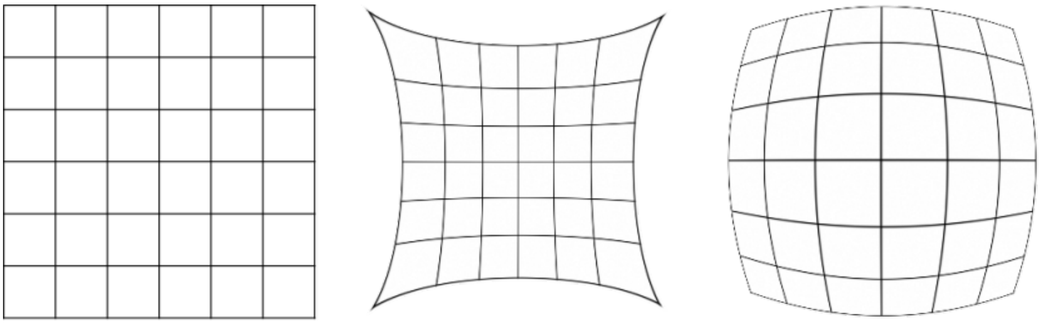


Figure 3: Representation of an undistorted grid (left), grid exhibiting pinhole distortion (centre), and grid exhibiting barrel distortion (right). Adapted from Hoonjong Kang, Elena Stoykova, Jiyung Park, Sunghee Hong, and Youngmin Kim. Holographic printing of white-light viewable holograms and stereograms. *Intech, Rijeka*, pages 171–201, 2013.

3.2 Transforming 3D coordinates to pixels

Several well-known calibration algorithms from Zhang, Tsai and Heikkilä [5, 19, 27] describe a set of transformations that map pixel coordinates to rays, given the camera's position and orientation. Though many other calibration algorithms use the same or similar transformations, Zhang, Tsai and Heikkilä's are particularly well-documented, tested, and publicly available. The transformations provide an insight into the mapping of rays to pixels, as well as the modelling of radial distortion.

The first transformation is from world coordinates $\boldsymbol{V_w} = (X_w, Y_w, Z_w)$ via camera coordinates $\boldsymbol{V_c} = (X_c, Y_c, Z_c)$, where $\boldsymbol{R}$ and $\boldsymbol{t}$ are the extrinsic camera parameters representing rotation and translation. This extrinsic transformation is well-known and does not differ between calibration methods.

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = \boldsymbol{R} \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} + \boldsymbol{t} \tag{3}$$

The second transformation translates camera coordinates to ideal undistorted image coordinates (x_u, y_u). Approaches differ in form from this translation onward, so transformations following Zhang's method are shown. Other methods, such as Tsai's and Heikkilä's, scale the undistorted coordinates by the camera's effective focal length f.

$$\begin{bmatrix} x_u \\ y_u \end{bmatrix} = \frac{1}{Z_c} \begin{bmatrix} X_c \\ Y_c \end{bmatrix} \tag{4}$$

The third transformation, followed by methods which model radial distortion, translates the undistorted image coordinates to distorted image coordinates (x_d, y_d). Several radial terms and coefficients are introduced. The order of the radial terms varies between methods. Zhang's model for distorted coordinates is below, which uses a fourth-order radial term. Lavest et al's calibration approach for underwater applications adds a sixth-order radial term [9]. The effect of this on calibration accuracy has been discussed and evaluated by Sun and Cooperstock [16].

$$\begin{bmatrix} x_d \\ y_d \end{bmatrix} = (1 + k_1 r^2 + k_2 r^4) \begin{bmatrix} x_u \\ y_u \end{bmatrix} \tag{5}$$

where

$$r = \sqrt{x_u^2 + y_u^2}$$

The fourth and final transformation is to pixel coordinates, and refers to the pinhole camera model presented earlier:

$$\begin{bmatrix} x_p \\ y_p \\ 1 \end{bmatrix} = \boldsymbol{K} \begin{bmatrix} \boldsymbol{R} & \boldsymbol{T} \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \tag{6}$$

3.3 Procedures for single camera calibration

The classic photogrammetry approach [13] solves the calibration problem for a single camera using a two-step procedure. The first step is to estimate the intrinsic and extrinsic camera parameters linearly via a closed-form solution. The second step uses nonlinear minimisation to obtain the final values, generally via the Levenberg-Marquardt algorithm [11].

Significant research has been achieved which further develops this two-step approach in different ways. Zhang, Heikkilä, and Tsai have made such developments, and their methods and source code are publicly available [5, 19, 27]. Zhang's method is particularly flexible, as a good estimate of the camera parameters can be made by capturing images of a planar pattern from at least two orientations (usually the pattern is a checkerboard - the corners of each square act as convenient feature points). Sun and Cooperstock provide an overview and empirical evaluation of the accuracy of these well-known methods [16].

3.4 Procedures for light field camera calibration

Single camera calibration approaches such as Zhang's are on their own unsuitable for light field cameras. Rigid transformations between pairs of viewpoints become inconsistent when the cameras are calibrated independently [24]. This inconsistency causes inaccurate estimations of the distances between each camera. Specialised calibration methods are therefore required.

Ueshiba and Tomita describe an extension to Zhang's method for multi-camera systems, which recover the rigid displacements between cameras, as well as the intrinsic parameters [20]. The handiness and flexibility of Zhang's method is maintained, as only two captures of a known planar pattern at different orientations are needed. The algorithm presents a homography matrix to act between the camera image and planar pattern. This leads to a measurement matrix with an unknown scale. This matrix can then be factorised to find camera and plane parameters. However, lens distortion is not considered in this method.

Svoboda et al. present a convenient method to calibrate multi-camera arrays, which also uses this factorisation approach [17]. However, instead of using a planar reference pattern, the calibration object is a freely moving bright spot, such as one generated by a laser pointer. This method was designed for virtual environment applications, and deals with a fixed volume and static camera system, and therefore is inflexible to a more dynamic camera system such as the one in our project.

Xu et. al. present a method to calibrate a mobile camera array in which the working volume and viewpoints need not be fixed [24], which also performs in Zhang's style by moving a checkerboard pattern. The method is flexible enough to allow the user to assign the number of viewpoints, and global optimisation of the intrinsic parameters is optional. The method also models radial distortion and achieves accurate results.

Dansereau et al. describe a method to calibrate a lenslet-based camera [2]. The light field camera's initial pose is estimated by taking the mean or the median of each image's pose estimate by following a conventional single camera approach [5, 19, 27]. Following this, the camera's intrinsic parameters are estimated through a closed-form solution for the camera's intrinsic matrix. The estimates are then refined through an optimisation such as those used in conventional approaches. Finally, distortion parameters are introduced and a full optimisation takes place. This method also introduces a practical 4D intrinsic matrix and distortion model which relate the indices of pixels to corresponding spatial rays. The source code for this method is publicly available from Dansereau's *Light Field Toolbox* for MATLAB.

Vaish et al. present a method that uses a plane plus parallax framework to calibrate large camera arrays [21]. Assuming all cameras lie on a plane parallel to the reference plane, camera positions can be recovered (such as in Ueshiba and Tomiba's approach [20]). This is achieved by measuring the parallax of a single scene point that is not on the reference plane. The light field can then be parameterised as a light slab (Levoy's two-plane parameterisation) [10]. Since the method assumes that all cameras are on precisely the same plane, and only calculates projection to a reference plane in advance of calibration, the accuracy is somewhat diminished for certain setups. This approach is therefore suitable for applications such as synthetic aperture photography, where planar cameras are commonly used.

4 Occlusion Removal

Occlusion removal has been demonstrated in a range of contexts and using a variety of methods. It has also been demonstrated on light fields, an area of particular interest. In this section we explore occlusion removal in these differing contexts, categorised by capture medium and device setup.

4.1 Applied to a sequence of images

Perhaps the simplest occlusion removal method is the one that deals with a sequence of temporally sparce images, where some images contain occluded segments. This may occur when a tourist takes a number of photos of a monument as passerbys come in and out of view. If all photos are occluded by passerbys at different locations, it is clear that to remove the occlusions, at least one image is required for each occluded segment, wherein the occluder is not present. This has been formalised and demonstrated to be executed automatically on sequences of images [6]. Visually pleasing results can be achieved without necessarily capturing images from precisely the same position and orientation, as long as stitching algorithms [18] are used to reconstruct the unoccluded image.

4.2 Applied to video from a single view

Occlusion removal can be achieved on still video (e.g. on surveillance cameras) by exploiting the temporal axis to perform *background subtraction* (also called *foreground detection*). A common background subtraction method for video data involves thresholding the error between estimates of images with and without occlusions. This generally involves computing a confidence level for each pixel with past and future frames. This allows a background model, and therefore an occlusion layer to be built. The numerous approaches to this problem differ in the type of background model used, and the procedure used to update the model [4, 12, 14, 23]. For example, Stauffer and Grimson's method models each background pixel as a mixture of Gaussians, updating them via an on-line approximation [14].

4.3 Applied to light field stills

An occlusion removal method specific to light fields involves exploiting focusing techniques via a synthetic aperture. Given enough views and a sufficiently wide synthetic aperture, focussing on a region of interest can effectively blur out occluders in the reconstructed image to the point that they disappear. Vaish et al's plane plus parrallax calibration technique shows improved occlusion removal results via synthetic aperture focussing, compared to results from metric calibration techniques [21].

Vaish et al. have also explored occlusion removal through 3D reconstruction, using cost functions that are robust to occluders. This has been shown to improve the occlusion removal quality of synthetic aperture focusing [22]. Although this technique improves on the results of ordinary synthetic aperture focusing, its performance drops significantly in the case of complicated or severe occlusion.

To overcome these issues with severe occlusion, Yang et al. consider occluded object imaging a problem of light ray selection from optimal camera views [25]. An optimal camera selection algorithm and greedy optimisation is used to to propagate visible ray information from depth focus planes. This approach leads to a much clearer reconstruction of occluded objects.

4.4 Applied to camera array video

For the case where an array of cameras is used (though not necessarily a light field camera or cameras even on the same plane), synthetic aperture focusing can be combined with object detection and tracking algorithms. Joshi et al. uses a straightforward approach built from existing single camera tracking algorithms, which tracks moving objects through severe occlusions [7]. The method tracks objects with up to 70% occlusion on all cameras via detection aggregation across views.

Similarly, Yang et al. describe a method to remove occlusions from video via object tracking and synthetic aperture focusing [26]. A synthetic aperture imaging system is used to model precise locations of objects in a controlled scene (see Figure 4). Yang's method also enables the seamless interaction among detection, imaging and tracking modules via a hybrid framework, and introduces an improved synthetic focusing method. However, its use is limited to controlled scenes such as in their setup.

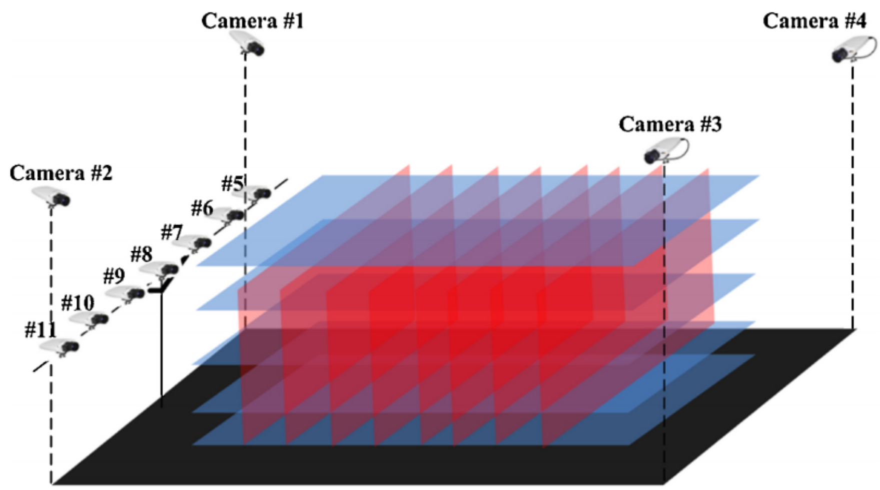


Figure 4: Hybrid synthetic aperture imaging system. Adapted from Tao Yang, Yanning Zhang, Xiaomin Tong, Xiaoqiang Zhang, and Rui Yu. A new hybrid synthetic aperture imaging model for tracking and seeing people through occlusion. *Circuits and Systems for Video Technology, IEEE Transactions on*, 23(9):1461–1475, 2013.

4.5 Applied to light field video

Research on occlusion removal on light field video is limited. However, it has been shown that some occlusion removal can be achieved through depth-velocity filtering. Edussooriya et al. have analysed the spectrum of a light field video corresponding to a Lambertian object with constant depth and velocity. They have showed that the object can be enhanced and modelled when occluded based on its depth and velocity via a 5D depth-velocity filter [3]. This method is applicable to a small set of applications, as it only works when objects are modelled at constant velocity and depth.

5 Conclusion

The aim of the project connected to this review is to develop and demonstrate an occlusion removal method on light field stills and particularly video data, that is more robust than the limited existing solutions. The camera setup available for the project requires an understanding of light field cameras and calibration methods in order to test and demonstrate findings. Additionally, as a novel increment or development is to be made in the occlusion removal research space, an understanding of existing and related methods is needed.

Important theoretical concepts such as the plenoptic functions, light field representations, and pinhole camera model have been presented. This led into an overview of camera calibration, including its purpose and motivation, discussions on radial lens distortion, and an outline of the transformations applied to translate from 3D world coordinates to 2D pixels. Existing procedures for single camera and light field camera calibration were then discussed and compared. Finally, occlusion removal was introduced along with its motivations and applications, before exploring work achieved in the area, which can be applied to sequences of images, single camera video, light field stills, camera array video, and light field video.

It is clear that current research in light field video is limited. The existing methods described are also relatively inflexible to more dynamic scenes. However, plenty of work has been done on occlusion removal in other contexts, and their concepts may be applied to produce a novel incrementation or new method. There is a great opportunity for a valuable addition to this area.

References

[1] Edward H. Adelson and James R. Bergen. The plenoptic function and the elements of early vision. In *Computational Models of Visual Processing*, pages 3–20. MIT Press, 1991.

[2] D. G. Dansereau, O. Pizarro, and S. B. Williams. Decoding, calibration and rectification for lenselet-based plenoptic cameras. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 1027–1034, June 2013.

[3] Chamira US Edussooriya, Donald G Dansereau, Len T Bruton, and Panajotis Agathoklis. Five-dimensional depth-velocity filtering for enhancing moving objects in light field videos. *Signal Processing, IEEE Transactions on*, 63(8):2151–2163, 2015.

[4] Nir Friedman and Stuart Russell. Image segmentation in video sequences: A probabilistic approach. In *Proceedings of the Thirteenth conference on Uncertainty in artificial intelligence*, pages 175–181. Morgan Kaufmann Publishers Inc., 1997.

[5] Janne Heikkila and Olli Silvén. A four-step camera calibration procedure with implicit image correction. In *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, pages 1106–1112. IEEE, 1997.

[6] Cormac Herley. Automatic occlusion removal from minimum number of images. In *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, volume 2, pages II–1046. IEEE, 2005.

[7] Neel Joshi, Shai Avidan, Wojciech Matusik, and David J Kriegman. Synthetic aperture tracking: tracking through occlusions. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8. IEEE, 2007.

[8] Hoonjong Kang, Elena Stoykova, Jiyung Park, Sunghee Hong, and Youngmin Kim. Holographic printing of white-light viewable holograms and stereograms. *Intech, Rijeka*, pages 171–201, 2013.

[9] Jean-Marc Lavest, Gérard Rives, and Jean-Thierry Lapresté. Dry camera calibration for underwater applications. *Machine Vision and Applications*, 13(5-6):245–253, 2003.

[10] Marc Levoy and Pat Hanrahan. Light field rendering. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 31–42. ACM, 1996.

[11] Jorge J Moré. The levenberg-marquardt algorithm: implementation and theory. In *Numerical analysis*, pages 105–116. Springer, 1978.

[12] Christof Ridder, Olaf Munkelt, and Harald Kirchner. Adaptive background estimation and foreground detection using kalman-filtering. In *Proceedings of International Conference on recent Advances in Mechatronics*, pages 193–199. Citeseer, 1995.

[13] Chester C. Slama, Charles Theurer, Soren W. Henriksen, and American Society of Photogrammetry. *Manual of photogrammetry*. American Society of Photogrammetry, Falls Church, Va, 4th edition, 1980.

[14] Chris Stauffer and W Eric L Grimson. Adaptive background mixture models for real-time tracking. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, volume 2. IEEE, 1999.

[15] Peter Sturm. *Computer Vision: A Reference Guide*, chapter Pinhole Camera Model, pages 610–613. Springer US, Boston, MA, 2014.

[16] Wei Sun and Jeremy R. Cooperstock. An empirical evaluation of factors influencing camera calibration accuracy using three publicly available techniques. *Machine Vision and Applications*, 17(1):51–67, 2006.

[17] T. Svoboda, D. Martinec, and T. Pajdla. A convenient multicamera self-calibration for virtual environments. *Presence*, 14(4):407–422, Aug 2005.

[18] Richard Szeliski and Heung-Yeung Shum. Creating full view panoramic image mosaics and environment maps. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 251–258. ACM Press/Addison-Wesley Publishing Co., 1997.

[19] R. Tsai. A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Journal on Robotics and Automation*, 3(4):323–344, 1987.

[20] Toshio Ueshiba and Fumiaki Tomita. Plane-based calibration algorithm for multi-camera systems via factorization of homography matrices. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 966–973. IEEE, 2003.

[21] V. Vaish, B. Wilburn, N. Joshi, and M. Levoy. Using plane + parallax for calibrating dense camera arrays. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 1, pages I–2–I–9 Vol.1, June 2004.

[22] Vaibhav Vaish, Marc Levoy, Richard Szeliski, C Lawrence Zitnick, and Sing Bing Kang. Reconstructing occluded surfaces using synthetic apertures: Stereo, focus and robust measures. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 2331–2338. IEEE, 2006.

[23] Christopher Richard Wren, Ali Azarbayejani, Trevor Darrell, and Alex Paul Pentland. Pfinder: Real-time tracking of the human body. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(7):780–785, 1997.

[24] Yichao Xu, Kazuki Maeno, Hajime Nagahara, and Rin-ichiro Taniguchi. Mobile camera array calibration for light field acquisition. *arXiv preprint arXiv:1407.4206*, 2014.

[25] Tao Yang, Yanning Zhang, Xiaomin Tong, Wenguang Ma, and Rui Yu. High performance imaging through occlusion via energy minimization-based optimal camera selection. *International Journal of Advanced Robotic Systems*, 10, 2013.

[26] Tao Yang, Yanning Zhang, Xiaomin Tong, Xiaoqiang Zhang, and Rui Yu. A new hybrid synthetic aperture imaging model for tracking and seeing people through occlusion. *Circuits and Systems for Video Technology, IEEE Transactions on*, 23(9):1461–1475, 2013.

[27] Zhengyou Zhang. A flexible new technique for camera calibration. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(11):1330–1334, 2000.