Removing occlusions from light field data
Literature Review

Ashley Stewart

October 28, 2016

Contents

1 Introduction **2**

2 Light fields and light field cameras **3**

2.1 The plenoptic function 3

2.2 The 4D light field 4

2.3 Pinhole camera model 5

3 Camera calibration **6**

3.1 The effects of radial lens distortion 6

3.2 Transforming 3D coordinates to pixels 7

3.3 Procedures for single camera calibration 9

3.4 Procedures for light field camera calibration 9

4 Occlusion Removal **11**

4.1 Applied to a sequence of images 11

4.2 Applied to video from a single view 11

4.3 Applied to light field stills 12

4.4 Applied to camera array video 13

4.5 Applied to light field video 14

5 Conclusion **14**

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.

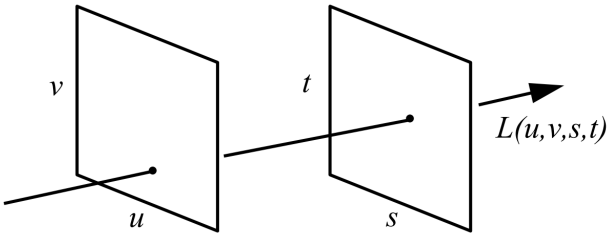

Figure 1: Levoy's light slab parameterisation. Adapted from Marc Levoy and Pat Hanrahan. Light field rendering. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 31–42. ACM, 1996.

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).

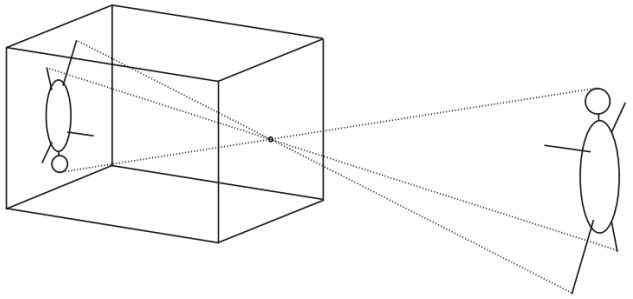

Figure 2: Pinhole Camera Model sketch. Adapted from Peter Sturm. *Computer Vision: A Reference Guide*, chapter Pinhole Camera Model, pages 610–613. Springer US, Boston, MA, 2014.

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.

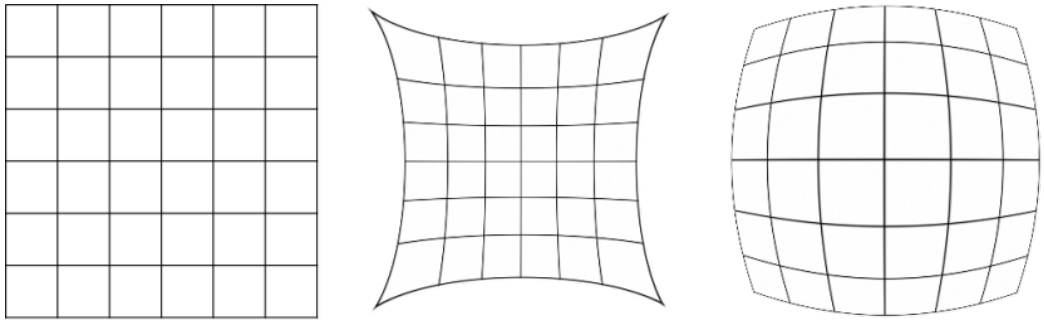

Figure 3: Representation of an undistorted grid (left), grid exhibiting pinhole distortion (centre), and grid exhibiting barrel distortion (right). Adapted from Hoonjong Kang, Elena Stoykova, Jiyung Park, Sunghee Hong, and Youngmin Kim. Holographic printing of white-light viewable holograms and stereograms.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

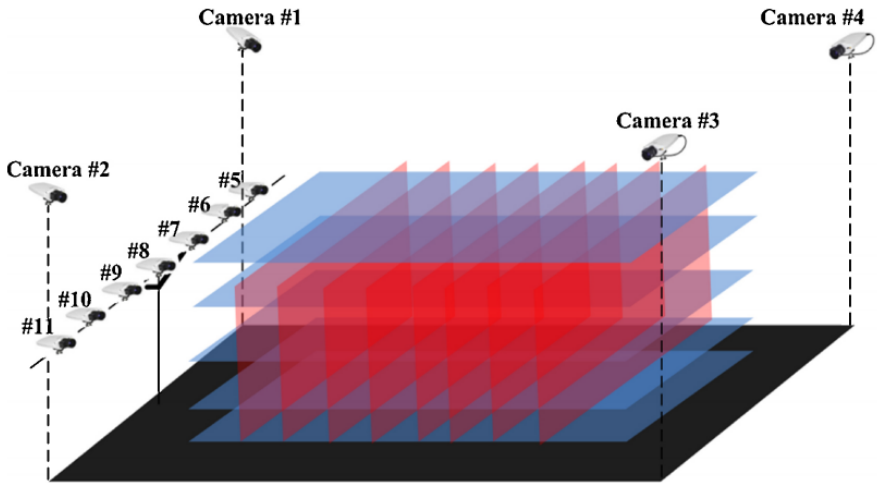

Figure 4: Hybrid synthetic aperture imaging system. Adapted from Tao Yang, Yanning Zhang, Xiaomin Tong, Xiaoqiang Zhang, and Rui Yu. A new hybrid synthetic aperture imaging model for tracking and seeing people through occlusion. *Circuits and Systems for Video Technology, IEEE Transactions on*, 23(9):1461–1475,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