ABSTRACT

Title of Document: Light field Analysis and Its Applications in Adaptive Optics and Surveillance Systems

Mohammed Ali Eslami,

Doctorate of Philosophy, 2012

Directed By: Professor, Christopher C. Davis, Department of Electrical and Computer Engineering

An image can only be as good as the optics of a camera or any other imaging system allows it to be. An imaging system is merely a transformation that takes a 3D world coordinate to a 2D image plane. This can be done through both linear/non-linear transfer functions. Depending on the application at hand it is easier to use some models of imaging systems over the others in certain situations. The most well-known models are the 1) Pinhole model, 2) Thin Lens Model and 3) Thick lens model for optical systems. Using light-field analysis the connection through these different models is described. A novel figure of merit is presented on using one optical model over the other for certain applications.

After analyzing these optical systems, their applications in plenoptic cameras for adaptive optics applications are introduced. A new technique to use a plenoptic

camera to extract information about a localized distorted planar wave front is described. CODEV simulations conducted in this thesis show that its performance is comparable to those of a Shack-Hartmann sensor and that they can potentially increase the dynamic range of angles that can be extracted assuming a paraxial imaging system.

As a final application, a novel dual PTZ-surveillance system to track a target through space is presented. 22X optic zoom lenses on high resolution pan/tilt platforms recalibrate a master-slave relationship based on encoder readouts rather than complicated image processing algorithms for real-time target tracking. As the target moves out of a region of interest in the master camera, it is moved to force the target back into the region of interest. Once the master camera is moved, a precalibrated lookup table is interpolated to compute the relationship between the master/slave cameras. The homography that relates the pixels of the master camera to the pan/tilt settings of the slave camera then continue to follow the planar trajectories of targets as they move through space at high accuracies.

LIGHT FIELD ANALYSIS AND ITS APPLICATIONS IN ADAPTIVE OPTICS AND SURVEILLANCE SYSTEMS

By

Mohammed Ali Eslami

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctorate of Philosophy
2012

Advisory Committee:
Professor Christopher C. Davis, Chair
Professor Gilmer Blankenship
Professor Rama Chellappa
Professor Thomas E. Murphy
Professor John T. Fourkas, Dean's Representative

Dedication

To my parents for keeping me in their prayers…

To my brother for granting me his support…

To my wife for giving me the final push…

And to Mohebban-e-Fatemeh (SA) for granting me purpose…

Acknowledgements

I begin this dissertation in the name of Allah (SWT) for all praise and thanks belongs to Him. He has truly showered my life with blessings and mercy.

I would like to next thank my adviser Dr. Chris Davis for his continued support, both through research fellowships and mentoring, that ultimately guided me to putting the appropriate "optics" touch on this project. I would like to thank the committee members for their participation in reviewing this thesis, in particular to Dr. Rama Chellappa who pulled my project out of a simple "calibration" oriented thesis to a more optics oriented look at camera systems during my proposal. This push along with Dr. Davis' expertise in the field helped pave the way for me to add my niche in the world of science and technology.

I also must thank TRX Systems for providing the funding to the Maryland Optics Group from the awarded DARPA project to make this thesis possible. I truly appreciate all the time and mentorship provided by the CTO and Chairman, Dr. Carole Teolis. I would also like to thank TRX engineers Dr. John Karvounis and Jared Napora particularly for their help in understanding OpenCV and guiding a MATLAB programmer through C# and then through C++.

I would also like to thank all the members of the Maryland Optics Group, in particular Dr. Quirino Balzano, for his help on all of my RF related projects, Dr. Navik Aggrawal for the opportunity to bounce ideas off of him and extend different areas in my thesis, and last but not least, John Rzasa who led all the hardware design for the high-resolution gimbals and assisted with the acquisition of the appropriate cameras and lenses we used in the experiments of this project. I can say with the

utmost certainty that had John not been around, my project would have probably been too expensive/difficult for any university student to be able to complete. I promised him a page long dedication for all his help in setting up experiments/hardware in my thesis but I'm sure this will suffice in him knowing that I am truly grateful for all his help throughout these past years particularly the evenings and weekends we would meet for experiments.

On a more personal note, I would like to thank my family for all of their support and push to even begin graduate school. My parents have always provided me with everything I needed to ensure that I could focus on my school work and not worry about anything else. My brother was a great listener at times when I was stuck and even though he rarely understood what I was saying managed to say one to two useful things to have that light-bulb in my mind turn on. In the last year of my graduate school, I was blessed to marry a wonderful wife who provided me with that final push/burst of motivation to ensure that I had the right focus to finish my thesis. Although TRX Systems was waiting for monthly reports, my wife was waiting for nightly reports every evening to ensure I completed enough throughout the day to be able to graduate. I would also like to thank her for drawing all of my figures.

Lastly, I end this section by acknowledging a group that I grew up with, the group mentioned in the dedication of this thesis, Mohebban-e-Fatemeh(SA). Watching the members of that group grow both in quantity (from 2 to 60 in five years) and quality always kept my head on straight to ensure that I never lost sight of my ultimate goal, and that was to become an accomplished, successful engineer to be able to help mentor /guide others along the path that I traveled. Thank you all.

Table of Contents

List of Figures

1 Introduction

1.1 Brief History

Too often in computer vision applications optical models for lens systems are over simplified to a common pinhole model. These models project a point in the world along a straight line to the imaging plane preserving the angle the ray makes with the optic axis at the input and output planes of a lens. In such systems, linear and angular magnifications are fixed to unity.

Complex lens structures, specifically large magnification zoom lenses are finding their way more and more into many applications in the computer vision field. PTZ camera networks for tracking objects at long rages, target identification, and handoff are the next step in camera network systems. Often in a camera system the more expensive parts are the optics involved to minimize aberrations and provide larger fields of views as well as angular magnifications. It then seems too rudimentary to model such structures as simple pinholes for the mapping between world to image coordinates.

Common papers in the area of PTZ camera networks claim that (1) (2) the cameras can be modeled as pinholes at each specific zoom setting. A look-up-table is formed and then interpolated for real time applications. But the question then arises as to how good their target localization accuracies actually are? Rarely do we find papers that compare the accuracies of their vision systems to the accuracies of a laser range finder for example. Since the models work, and are claimed to work reasonably well, is there any figure of merit that can provide a detailed look into the transformation of the rays that are coming into the camera system entrance pupil and leaving out at the exit pupil?

A figure of merit between these optical systems gives better insight into the different models that need to be used for various applications. For instance, FSO communications call for realignment at micro radian accuracies. Could vision systems aboard UAVs acting as a base station track multiple ships at sea to provide the accuracies for these links? If so, what accuracies are expected from the pointing errors due to a chosen optical model for the vision system? Is there a best possible pointing limit that can be achieved once a specific model is chosen?

This thesis explores these questions using the future of cameras, the Light Field or Plenoptic camera. Extending the model of image formation within these cameras to complex lens structures provides an insight into how optical models compare with one another. It provides the measurement criterion to be able to compare *apples to apples.*

1.2 Light Field Analysis

Adelson and Wang (3) in their paper introduce the plenoptic function as a periodic table for vision systems. The plenoptic function is a measure of irradiance along a light ray as it travels from a radiating source to a point of interest. Cameras that capture such light fields are only recently hitting the market via a small startup company called Lytro Technologies, founded by Ren Ng of Stanford University. Earlier work at Stanford captured light fields using an array of webcams that were synchronized to all take images of an object at the same time. Ng retrieved this ray information that is lost in conventional cameras by placing a micro lens array, whose *f/#* matches that of the objective lens of the camera, in front of the image sensor plane. A diagram of his camera is shown in Figure 1**.**

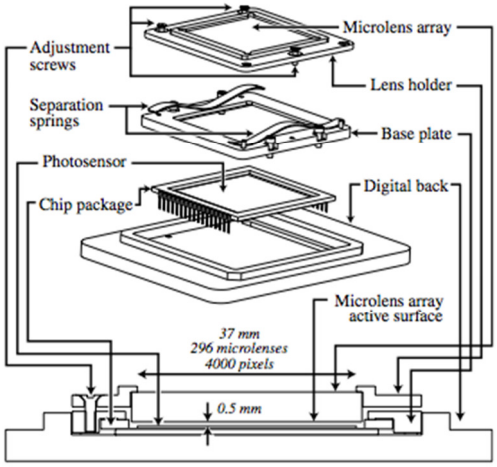


Figure 1: Plenoptic Camera schematic (4)

Retrieving the ray information of the incoming light coming into the camera system allows for multiple focus points by a simple mathematical change of bases in a post processing algorithm that can be run after the image is taken. That is, the image capture retrieves multiples depth of focus from a single image. In a conventional camera this would have required the photographer to take an image, change his zoom/focus, take another image, change zoom/focus, take another image, etc.

Wang (5) extended the work of light field analysis using simple paraxial optics to model image formation from a Lambertian object as it travels a distance *d* and passes through a thin lens of focal length *f*. He showed that by a simple change of basis, the light field at the sensor plane can be put in terms of the light field at the object using the paraxial optics matrices that are derived for simple systems. In addition, any apertures within the optical system simply modulate the light field as it passes through the structure.

This thesis extends the work of Wang to model image formation using paraxial optics for any complex lens structure. Vendors often do not provide the prescription of the lens that is used within their device. By modeling the lens as a black box and using the effective

focal length, back focal length, and front focal length, which are all parameters that must be provided by the vendor, one can model image formation for a light field as it travels through the lens structure. In addition, using such an analysis of the light field as it enters/exits a complex lens structure, a figure of merit is provided to compare that lens system to any other model for a camera. It finally resolves the issue of the optical equivalence of the pinhole model's usage for any lens system.

Applications for such systems are in adaptive optics for astronomical measurements of distortions in wave fronts as they travel along multiple channels from the source to image sensor; as well as surveillance applications for accurate tracking of a target as it moves through the field of view of a control camera. Simulations and experiments are conducted in both of these areas and are summarized in coming sections.

1.3 Adaptive Optics

Adaptive optics is a field that is primarily concentrated in astronomy to measure degradations in the wave fronts that are incident on optical systems and use structures such as deformable mirrors to correct for the distortions in wave fronts. Shack Hartmann sensors are constructed by micro lens arrays placed in front of a sensor plane to locally measure the distortions of incident waves from planar wave fronts. A wave is sampled through an array of pinholes before hitting a sensor plane and the micro lens elements simply add up the bundle of rays traveling into the sensor to provide a brighter image. Shack Hartmann sensors are limited to the angular resolution given by the *f/#* of the micro lens array that is used to sample the wave front at the aperture plane. Methods for increasing their angular resolution are of great interest in adaptive optics (6) (7) (8).

The principle of operation behind Shack Hartmann sensors is to be able to sample the input wave front and measure the distortions in the local wave front by measuring the energy vector normal to the curvature (9) of the wave at the point. The light field camera follows a very similar principle of operation. The micro lens array placed in the back of a plenoptic camera is meant to retrieve the ray information that is lost by a conventional camera. The major difference between these two (Shack Hartmann sensors and Light Field cameras) are the price where the former is on the order of $1,500, while the latter is on the order of $400 (depending on the size of memory). Could the Light Field camera then be used for applications in adaptive optics?

After a thorough literature review it was found that there is only one group (in Spain) that is looking into measuring wave front distortions using a Light Field camera. CAFADIS (9)is a plenoptic camera with interchangeable object lenses that achieves focal ratios on the order of *f/1.4* and *f/2.8* without being forced to place the micro lens array only micrometers from the image sensor plane.

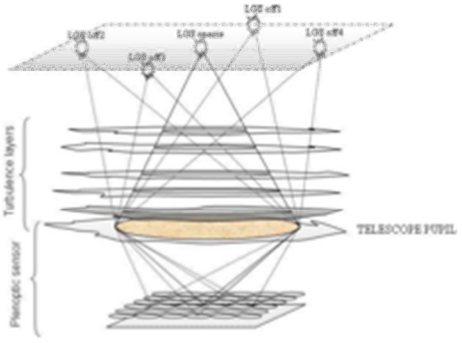


Figure 2: Plenoptic wave front sensor measuring beams from LGS passing through six turbulent atmosphere layers (9)

Sodium artificial laser-guide stars (LGS) at 90km ranges are used to obtain the system's reference wave front phase and optical transfer function. This is a dual to the Shack

Hartmann sensor by placing the micro lens array at the focal plane rather than at the aperture plane. The authors classify CAFADIS to be where Shack Hartmann and pyramid sensors meet. A pyramidal sensor consists of a lens relay and an oscillating pyramidal shaped prism to control the gain and sampling of the wave front sensor. In the pyramidal sensor, a prism is placed in front of an objective to form four images of the entrance pupil on the sensor plane. Adding these images together provides a brighter image for measurements, exactly like the plenoptic camera. In this particular case though, rays of light are not used for the measurements but more blurs measured from the oscillations of the prism.

Preliminary simulations for a novel technique using paraxial imaging to retrieve the input wave front using CodeV are presented in this thesis. Applying the same idea of light field analysis proposed in the previous section, knowing the light field at the image sensor plane can be inverted to find the input light field at the entrance pupil of the optical system. It is shown in this thesis that such a wave front reconstruction retains the accuracies of a Shack Hartmann sensor. Lytro has only recently released its Light Field camera (late February 2012) and experiments are going to be conducted to verify these simulations once a plenoptic camera is fabricated or a Lytro camera is obtained.

1.4 Surveillance Systems

As mentioned earlier, another application of light field analysis is in surveillance. Surveillance systems can be categorized into three areas:

1. Stationary object tracked by a stationary camera (no fun). This just amounts to finding the 3D position of an object from stereoscopy.

2. Stationary camera tracking a moving object, or, equivalently, a dynamic camera tracking a stationary object. The former is a surveillance system while the latter can be used as an inertial navigation unit.
3. Dynamic camera tracking a dynamic object (this is extremely difficult and the focus of this research).

Zoom lenses would increase the accuracy in any one of the three scenarios by focusing on an object of interest to allow it to take more pixel-space within the image. By increasing the focal length, we are thereby increasing the zoom, and will be able to track objects at far distances, or for that matter, objects that move to far distances, thereby keeping them within the field of view of the camera system over a long period of time. A simple solution to the third scenario of a surveillance system is provided by Badri et al, where a hybrid system consisting of a static camera with a wide field of view is used to look at a scene, which then sends coordinates of objects to a zoom camera that is free to pan, tilt, and zoom to focus in on the object of interest caught by the static camera. In other words, two cameras are used, one static and the other dynamic, with the dynamic one obtaining its target information from the static one, thereby adjusting the pan-tilt and zoom to focus on the object. In this paper, an automatic and autonomous solution is presented for a non-calibrated pair of cameras, by forming a look-up table (LUT) of static camera position with respect to zoom setting, of another LUT of zoom setting with respect to camera parameters. To determine the LUT to use for an object located in a specific region of the static image, use the first LUT to determine the zoom setting, and once that is achieved, use the second LUT (or an extension of the second LUT) to determine the parameters (intrinsic and extrinsic) to use for the camera. The goal is to find a function *f*, that maps

$$(p_{Z,}\, t_Z) = f(x_s,\, y_s,\, Z_j)$$

where $(p_{Z,}\, t_Z)$ are the pan and tilt settings at a specific zoom Z, and $(x_s,\, y_s,\, Z_j)$ are the static image coordinates in pixel units at zoom Z_j. The paper found the homography mapping between a number of feature points matched between the static dynamic images and set the pan/tilt of each node point in the static camera to the point at which it centers that pixel coordinate in the image of the dynamic camera. The experiment used 704 x 576 resolution, 42deg FOV, 26x optical zoom dynamical camera, and a 640 x 480, 90 degree static camera. Their gimbal resolution was 0.11 degrees. They were able to show that their accuracy was limited by this mechanical factor and received the following errors for their pan/tilt angles from tracking points on three different ellipse locations of which the angles were known (1):

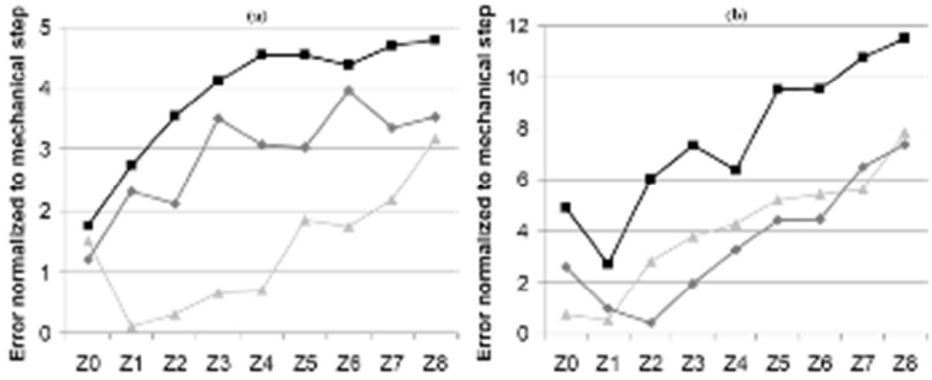


Figure 3: Error normalized to mechanical step for (a) pan and (b) tilt parameters for several zooms using (light gray) triangle (mid gray) diamond and (black) square to segment features

The paper does not provide any information on the depth of the target and the system is limited to the FOV of the static camera.

In contrast, Chen, Yao and Page interpolate a function that maps each pixel coordinate in the static image to a specific setting of the pan-tilt-zoom camera.

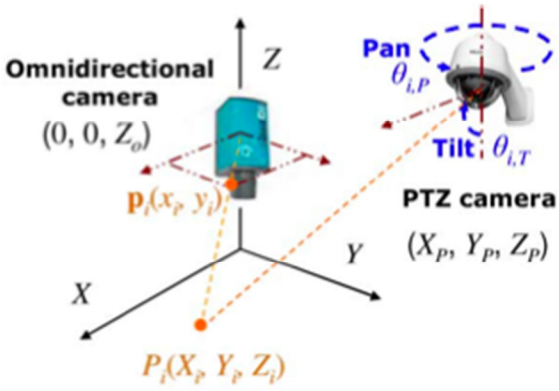


Figure 4: Typical setup for dual-camera surveillance system (10)

That is, they find functions

$$\theta_P = f_P(x_i, y_i) \qquad \theta_T = f_T(x_i, y_i) \qquad f_Z = f_Z(x_i, y_i)$$

so that when the target is in a particular region of the wide-field of view camera, the functions set the motors of the pan-tilt-zoom to focus in on that particular object. Once again, the area of interest is solely determined by the field of view of the static camera and once the target is out of sight from that image, then these functions will be rendered useless. Using an IQeye3 omnidirectional camera paired with a Pelco PTZ Spetra III SE dome camera, targets were placed between 0.2-5m from the cameras. Eight images at resolutions of 820x720 were taken of a planar 7x7 checkerboard. The results obtained showed that the larger the tilt angle, the higher the errors for the two calibrations methods that they tested, however the pan angle did not affect calibration errors (10).

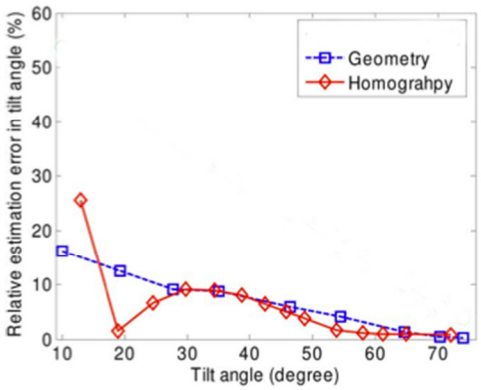


Figure 5: Comparison between methods based on relative errors in the estimated tilt angle (10)

An extension to this work is that of Chen, Yao, and Driria in which they calibrate two zoom cameras and interpolate a function between the two cameras parameters. Thus, from then on, all they need to do is calibrate one camera and send the coordinates of that camera to the function which will output the coordinates of the second camera. A block diagram of their system is shown below:

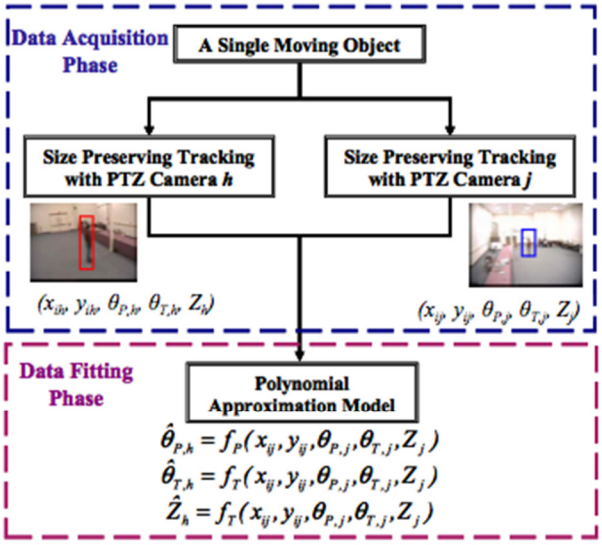


Figure 6: Cooperative mapping of multiple zooms block diagram (2)

They tested their results with two Pelco PTZ cameras of 640x480 resolution with 825 image samples uniformly distributed over their scene. The object is located at most 18m from the camera and as close as 3m (2).

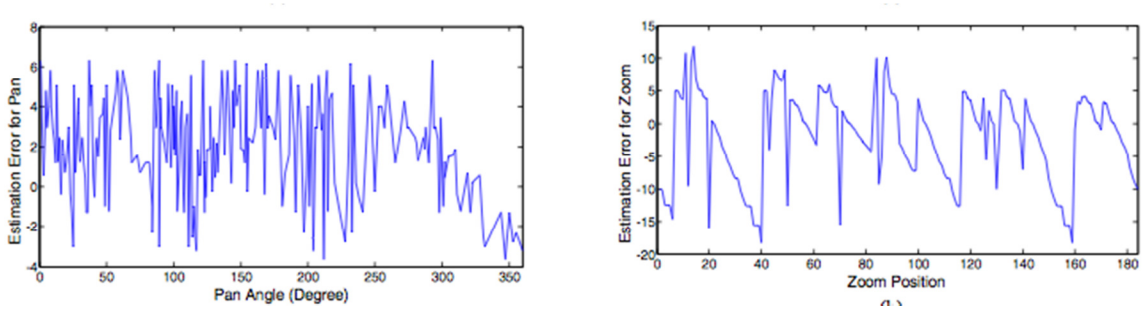


Figure 7: Error in Pan Angle and Zoom position using the mapping from one camera to the other (2)

Errors were retrieved by comparing their algorithm to that of Chen and Wang, who calibrate each camera independently by using known fixed objects within a scene.

This thesis explores an extension of the work by Chen and Wang for two dynamic cameras tracking a moving object. A similar transformation is found but accuracies are increased by using machine vision cameras that output uncompressed images, high resolution rotary platforms up to 0.00001 degrees, and zoom lenses with preset 16 bit encoders to accurately map the trajectory of a known target. A master/slave setup allows the master camera to keep the target within the center of its image and control the slave camera for accurate measurements at long ranges. Simulations show that by understanding the optics within the lens model increase the accuracies of target localization.

This thesis is organized as follows: Chapter 2 provides a complete background that covers all the different lens models that are used in optics. Chapter 3 uses light field analysis to compare the models presented in Chapter 2. Chapter 4 extends the use of

light field cameras in adaptive optics, while Chapter 5 talks about their use in surveillance systems. Chapter 6 then concludes the thesis with a summary of the novelties presented as well as future work that is to be conducted.

2 Review of Imaging Systems

2.1 Introduction

In this chapter we will introduce the different models that exist for imaging systems. An imaging system is a system that takes a 3D world coordinate to a 2D image plane. This can be done through both linear and non-linear transfer functions. In certain situations one is often preferable over the other. The most well-known models are the 1) Pinhole model, 2) Thin Lens Model and 3) Thick lens model of optical systems. The novelty presented in this chapter is the evolution of these models from one to the next. The thin lens model is an extension of the pinhole model and the thick lens is an extension of the thin lens model . Reduction to a small angle approximation involves linear transforms using Snell's Law of refraction. The pinhole model for an optical system is a linear transformation from homogeneous world to image coordinates, while the thin lens uses the radius of curvature and index of refraction of the material to bend the light towards the focal plane of the lens. The thick lens is an array of "thin lenses" all concatenated together to form any complex optical system that is desired.

We will derive all of the necessary equations that are consequences of each assumption and discuss the effect and applications of each model. The models include: 1) pinhole model of a camera 2) thin lens imaging system 3) thick/compound lens model.

2.2 *Pinhole Model for Optical System*

The most well-known model used in computer vision for camera calibrations is the pinhole model of the camera. This model assumes that the camera acts as a pinhole, which allows the rays coming from the object to pass through it and preserves the angle the ray makes with the optical axis.

2.2.1 *Introduction*

The perspective model of the camera is the simplest model to understand the formation of an image from an object.

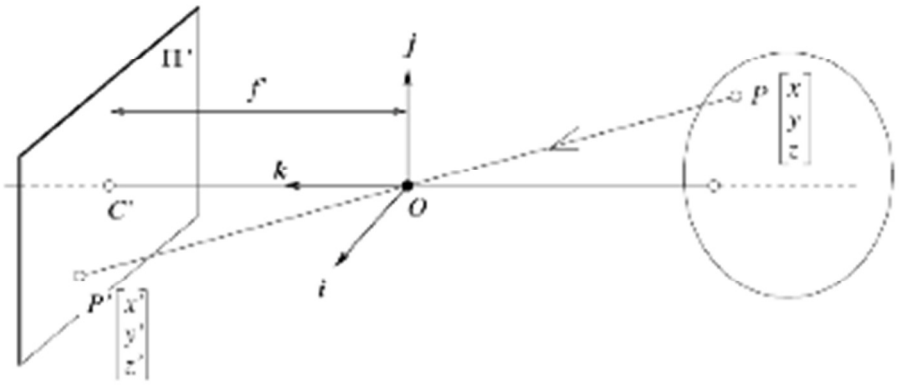


Figure 8: Pinhole model for an imaging system that maps a world point P to an image point P'

This model maps a point in space with coordinates P(x, y, z) in a straight line passing through the pinhole of the camera (the origin shown above) to its corresponding point in the image plane. From Figure 8, it seems that the image of the world will be inverted, but without loss of generality, the coordinate system can be moved behind the image plane at a distance *f*, and the mathematics remains unchanged.

The third coordinate in the image plane is always fixed to the focal length of the camera, and therefore provides no useful information and can be ignored. By use of

similar triangles and the preservation of the angle θ between the ray and the optical axis, it can be seen that the equation that relates the world point P to the two dimensional image point P' can be given by:

$$x' = f\frac{x}{z-f}$$
$$y' = f\frac{y}{z-f} \qquad 2.1$$

In matrix form, these equations can be written as in :

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = s\begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & -f \end{bmatrix}\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \qquad 2.2$$

where s is a constant that is adjusted (and depends on the distance z) to force the third coordinate to be equal to one. From the matrix equation above, the notion of homogeneous coordinates is introduced to define the mapping between world and image coordinates.

As the camera is translated and rotated about the world frame, the way it perceives the world coordinates changes. In addition, the model above ignored any skew that could be present in the CCD, the real image center, and any distortion that arise from the properties of the camera. The former rotations and translations are known as the extrinsic parameters of the camera, while the latter properties are known as the intrinsic properties of the camera.

To complete this calibration, a mapping is first found from the world coordinates to the camera coordinates. This mapping uses the extrinsic parameters to provide the world coordinates in the reference frame of the camera as its pose and movement are adjusted. The intrinsic matrix is then applied to the point that is now in camera coordinates to give the pixel coordinates of the point. This is outlined in Equation 2.3:

$$x = \mathrm{K}\,[\mathrm{R}|\mathrm{t}]\,\mathbf{X} = \mathrm{P}\mathbf{X} \qquad 2.3$$

$\boldsymbol{x}$ are the image-pixels, and $\boldsymbol{X}$ the world in homogenous coordinates, K represents the intrinsic matrix of the camera given by Equation 2.4:

$$K = \begin{bmatrix} f_x & \gamma & o_x \\ 0 & f_y & o_y \\ 0 & 0 & 1 \end{bmatrix} \qquad 2.4$$

and [R|t] is a 3x4 matrix composed of a 3x3 rotation around the three camera axes and t its translation. Calibration therefore calls for finding the matrix P, a 3x4 matrix that is the product of these two matrices. There are 11 degrees of freedom (focal length, pixel spacing, CCD skew, image x/y center, three rotations, and three translations), and thus 11 unknowns must be found, which means the matrix P is only found up to a scale factor, eliminating one of the unknowns. The two ways primarily used to find the matrix P are known as the planar homography and infinite homography solutions. The infinite homographies are beyond the scope of this thesis but follow the same principles after calibration.

2.2.2 Calibration of Imaging System Using the Pinhole Model

Camera calibration is defined as finding the extrinsic (rotation, translation) and intrinsic parameters (focal length, sensor skew, and distortion) of the camera that yield the mapping of the world coordinate to the image-pixel coordinates.

Planar calibration uses a planar target, like a chessboard, to form the world-image point correspondences to extract the necessary unknowns from the calibration. Well-developed toolboxes like OpenCV (11) and the Matlab toolbox by Bouget (12) find a chessboard within a scene based on Harris corners. The size of the chessboard is given to the algorithms to extract all corners that lie on a line in the pattern of a chessboard. Each corner is found and a 360 degree revolution of intensity variations around a region R of the gray scale image is analyzed. If there are four "adequate" intensity changes, then the corner is the corner of a chessboard. The toolbox then seeks to find MxN such variations that are collinear in the horizontal and vertical directions to identify the chessboard. Once the MxN pattern is found, a logic one is returned and the corners are saved in a matrix of image points.

A flexible technique for camera calibration is one that only requires the camera to observe a planar pattern in at least two different orientations; the motion need not be known (13). A checkerboard is used with its plane in the real world set to Z=0, forcing the 4x3 K[R|t] matrix into a 3x3 matrix H, which is known as the homography (14). The X and Y coordinates of the world points as shown in Figure 8, are determined by spacing the corners of the chessboard in a way where they lie from (0,N-1) on the y-axis, and from (0,M-1) on the x-axis. Note that OpenCV arranges its points in a way where the board is a parallelogram as shown in Figure 9.

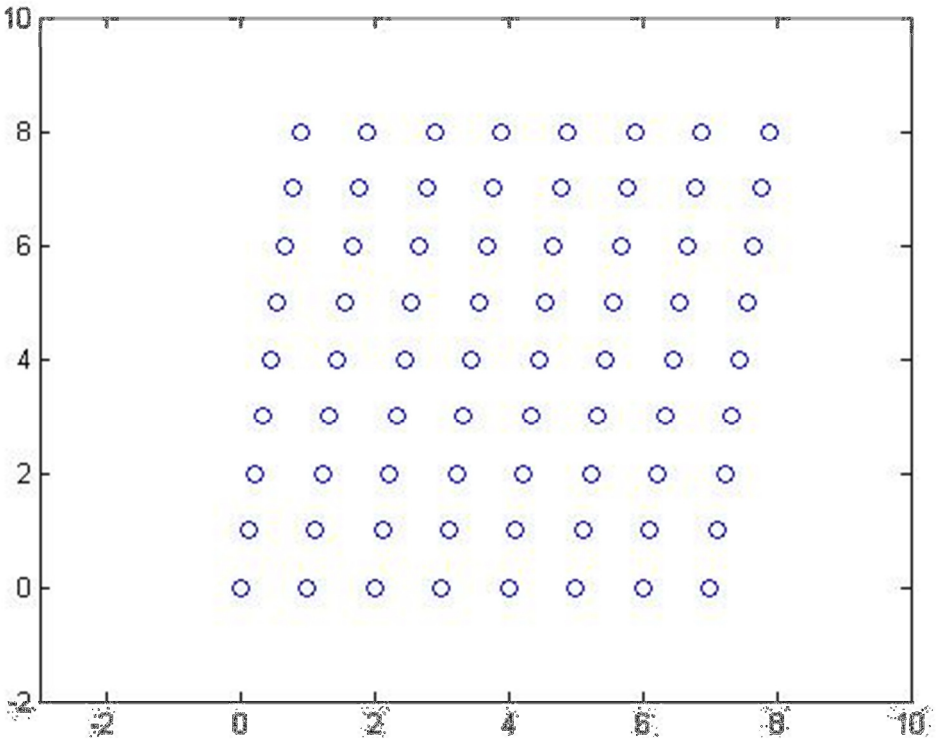


Figure 9: OpenCV corner arrangements of chessboard

Matlab's toolbox however differs slightly in that the world points are all equally spaced and kept within the grid [0,1]x[0,1]. In addition, all corners found by the Matlab toolbox are affine normalized to find the homographies.

After point correspondences are made between the world and the image domains, without loss of generality, we can set the Z=0 axis to be at the checkerboard, and thus eliminate three unknowns of the P-matrix (i.e. the third column of the matrix P) from our list of unknowns. We can call this new matrix H, and write Equation 2.5:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \qquad 2.5$$

where x' = x_1/x_3 and y'=x_2/x_3. Rearranging these equations yields Equation 2.6:

$$\begin{aligned} x'(-H_3-)X_W - (-H_1-)X_W = 0 \\ y'(-H_3-)X_W - (-H_2-)X_W = 0 \end{aligned} \qquad 2.6$$

which can then be written as the matrix equation in 2.7 :

$$\begin{bmatrix} -X_W^T & 0^T & x'X_W^T \\ 0^T & -X_W^T & y'X_W^T \end{bmatrix} \begin{bmatrix} H_1^T \\ H_2^T \\ H_3^T \end{bmatrix} = 0 \qquad 2.7$$

where H_i is the i^{th} row of the matrix H. These image to world pair coordinates are manually selected in the image, and since we know the grid spacing between consecutive patterns, we can map a standard grid with corners (0,0), (1,0), (1,1), and (1,0) to those pixel coordinates. With four of these correspondences, we would have enough equations to solve the system for the homography.

After finding the SVD of the corresponding matrix, the last column of V is the eigenvector that has eigenvalue equal to zero, which is the only vector in the null space of this matrix (up to a scale factor). Therefore, by imposing the condition that the third row of H must have a magnitude of unity, defines the unique homography that maps the world's coordinates to the image's coordinates.

Finally after finding the mapping between the world to image points, we will need to extract the camera parameters (the aforementioned intrinsic and extrinsic) from the matrix. To do so, we make note of the fact that images of orthogonal vanishing points are mutually orthogonal and thus can derive 2.8.

$$(K^{-1}v_1)'(K^{-1}v_2) = 0 \qquad 2.8$$

where v_1 and v_2 are known as vanishing points. A vanishing point is a point that is the image of an ideal point in the world (the point of intersection between parallel lines in homogeneous coordinates), where the ideal points are given by (1,0,0), (0,1,0), ½(1,1,0) and ½(1,-1,0). These ideal points are orthogonal and thus their images in the camera domain are also mutually orthogonal. To find these vanishing points, we simply multiply the aforementioned orthogonal ideal points to the matrix H that we have found form the world-image correspondences. It turns out that for these specific ideal points chosen, the vanishing points are simply linear combinations of the columns of H.

Assume first a skew and initial distortion of zero, and the principal point (intersection of the optical axis with the sensor plane) is in the center of the image. Denoting the vanishing points by $v_i = (a_i, b_i, c_i)$ gives the following two equations:

$$\frac{a_1 a_2}{f_x^2} + \frac{b_1 b_2}{f_y^2} + c_1 c_2 = 0$$
$$\frac{a_3 a_4}{f_x^2} + \frac{b_3 b_4}{f_y^2} + c_3 c_4 = 0 \qquad 2.9$$

which gives the initial settings of the horizontal and vertical focal lengths. If the spacing between pixel coordinates in the horizontal and vertical directions is the same, then the f_x and f_y will be the same.

These provide the initial guesses for the intrinsic matrix. Multiplying the inverse of the K matrix to the imaged corners of the chessboard gives us the chessboard in the camera reference frame. We compare this to a chessboard whose normal is parallel to the optical axis (the previously defined source points) to extract the initial rotation matrix.

With 11 unknowns, only six points are needed from a single image. But to choose the best matrix that fits the data, an optimization of the cost function 2.10.

2.10

provides more accurate results. Matlab's toolbox minimizes the cost function using the Gradient Descent method, while OpenCV uses a Levenberg Marquadt optimization to minimize the function. The standard deviations between the results provided reduce as you use more images, but as with all optimizations, their results do not change much after passing a certain threshold of images used, which was simulated to be about 30 images.

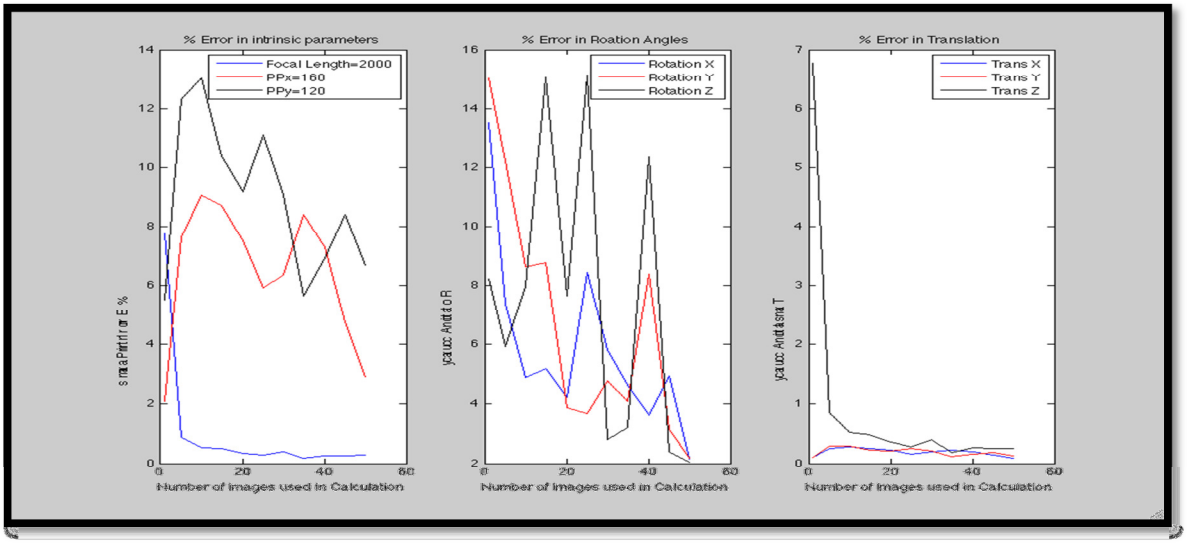


Figure 10: Relative percent error of intrinsic and extrinsic matrices vs. number of images used for each calibration

The OpenCV handbook recommends using at least 10 images of an 8x9 chessboard whose views take up over 50% of the image and views are not too similar to one another. This becomes rather difficult in calibrating a lens at far distances, as will be seen in Chapter 5.

2.3 *Thin Lens Model of Optical System*

Expanding the pinhole model to include more light rays gives the thin lens model of an imaging system. To be put plainly, the thin lens is the most simplistic lens whose index of refraction and radii of curvature determine the focal length of the lens.

2.3.1 *Introduction*

The thin lens is the simplest optical system that forces light rays to bend towards a focal point. The thickness of the lens is assumed to be infinitesimal so that light rays which hit the center of the lens and make an angle θ, with the optical axis has its angle preserved.

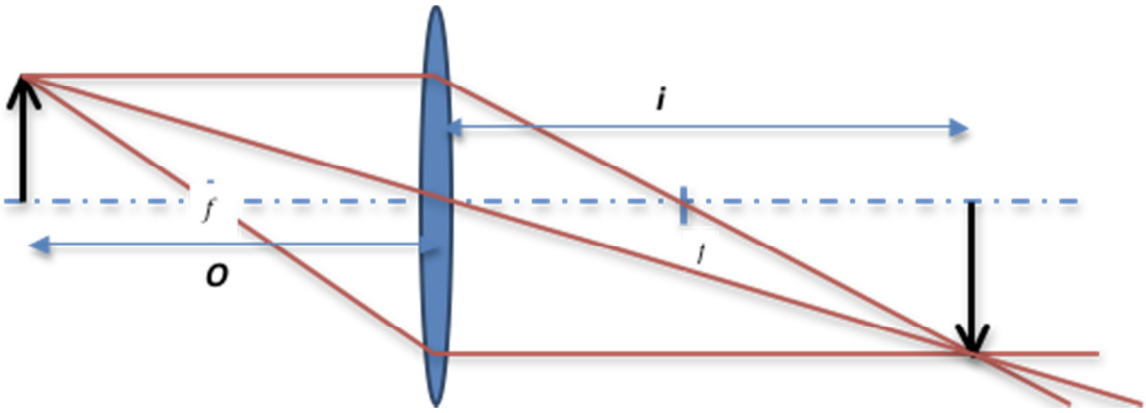


Figure 11: Thin lens ray tracing diagram showing chief (centered) and marginal (edge) rays

In all imaging systems, we have light travelling from one point to another. A ray of light is defined as a trajectory or path the light will take to go from Point A to Point B. As light travels from one medium to another, the speed of the wave changes, and as a consequence, the light ray bends at an angle dependent on the angle to which it enters the system. This is a consequence of phase matching in superposition of waves and also Fermat's principle for optical paths. The input and output bend angles are

related by Snell's Law of refraction which can be derived from Fermat's principle and is given by 2.11:

$$n_1 \sin(\theta_1) = n_2 \sin(\theta_2) \qquad 2.11$$

where the angles are measured from the normal of the surface that the light ray intersects. Let's take a close look at the principle of refraction at a curved surface so that we can derive the paraxial equation for a thin lens.

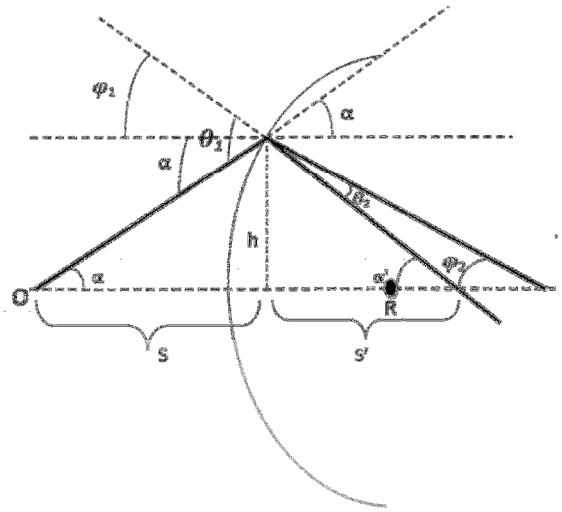


Figure 12: Derivation of thin lens equation by ray tracing at a spherical surface of radius R

First, by assuming that the angles are small enough to linearize the sine function, we can write Snell's Law as:

$$n_1\theta_1 = n_2\theta_2 \qquad 2.12$$

From the diagram above it is clear that the angles can be written as shown in 2.13.

$$\theta_1 = \alpha + \varphi$$
$$\theta_2 = -\alpha' - (-\varphi) = \varphi - \alpha' \qquad 2.13$$

Assuming the distance *v*, the distance from the lens to the image plane, to be small enough so that it can be ignored, we can write:

$$\alpha = \frac{h}{u_1}$$

$$\alpha = \frac{h}{v_1} \qquad 2.14$$

$$\varphi = \frac{h}{R_1}$$

Plugging all of the relationships given in 2.14 into Snell's Law gives,

$$n_1(\frac{h}{u_1} + \frac{h}{R_1}) = n_2(\frac{h}{R_1} - \frac{h}{v_1}) \qquad 2.15$$
$$n_1(\frac{1}{u_1} + \frac{1}{R_1}) = n_2(\frac{1}{R_1} - \frac{1}{v_1})$$

R is the radius of curvature of the surface, *s* is the object distance, and *s'* is where the ray will hit the optical axis. Every lens has to interfaces between itself and air, so if we were to model the opposite side of the lens with a radius of –R (standard sign convention dictates that concave bends have positive radii while convex ones are negative) and interchanging the places of n_1 and n_2 gives:

$$n_2(\frac{1}{u_2} - \frac{1}{R_2}) = n_1(-\frac{1}{R_2} - \frac{1}{v_2}) \qquad 2.16$$

Finally, using the property of a thin lens that says the image point from the first radius of curvature is the negative of the object distance to the second radius of curvature yields 2.17.

$$v_1 = -u_2 \qquad 2.17$$

and combining equations 2.16 and 2.17 gives:

$$\frac{1}{u_1}+\frac{1}{v_2}=\frac{n_2-n_1}{n_1}\left(\frac{1}{R_1}-\frac{1}{R_2}\right)$$

$$\frac{1}{u}+\frac{1}{v}=\frac{1}{f} \qquad 2.18$$

This is the famous lens maker's equation that describes rays coming from an object at a distance of *u*, and creating an image at *v* after passing through a thin lens whose focal length is *f*. Note that *f* can be negative depending on the signs of R.

2.3.2 Extension to the Pinhole Model

As a consequence of the lens maker's equation, we see two immediate differences between the thin lens model and the pinhole model of a camera.

Multiple light rays can now be considered from leaving a source point rather than only the light ray that passes through the center of the lens.

The point of image generation depends on the depth of the object. In the pinhole model, all points along a line, regardless of depth, would all image to the same point.

The second difference is of particular interest since now a source point can map to more than one image point. This will happen if all rays do not hit the sensor plane at the same location and will cause a blur in the image. Therefore, if the sensor plane can be set at the position where the image is in perfect focus, we can derive the depth of the point by solving the lens maker's equation.

Subbarao and Surya (15) were able to extend this concept further and use the blur of the image to extract a depth of field for the scene. Suppose the optical setup is the one shown in Figure 13.

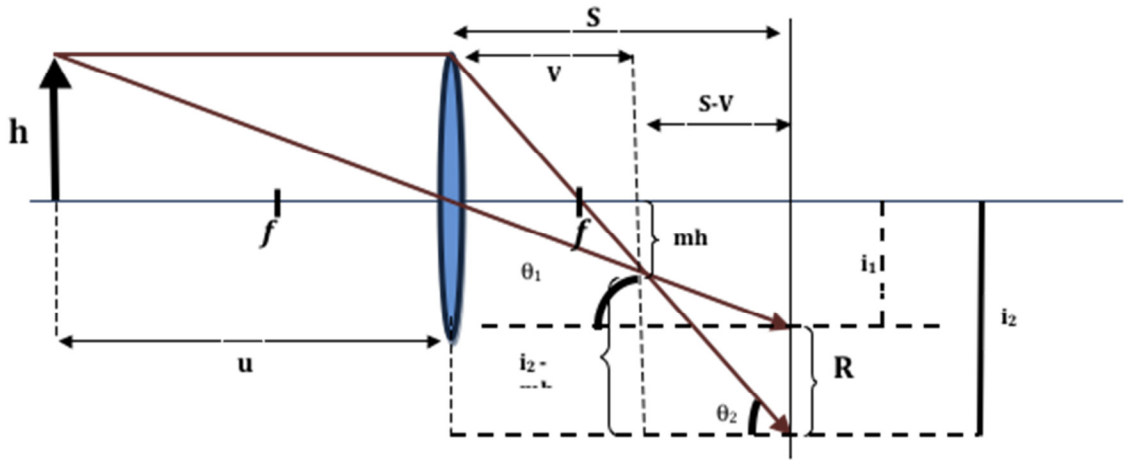


Figure 13: Thin lens ray tracing diagram used to derive depth u from blur radius R

One of the major challenges with depth from defocus is measurement of the blur within an image (16). Suppose the blur radius could be found using techniques described in (17) (18) (19). Using similar right triangles equation 2.19 gives an expression for θ_1:

$$\tan(\theta_1) = \frac{h + i_2}{s} = \frac{i_2 - mh}{s - v} \qquad 2.19$$

Solving the expression for i_2 yields:

$$i_2 = -\frac{h}{v}(-s + v - \frac{vs}{u}) \qquad 2.20$$

Again from similar triangles, we can write an expression for i_1:

$$\frac{h + i_1}{s + u} = \frac{i_1}{s} \rightarrow i_1 = \frac{hs}{u} \qquad 2.21$$

Using expressions 2.19, 2.20, and 2.21 we can define the blur radius R and solve for the depth u in terms of the blur radius R as shown in 2.22.

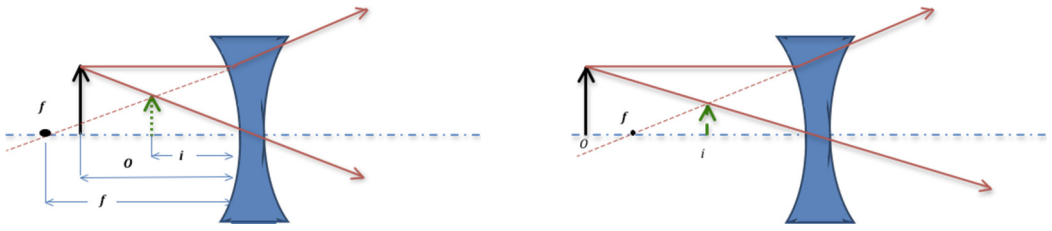
$$R = i_2 - i_1 = -\frac{h}{v}(-s + v - \frac{vs}{u}) - \frac{hs}{u} \qquad 2.22$$

Writing all of the expressions in terms of known quantities, u can now be solved for in terms of depth:

$$R = hs(\frac{1}{f} - \frac{1}{u} - \frac{1}{s}) \qquad 2.23$$

2.3.3 Positive and Negative Thin Lenses

To complete the discussion on thin lenses and their differences with the pinhole model, let's take a look at two primary types of thin lenses (positive power and negative power lenses) whose sign depends on the curvature, R, of the lens, see Figure 7. Biconcave lenses are lenses with a negative radius, and thus a negative focal length, while biconvex ones have positive radii and thus have a positive focal length. The index of refraction of the material of which they are made is always higher than that of air. For operation in the visible and near infrared regions of the spectrum they are primarily made from glasses or plastic whose indices vary between 1.4 and 1.8. A thin lens's power is measured in diopters and is the reciprocal of the focal length.



(a) (b)

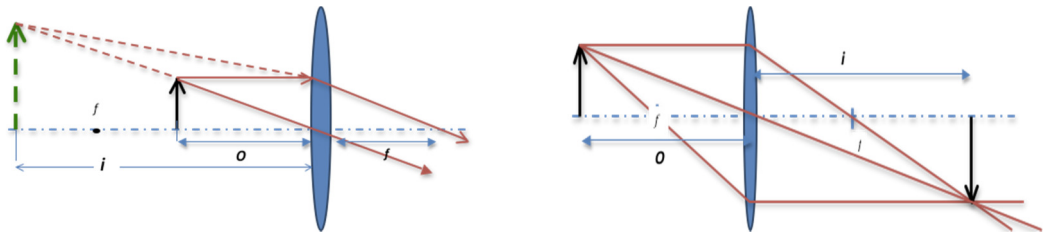


Figure 14: Ray tracing diagram for concave thin lenses with object (a) inside and (b) outside focal length, and convex lenses with object (c) inside and (d) outside focal length.

The ray diagrams in the figure are all derived from the lens maker's equation. We focus on three primary rays:

1. Any ray coming from infinity that is parallel to the optical axis ($u=\infty$) will have its ray pass through *f* on the imaging side. Most rays from distant objects fall into this category.
2. Any ray that passes through the point *f* on the object side ($u=f$) will emerge parallel to the optical axis on the imaging side ($v=\infty$)
3. Any ray coming from the object that hits the center of the lens, since we are assuming a thin lens, does not have its angle changed and thus is unperturbed on the imaging side. This is called the "chief ray."

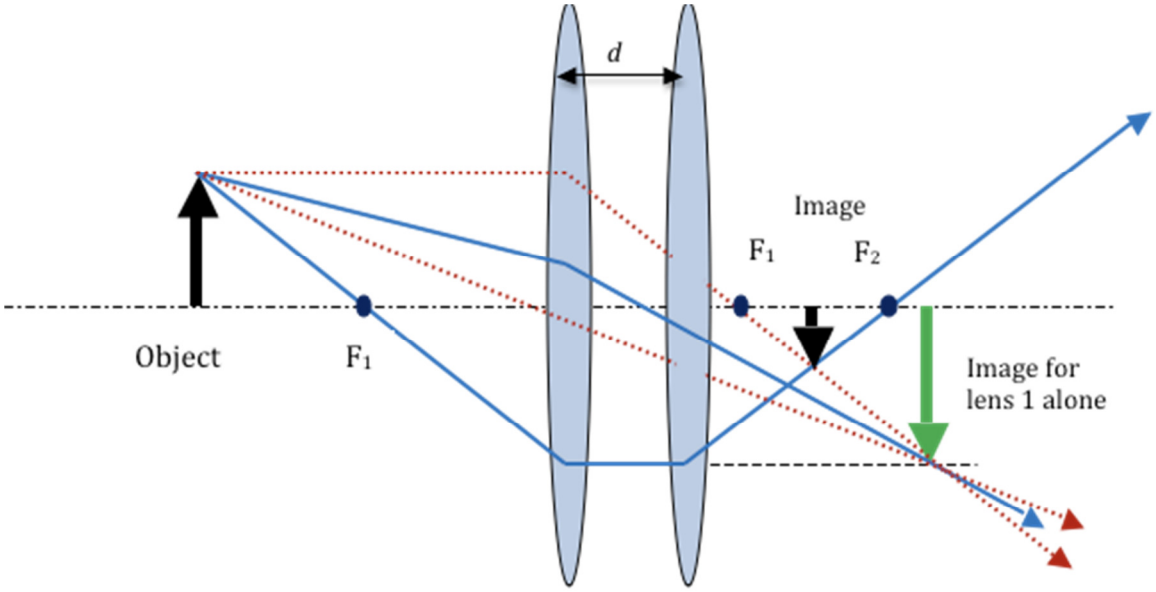


Figure 15: Combining thin lenses together to form an optical system

The last most simple thin lens setup we can look at is a combination of two thin lenses as seen in Figure 15. Often thick or compound lens systems such as zoom lenses can be modeled by varying the spacing between these two thin lenses. The light rays pass through lens A and an image is formed, which acts as the object that is imaged through lens B. The image of lens B is the image of the overall system. From the first half of this section, depending on the focal lengths of these lenses, we can have a combination of real and virtual images simply by changing the spacing in between the lenses. In addition, as with a zoom lens, the magnification is also changed based on the spacing between lenses.

The lens maker's equation for each lens can be written as:

$$\frac{1}{u_1}+\frac{1}{v_1}=\frac{1}{f_1} \qquad 2.24$$

$$\frac{1}{u_2}+\frac{1}{v_2}=\frac{1}{f_2}$$

The object distance u_2 and the image distance v_1 can be related together with the distance d between the two lenses as:

$$u_2 = d - v_1 \qquad 2.25$$

We are looking for a thin lens analog to this system, so we will assume that the distance between the two lenses is very small compared to the focal length, then the above expressions can be written as shown in 2.26:

$$\frac{1}{u_1} + \frac{1}{v_2} = \frac{1}{f_1} + \frac{1}{f_2} \qquad 2.26$$

From this we can see that at short distances the optical powers of the lenses simply add and the equivalent system can be modeled with a thin lens of that optical power. The lenses work together to keep the focus position unchanged as the magnification is varied. The change in the field of view results in a change of the magnification, which changes the intrinsic parameters of the camera. The goal is now to minimize the variations in aberrations and distortions over a range of settings rather than minimizing them for a single, fixed set of parameters. To do so, a new lens model needs to be analyzed, that of a thick or compound lens system, which provides the zoom capabilities.

2.4 Thick/Compound Lens Model

In the previous sections we saw the first extension of the pinhole model of a lens to a thin lens and were witness to a number of different properties in the model. The biggest factor was the fact that the thin lens model incorporated more than a single ray from an object point to the imaging plane. Thin lenses, however, condense all planes and points of interest in one location, namely, the location where the lens is

located. But thick/compound lens systems include more lenses all of which may not be categorized as a "thin" lens. This calls for a further extension of the model to incorporate the fact that our lens systems may be more complex and not include all thin lenses.

2.4.1 Introduction

In an imaging system, particularly one with a number of lenses as shown below:

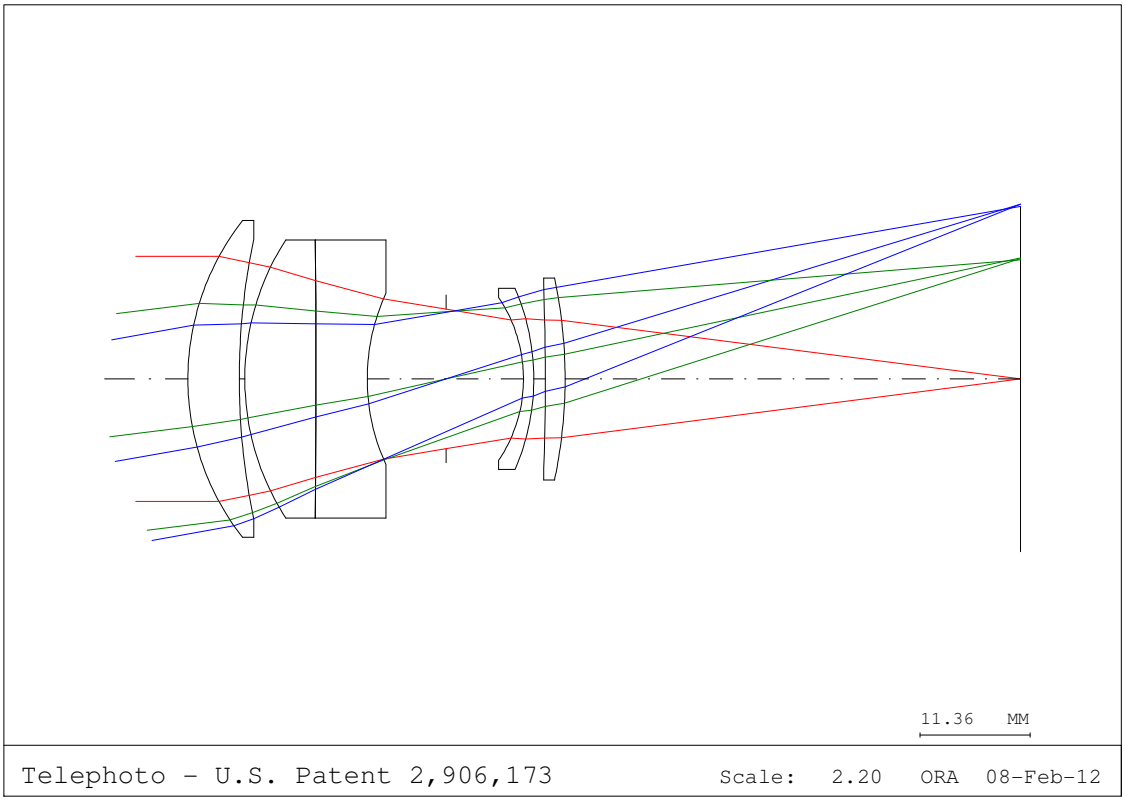


Figure 16: Telephoto zoom lens Patent 2906173 YZ-plane view

A number of questions arise:

1. What limits the amount of light that enters the system?
2. Since light is hitting a number of interfaces, how do we define focal length, object distance, and image distance?

3. Now that light is passing through a number of interfaces, it experiences many more bends, are there analogs to the rays that were described for thin lenses?

For a thin lens model, the answer to all of the questions above is, "The Thin Lens." All planes, distance measurements, rays, and amount of light entered, exited, and limited are all results of the thin lens. With thick/compound lenses, we are going to open up the model to find new entrance and exit planes, points of interest, and apertures.

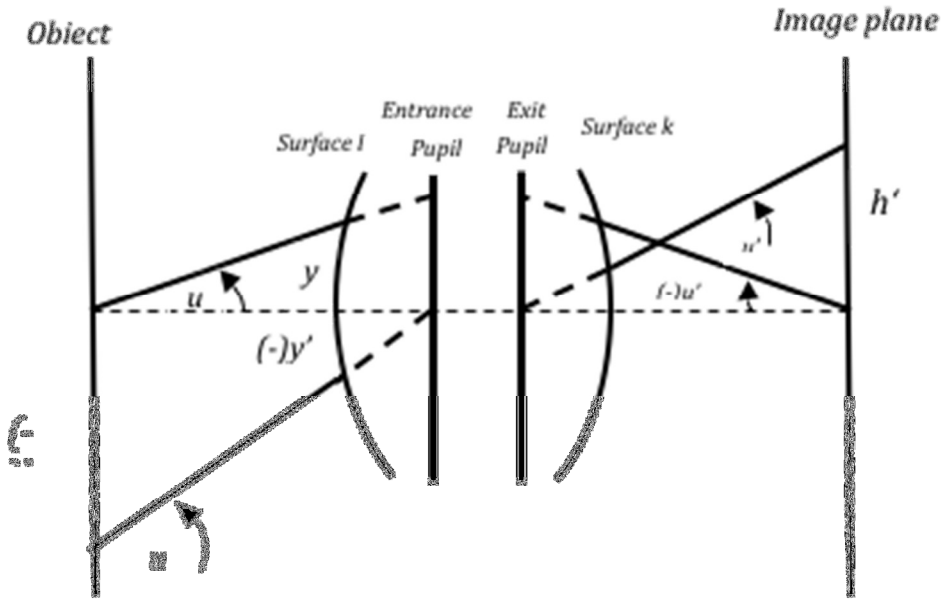


Figure 17: Black box representation of optical system

Every lens system can be analyzed as a black box where inside the box there are a number of different lens elements. The element within the black box that limits the amount of light the system gathers is known as the aperture stop. The image of this aperture stop through all elements that lead to the object is called the entrance pupil and the image of the aperture stop through all elements towards the image screen is known as the exit pupil. The two pupils are thus the two ends of the black box and

are the primary parameters that characterize the input-output relationship of the system.

In between the entrance and exit pupils, which define the input/output of the lens system, are two principal planes. These principal planes are the "front" and "back" of the lens, respectively. They are labeled P1 and P2 in the figure. These are the planes where the light rays bend towards the focal point if they are coming in parallel to the optical axis. They are hypothetical planes for refraction of light rays through the lens system. The locations of the principal planes are determined by the focal lengths of the system.

There are three focal lengths associated with every thick/compound lens system: a) Front Focal Length b) Back Focal Length c) Effective focal length. The front and back focal lengths are found by finding the intersection of the rays parallel to the optical axis as they come from the left/right hand sides of the optical system, and are measured from the left/right most lens intersection with the optical axis, respectively. The effective focal length of the system is the distance from the each principal plane to its respective focal length. That is, the principal planes are placed in a way where they are equidistant to the front/back effective focal points. Namely, the distance between PP1 and F1 is equal to the distance between PP2 and F2. This distance is known as the effective focal length of the system and is what is found by the calibrations of the pinhole model. Expressions for each of these focal lengths are derived in the next section using matrix optics.

2.4.2 Matrix Optics

We want a way to deal with more complicated optical systems. The law of refraction described in the previous section at surface interfaces is true for all surfaces and interfaces but it is too complex to analyze for each surface. Note that ray tracing program like CodeV or Zemax do exactly that to find the ray trajectories of an optical system. But in a paraxial ray analysis, where the angles of the incoming light are small enough to approximate Snell's law as 2.12 there is a linear system relationship between the input height/angle and the output height/angle. This section will derive a number of the common trajectories that take place.

First, denote the input rays as a 2D vector $\begin{bmatrix} y_i \\ \theta_i \end{bmatrix}$, where $i \in [0,1,...,N]$, denoting the surface height of incidence *y* measured from the optical axis, and the angle it makes with the normal as θ. Thus the linear system can be written as:

$$\begin{bmatrix} y_N \\ \theta_N \end{bmatrix} = M \begin{bmatrix} y_i \\ \theta_i \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} y_i \\ \theta_i \end{bmatrix} \qquad 2.27$$

Consider these matrices for two situations: a) Traveling through a medium of refractive index n and b) hitting a refractive surface of curvature R, can be used to describe any lens system. Then from these matrices, we can derive all the optical parameters discussed in the previous section (effective focal length, principal plane locations, front/back focal lengths).

A ray traveling in a space of refractive index *n*, as shown in Figure 18,

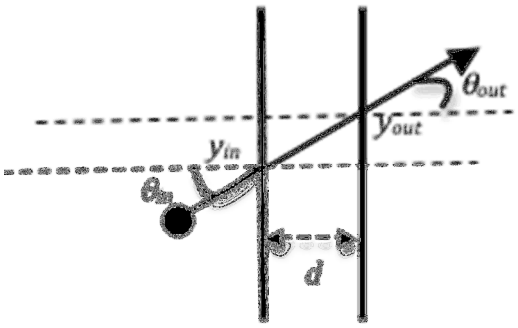


Figure 18: Ray traveling a distance d through media with refractive index n

forms a right triangle whose output height and angle is related to the input height and angle as:

$$\begin{aligned} y_{out} &= y_{in} + d\sin(\theta_{in}) \\ \theta_{out} &= \theta_{in} \end{aligned} \qquad 2.28$$

Linearizing the above equations and writing them as a system gives:

$$\begin{bmatrix} y_{out} \\ \theta_{out} \end{bmatrix} = \begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_{in} \\ \theta_{in} \end{bmatrix} \qquad 2.29$$

A ray that hits a refractive surface of radius of curvature R should be analyzed right at the surface as shown in Figure 19**.**

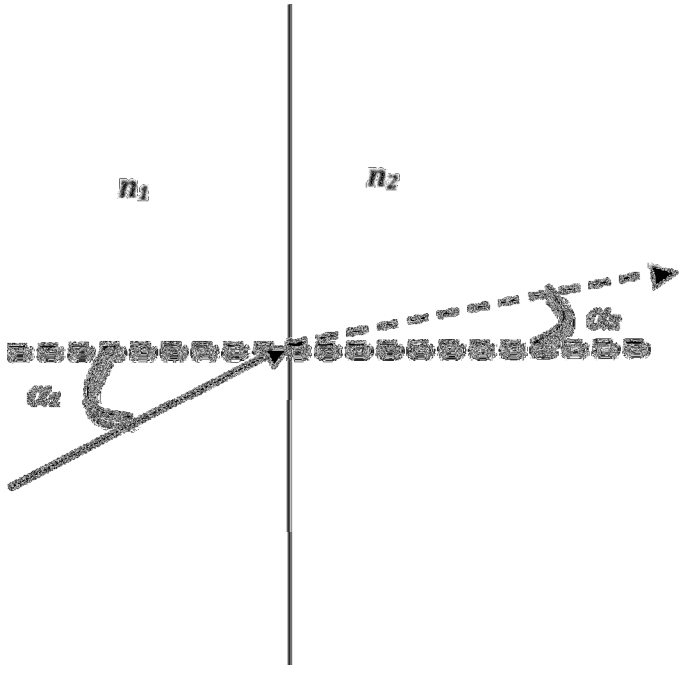


Figure 19: Refracting of a ray at boundary between materials of two different indices of refraction

The height immediately after the refractive surface remains unchanged:

$$y_{out} = y_{in} \qquad 2.30$$

The angles, on the other hand are related by Snell's Law, which from Figure 19 is given by:

$$n\alpha_1 = n\alpha_2 \qquad 2.31$$

where, $\alpha_1 = \theta_{in} + \phi_1$ and $\alpha_2 = \theta_{out} + \phi_2$. It is clear that $\phi_1 = \phi_2$, since they are alternate angles, and we can approximate it to be:

$$\phi = \frac{y_{out}}{R} = \frac{y_{in}}{R} \qquad 2.32$$

Substituting all of the above into Snell's Law an solving for θ_{out}, gives:

$$\theta_{out} = \frac{1}{R}(\frac{n_1}{n_2} - 1)y + \frac{n_1}{n_2}\theta_{in} \qquad 2.33$$

Putting the expressions for the height and angle in a linear matrix system gives:

$$\begin{bmatrix} y_{out} \\ \theta_{out} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \frac{1}{R}(\frac{n_1}{n_2} - 1) & \frac{n_1}{n_2} \end{bmatrix} \begin{bmatrix} y_{in} \\ \theta_{in} \end{bmatrix} \qquad 2.34$$

The radius *R* follows the same sign convention from the derivation of the thin lens equation.

These ray transfer matrices can now be used to derive equivalent optical matrix of any thick/compound lens system. Therefore, for the thick lens system shown in Figure 20,

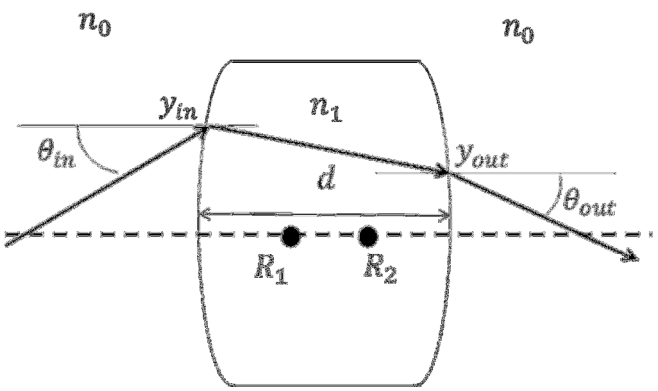


Figure 20: Thick lens system with radii and and thickness t

the matrix formulation is:

$$M_{tot} = M_{R_2} M_d M_{R_1} \qquad 2.35$$

where the rightmost matrix is the interface that the ray hits first and the matrices follow in order of incidents so that the final matrix can be given as:

$$\begin{bmatrix} y_{out} \\ \theta_{out} \end{bmatrix} = M_{tot} \begin{bmatrix} y_{in} \\ \theta_{in} \end{bmatrix} \qquad 2.36$$

Combining all of the expressions above for a thick lens yields:

$$\begin{bmatrix} y_{out} \\ \theta_{out} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \frac{1}{R_2}(\frac{n_1}{n_0} - 1) & \frac{n_1}{n_0} \end{bmatrix} \begin{bmatrix} 1 & \frac{d}{n_1} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \frac{1}{R_1}(\frac{n_0}{n_1} - 1) & \frac{n_0}{n_1} \end{bmatrix} \begin{bmatrix} y_{in} \\ \theta_{in} \end{bmatrix} \qquad 2.37$$

If we define,

$$D_1 = \frac{n_1 - n_0}{R_1}$$

$$D_2 = \frac{n_0 - n_1}{R_2} \qquad 2.38$$

€

as the optical power of the refracting surface then the matrices multiply out to 2.39:

2.39

Assume n_0=1 for free space and $d \rightarrow 0$ to give a thin lens. The C-element of the matrix (2,1) yields:

$$-(D_1 + D_2) = -(n_1 - 1)(\frac{1}{R_1} - \frac{1}{R_2}) = -\frac{1}{f} \qquad 2.40$$

Considering the appropriate sign convention taken, this is the exact expression of the focal length of a thin lens that was derived in the previous section. This yields that the transformation matrix of a ray traveling through a thin lens is given by:

$$M_{ThinLens} = \begin{bmatrix} 1 & 0 \\ -\frac{1}{f} & 1 \end{bmatrix} \qquad 2.41$$

As a matter of fact, any optical system can be defined by:

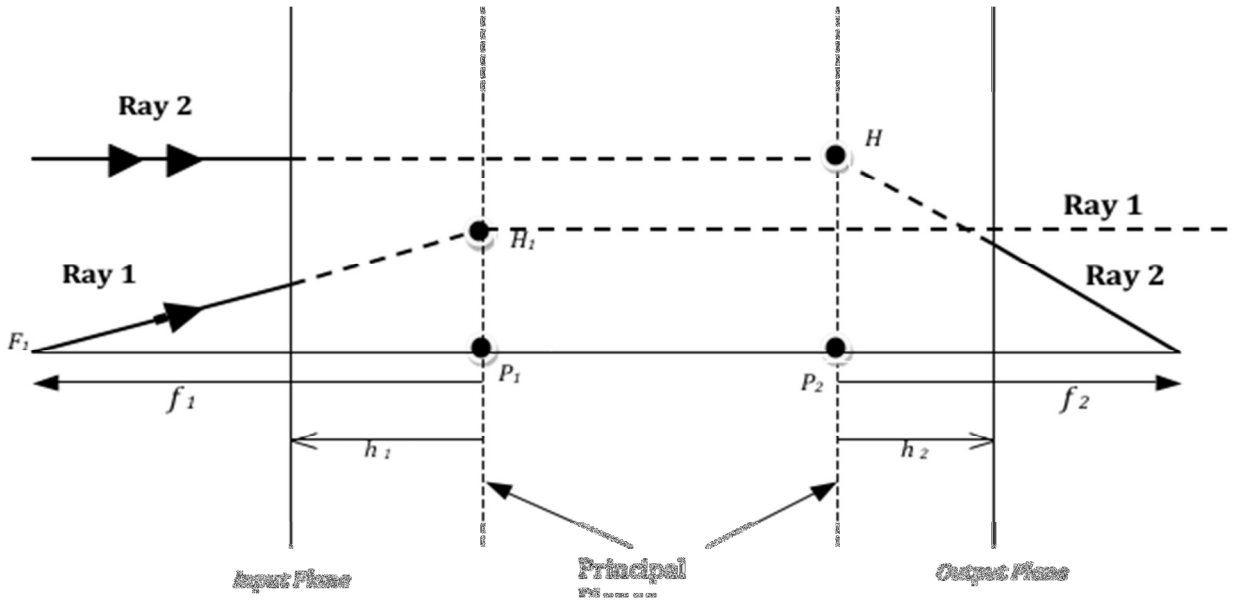


Figure 21: Black box of optical system used to find the ray transfer matrices between Rays 1 and 2 from input to output planes

The ray transfer matrix from the input to output plane can be written:

$$\begin{bmatrix} y_N \\ \theta_N \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} y_i \\ \theta_i \end{bmatrix} = \begin{bmatrix} 1 & h_2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} A' & B' \\ C' & D' \end{bmatrix} \begin{bmatrix} 1 & h_1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_i \\ \theta_i \end{bmatrix} \qquad 2.42$$

where h_i is the location of each principle plane and the primed letters denote the ray transfer matrix of the black-box optical system in between the input and output planes. From this expression we can derive the relationship between the black-box optical system and the overall ray transfer matrix to be:

$$\begin{bmatrix} A' & B' \\ C' & D' \end{bmatrix} = \begin{bmatrix} 1 & -h_2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} 1 & -h_1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} A - h_2 C & B - h_1 A - h_2 (D - h_1 C) \\ C & D - h_1 C \end{bmatrix} \qquad 2.43$$

Since the system above holds true for any input ray, Ray 2 does not have its height change as it travels through h_1 therefore $A' = 1$. Therefore, the location of the second principal plane with reference to the entrance pupil is given by:

$$h_2 = \frac{A - 1}{C} \qquad 2.44$$

In addition, we know that for Ray 1, the height should not change depending on the input angle, therefore B'=0. For the ray transfer matrices derived for refraction and translation we note that:

$$\det\left(\begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix} \right) = 1 \qquad 2.45$$

$$\det\left(\begin{bmatrix} 1 & 0 \\ \frac{1}{R}\left(\frac{n_1}{n_2} - 1\right) & \frac{n_1}{n_2} \end{bmatrix} \right) = \frac{n_1}{n_2} \qquad 2.46$$

Thus any multiplicative combination of these should have their determinants all multiply. It is clear then that if an imaging system is placed in a medium where the refractive index is the same on both sides, then when their determinants are all multiplied, the product should be 1. With the lens system in Figure 10 placed in air, we can derive the expression for the location of the first principle plane as:

$$h_1 = \frac{D-1}{C} \qquad 2.47$$

Lastly, from the paraxial ray analysis, Ray 2 has the property that

$$-\theta_{out} = \frac{y_1}{f_2} \longrightarrow C' = -\frac{1}{f_2} \qquad 2.48$$

Using Ray 1, this is shown to be:

$$-\theta_{out} = \frac{y_2}{f_1} \longrightarrow C' = -\frac{1}{f_1} \qquad 2.49$$

Since $C'=C$, we say that $f_1=f_2$, which is the effective focal length of the system. This completes the derivation of all of the unknowns from the previous section using Matrix optics.

2.4.3 Other Optical Considerations

After developing a complete set of models for lenses, it should be clear now that as light rays leave a source object and hit the lens system at different heights and angles, they may not all converge to the same point in the image plane, particularly if the image sensor is not placed at the location of best focus. The quantitative measure of spread of the ray intercepts on the image plane is known as the aberration. This spread of rays limits the spatial resolution of objects that can be captured with an optical system. In addition to the blur, rays coming from a line may not all converge

to a line in the image. These two problems bring forth the two primary aberrations of an optical system: Spherical Aberration and Radial Distortion.

In many optical systems, including the pinhole model, planar objects are projected as curved surfaces on the image plane (20). Mathematically (14), this can be written as

2.50

where *LI* is the distortion factor, *r* is the distance from the image origin to the pixel coordinate, x_i' are the coordinates as would be obtained if there were no distortions, and x_d and y_d are the distorted pixel coordinates as seen by the camera sensor array. A Taylor series of *LI* gives,

2.51

where k_i are the first five distortion parameters. Zisserman provides a method of correcting for the distortion by,

$$x_1' = x_c + L(r)(x_d - x_c)$$
$$x_2' = x_c + L(r)(y_d - y_c) \qquad 2.52$$

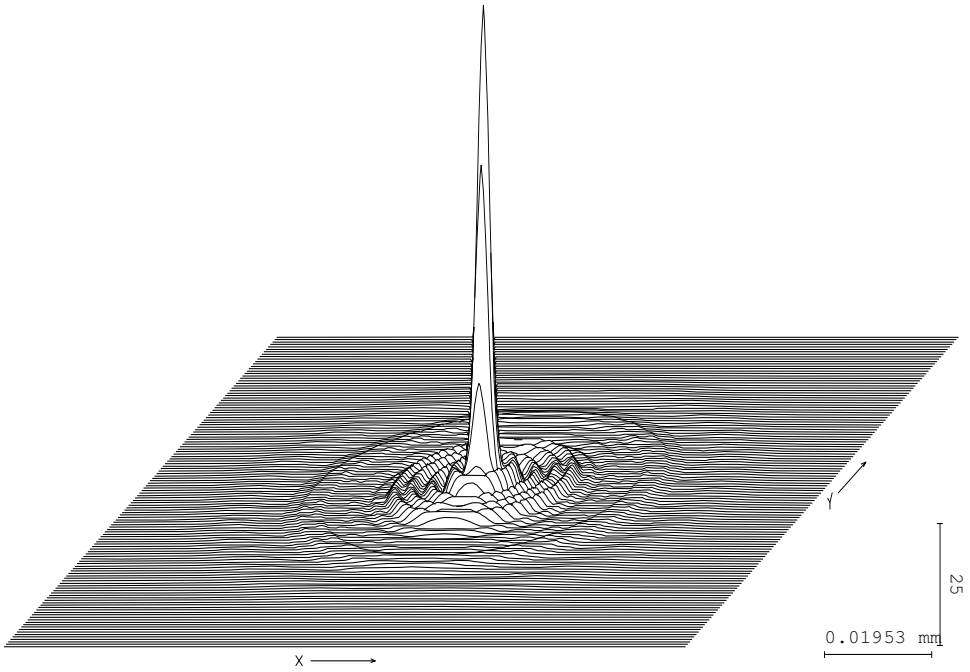
where x_c and y_c are the camera center coordinates. Using the MATLAB toolbox (12) the five distortion parameters that are provided are the coefficients of this Taylor series.

In addition to radial and tangential distortion, the response of an imaging system can add blur to an image of a scene. The imaging system can be described by its point spread function, or PSF. The PSF is the impulse response of the imaging system at hand since it describes what happens to a point source as it is image through the system. Every object can be thought of as multiple point sources that are imaged through a linear system, given the lens system has no nonlinear optical elements

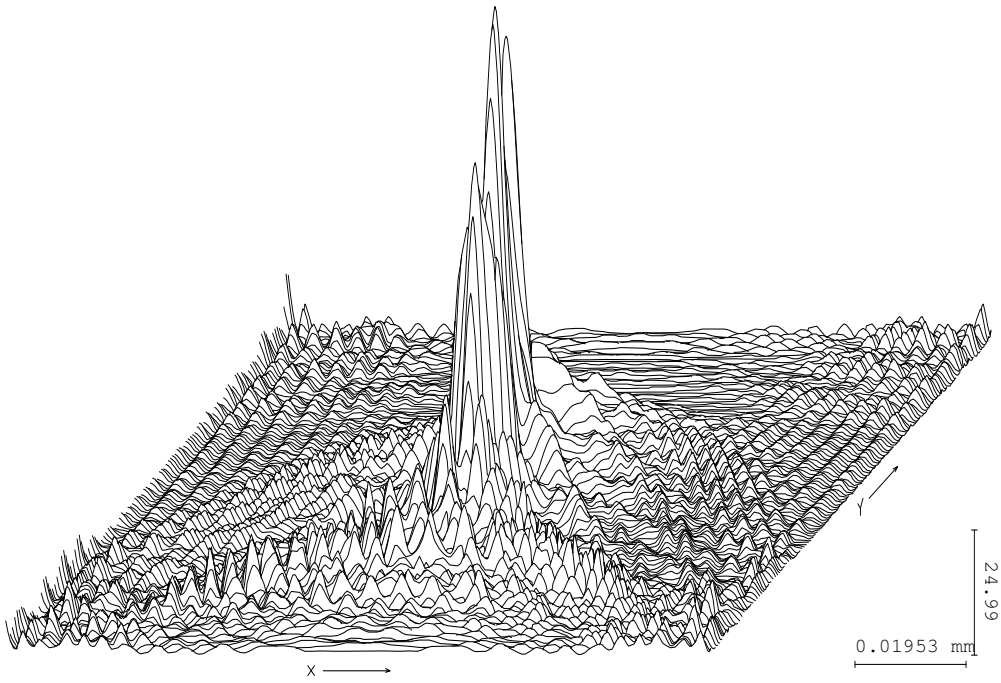
within it, that do not form a point on the sensor plane. With additive noise, η(x,y), any image g(x,y) can be described as:

$$g(x,y) = \int\int h(x,y;X,Y) f(X,Y) dXdY + \eta(x,y) \qquad 2.53$$

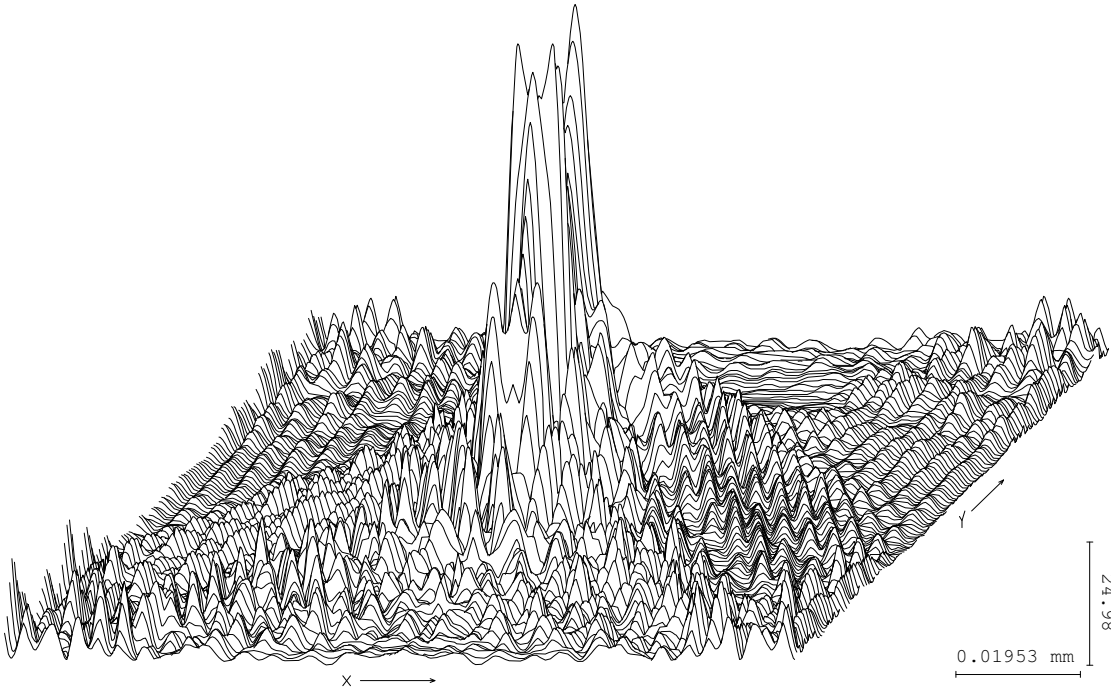
where *h(.)* represents the PSF, and *f(X,Y)* is the object intensity at spatial coordinates (X,Y). With a pinhole model of a camera, every point maps to a point, so *h(.)* is simply the delta function. With a thin lens model, the radius blur that was derived is analogous to the PSF, where a point in space created a blur on the image that was due to the different rays from the source focusing at different locations on the image plane. The PSF for the telephoto lens in Figure 22 is shown below.



(a)



(b)



(c)

Figure 22: Point spread function of telephoto zoom lens at (a) 0, (b) 7, and (c) 10 degrees

In general imaging system, the PSF is the linear response of the imaging system and it can be used to remove blur from the image. As is clear from Figure 15, the response is highly dependent on the input angle of the ray that hits the input surface, a fact overlooked by both thin lens and pinhole models. Its Fourier transform is known as the Modulation Transfer Function (MTF). The MTF describes the spatial resolution limits of an imaging system. It is the variation of contrast or visibility with spatial scale in the object. That is:

$$MTF = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} \qquad 2.54$$

where I_{max} and I_{min} are the maximum and minimum intensities of the image, respectively. Because of diffraction, if the entrance pupil diameter is similar in size to the wavelength of the light, fringe patterns are created on the sensor rather than a solid point. Thus every imaging system has a "diffraction limit" as shown in the MTF plots in Figure 23**.**

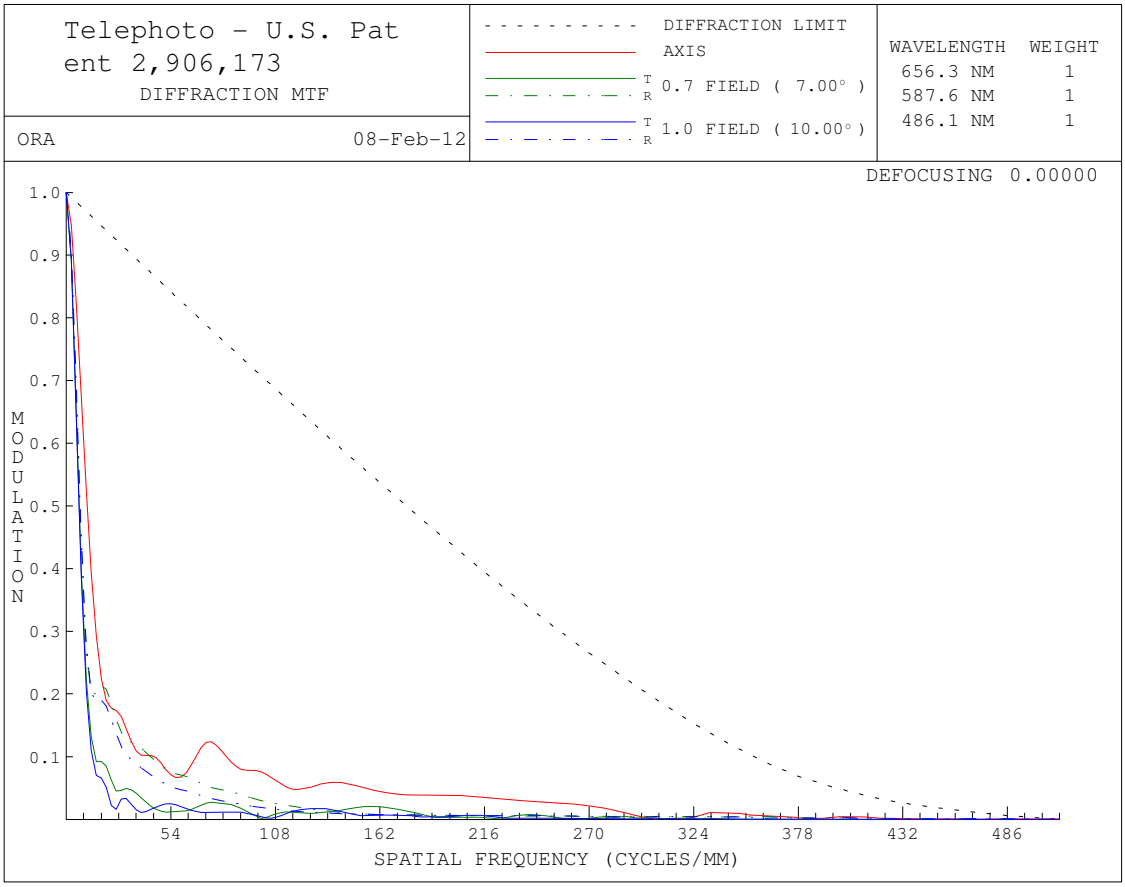


Figure 23: Modulation Transfer Function (MTF) of telephoto zoom lens

As an applicable example, a chessboard whose contrast changes from white to black for every check looks like a gray board depending on how it is imaged. The inability of the imaging system to describe such fine spatial resolutions of the object is described by the MTF.

2.5 Closure

This chapter summarized the evolution of different imaging systems. Vision problems can be divided into three parts.

1. 3D-world to 3D-camera coordinate transformations.
2. 3D-camera to 2D sensor transformations.

3. Algorithms performed to identify/classify certain features and/or objects within the 2D image.

The first part merely puts the camera's orientation as the reference frame and is not too interesting. The third part is where much of computer vision research is concentrated in finding different algorithms to identify certain features within an image: SIFT, SURF, HOG, Optic Flow, are all post-processing algorithms that only deal with the image and changes/features within it.

This chapter has focused on the development of the second part of the vision problems - a problem that is commonly overlooked and oversimplified, but is the bridge between the real world and the image. Accurately mapping pixels to real world metrics through these different imaging systems is of the utmost importance in surveillance and tracking systems. By understanding the evolution of these models new calibration procedures and mappings can be developed to provide better accuracy.

3 Light field Analysis

3.1 Introduction

The light field or plenoptic function is derived from the words in Latin *plenus* and *optic*, which means the complete optic function. The goal that Adelson and Wang (3) set out to accomplish was to form a periodic table where all *compounds* can be found as a combination of these vision elements. For example, a black and white image from a pinhole camera describes the intensity of a scene as viewed from a single point, at a single time, averaged over all the wavelengths of the visible spectrum. The function in its complete form can be described as a 6D function where any sample of this function is a photograph from a scene at a particular viewing direction. The plenoptic function is given by,

$$P = P(x, y, \lambda, t, V_x, V_y, V_z)$$

where (x,y) are the pixel coordinates on the sensor, λ the wavelength/color of the light, t the time of incidence on the sensor, and the triplet (V_x, V_y, V_z) the viewing direction the ray from the object makes with the normal of the viewer.

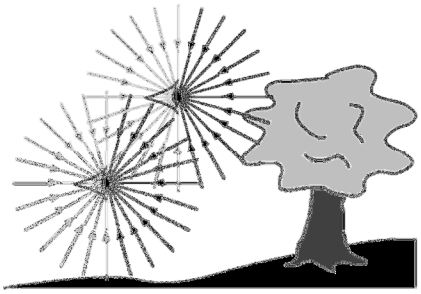


Figure 24: Conceptual look at Plenoptic function. Eyes represent sensor, rays are viewing angle to object. (3)

In its complete form, the plenoptic function describes a scene in space, viewed from any viewing angle, over all time and colors, at a 2D specific point on the sensor. Any sample of this function is a point on a color image at a specific view of the scene. Averaging over all colors, holding the scene static, and without loss of generality, always placing the viewer to be pointing along the z-axis, reduces the 6D domain of this function to a 4D function that can be described as,

$$P = P(s,t,u,v)$$

The quadruple (s,t,u,v) represents two points in space located on parallel planes separated by a normalized distance of unity. There exist different parameterizations of this function that are discussed later in the chapter.

This chapter focuses on deriving and capturing the light field as it travels through space and is incident on the three imaging system models described in Chapter 2. As far as the author has determined, only the pinhole and thin lens models have been used to find light fields (21) (5). Ren Ng (4) (22) has provided one of the most comprehensive works in capturing light fields with his plenoptic camera and has motivated their capture with complex/compound lens imaging systems but has not provided a clear derivation for modeling image formation with such lenses. Lian (5) has modeled image formation for pinhole and thin lens systems with matrix optics as derived in Chapter 2.

This chapter will expand on the concept of modeling image formation for more complex lens systems. Clearly, as light rays travel through different components of a complex lens systems with varying thicknesses and radii of curvature, the light rays bend in a number of different directions before actually hitting the image sensor

plane. In other words, the angle between the object and the optical axis is only preserved for the chief ray as it travels through the imaging system. This begs the question that often arises: Can complex lens structures be modeled as pinhole cameras? Kolb et al. have investigated this question by bringing to light a ray tracing algorithm to provide a camera model for computer graphics and have claimed that the pin-hole, thin, and thick lenses all differ in their optical properties and thus differ in the 3D to 2D coordinate transformations (23). However, in his thesis, Tordoff shows there are only minor differences between pinhole and zoom lens systems and uses a pinhole camera at each zoom to model his surveillance system (24). Using light field analysis, this chapter provides a figure of merit for the approximation that is often made to model complex lens systems as pinholes.

3.2 Mathematical Formulation

Light rays exist throughout all of space surrounding an object and if captured by a sensor without being occluded provide all the visual information that exists about the object. Light can only do one of two things:

- **Refract:** Bend as it passes from one medium to another leading to many illusions about the visualization of an object, like a bent straw as viewed from air to water.
- **Reflect**: Bounce off an object at the exact same angle of incidence to the local normal of the object.

If light is travelling in space and there are no occlusions, then there is no loss in electromagnetic energy as the light travels through space and all loss is only governed by the reflection coefficient of the object it hits.

3.2.1 Describing a Light Source and Reflecting Surface

The radiance of light is the amount of power per area radiated into a cone having unit solid angle and is measured in Watts/m2/steradian. Note that the area here is the area of the surface as seen by the source. The irradiance is the sum over the radiance provided by all sources and reflecting surfaces in a unit sphere, namely over all solid angles. This is measured by a CCD or CMOS sensor and digitized to assign the 8 or 16-byte value to a gray scale image.

Davis et al (8) provide the radiance analysis, when they look at coupling light from a source to an aperture. An object can be modeled as a series of point sources, where the image of the full object is the sum over all radiances from these point sources. In optics this property is also known as étendue, which characterizes the spread of light energy over a solid angle. A perfect optical system with reflective coatings to allow all light energy to pass, which most commercial camera lenses approximate, preserve the radiance as it travels through the optical system. This can be seen as a special case of the brightness theorem (8), which states the brightness of an image cannot be greater than the brightness of the object that was imaged. This is a consequence of the law of conservation of energy.

The light from a source and its trajectory to the camera sensors can be modeled as shown in Figure 25.

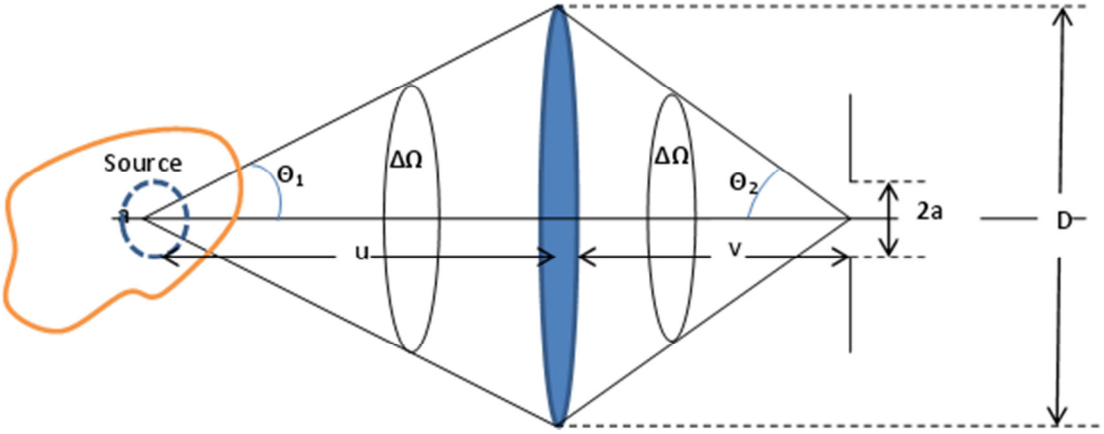


Figure 25: Geometry of an extended source and collection aperture used to describe the etendue (25)

The solid angle subtended by the source is:

$$\Delta\Omega_1 = 2\pi(1-\cos\theta_1)$$

where θ_1, is the angle the normal of the source makes with the edge of the object. For long ranges that are much greater than the entrance pupil of the lens, a circular aperture of radius a can have a solid angle approximation of:

$$\Delta\Omega = \frac{\pi a^2}{u^2}$$

where a is the entrance pupil diameter, and u is the range of the object to the entrance pupil. If the radiant power of the source is P watts then by the law of conservation of energy, the power at the receiver should be TP watts, where T is the transmission coefficient of the system between 0 and 1.

By its definition the radiance of this source can be given by,

$$L = \frac{P}{A\Delta\Omega}$$

where A is the area of the source. Suppose we have an source that is not faced with any occlusions all the way through the sensor plane. In other words, it is visible by the image sensor and is part of the image on the sensor plane. From the aforementioned conservation law of energy, we can write the powers in terms of radiance by,

$$L'(A'\Delta\Omega') = T(LA\Delta\Omega)$$

where the primes indicate the radiance on the sensor plane. Areas are solid angles are second order terms, thus they can be related to the source area/solid angle by the linear and angular magnifications as follows:

$$L'(m^2A)(m_a^2\Delta\Omega) = T(LA\Delta\Omega)$$

where, m and ma, are the linear and angular magnifications respectively. From geometrical optics, it is known that the product of the linear and angular magnifications is unity, and as a result we derive the brightness theorem to be,

$$L' = TL$$

When the light travels through a space where there are no occlusions, the brightness/radiance is preserved and there are no attenuations between the light source and the sensor, T=1, and thus $L'=L$.

The brightness theorem is of utmost importance in describing the plenoptic function and its application in image formation. The light source is described in terms of its radiance, where after traveling through any one of the lens models from Chapter 2, the final radiance pattern can be found simply by parameterizing the original plenoptic function as it is imaged through the lens system in the new basis of the plane of interest.

3.2.2 Parameterization of Light Fields

In this section we look at the different ways to describe the condensed 4D plenoptic function.

There are three common forms to describe a ray in space: 1) Plane Sphere Parameterization, 2) Two plane parameterization, 3) Two points on a unit sphere, all shown in Figure 26.

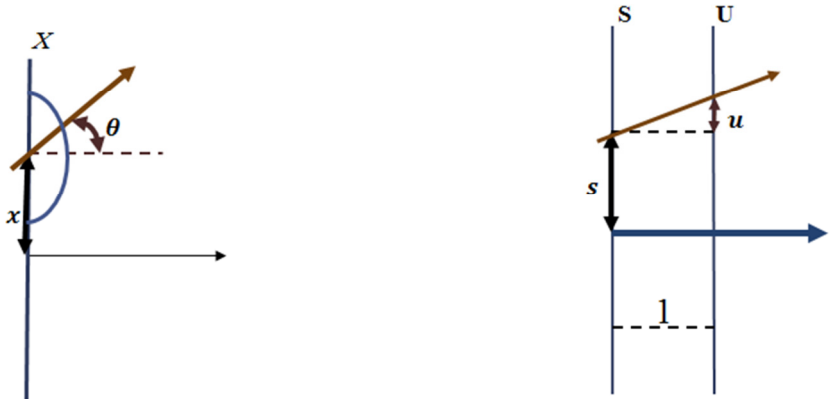


Figure 26: (Left) Plane angle and (right) Plane-plane parameterization of light field

The first describes a 2D point on some reference plane (without loss of generality, this can be assumed to be the xy-plane) and the pan/tilt angle that the ray makes with the normal of the plane. The 4D plenoptic function is then described by:

$$P = P(x, y, \theta, \phi)$$

The two plane parameterization, which is more commonly used, and will be used throughout this thesis, describes the point of intersection of a ray on two planes that are separated by a distance of unity between their normal vectors. The ray intersects the first plane at (s,t), which represent the lens coordinate of the lens on the lens plane, and the second one at (u,v), which represents the pixel locations on the sensor plane, to have the plenoptic function described as:

$$P = P(s,t,u,v)$$

The third parameterization is not used as often as the previous two, but for the sake of completeness is mentioned here. The parameterization is like longitude and latitude coordinates on a unit sphere and is given by:

$$P = P(\theta_1, \varphi_1, \theta_2, \varphi_2)$$

Depending on the application, it may be useful to use a certain parameterization. For instance, to describe a point light source, the initial trajectory is described by the plane-sphere parameterization, while the two-plane parameterization is used to describe the trajectory of the ray as it hits the camera system. As a look back, the two-plane parameterization can be scaled and modified to be thought of as a compound/thick lens system from the entrance to the exit planes of a camera system. Lastly, the points on the unit sphere are primarily used for stochastic process modeling to be able to uniformly choose two points along a unit sphere thereby forming a uniform probability distribution for any ray to be selected.

The Stanford Multi-camera Array, which will be discussed later in capturing light fields, uses the two-plane parameterization to build a camera array system making use of an array of pinhole cameras. Essentially, the two plane parameterization can be thought of as a series of pinhole cameras placed at the first plane whose pinhole coordinate is given by (s,t), each of which having an MxN pixel array sensor given by coordinates (u,v). Clearly, the quadruple (s,t,u,v) defines two points in space, and thus a unique ray that travels through the both of them, as shown in Figure 27**.**

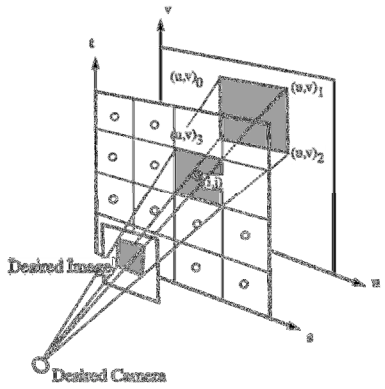


Figure 27: Light rays traveling through camera (s,t) and hitting pixel (u,v) (26).

As shown by Liang, et al (5), these parameterizations combined with the geometrical optics principles can be used to model image formation on a sensor plane. Most papers on light fields (21) (27) only assume pinhole models to simplify the (s,t,u,v) parameterization for modeling image formation. Liang (5) and Ng (28) expanded that to include different size apertures and thin lenses in their work with geometrical optics. Lastly, Ren Ng (4) of Stanford University began an analysis of more complicated imaging systems used as the main lens but did not incorporate matrix optics in their formulation. The next section will expand on their formulation for modeling image formation.

3.2.3 Modeling Image Formation

Liang and Shih (5) in their work have provided a unified framework in light field analysis for modeling image formation. Their work, as mentioned before, was motivated by Ren Ng who completed his image formation analysis by brute force, which is tracing each ray individually and following the radiance pattern along that ray (4). Liang and Shih used the basic principles of matrix optics in evaluating thin

lenses and traversing distances to evaluate the transfer of the light field through imaging systems as a whole rather than ray-by-ray.

From the brightness theorem, if there are no occlusions between the observer and the source, the radiance at the sensor plane must be equal to the initial radiance leaving the source, only different by the transformation between parameterizations. Mathematically, for a 2D slice of the 4D hyper plane that describes the plenoptic function, this is written as,

$$P_1(\begin{bmatrix} s_1 \\ u_1 \end{bmatrix}) = P_0(\begin{bmatrix} s_0 \\ u_0 \end{bmatrix}) = P_0(T_1\begin{bmatrix} s_1 \\ u_1 \end{bmatrix})$$

where the subscripts indicate the stage in the imaging system. For an imaging system with N stages, from the associative property of the multiplication of matrices this can be extended to the following result:

$$P_N(\begin{bmatrix} s_N \\ u_N \end{bmatrix}) = P_{N-1}(T_N\begin{bmatrix} s_N \\ u_N \end{bmatrix}) = P_{N-2}(T_{N-1}T_N\begin{bmatrix} s_N \\ u_N \end{bmatrix}) = \ldots = P_0(T_1T_2\ldots T_{N-1}T_N\begin{bmatrix} s_N \\ u_N \end{bmatrix})$$

The key to this analysis is to identify the transformation T_i from one side of the imaging system to the next. It is easily recognizable that this is merely a change of basis for the coordinates of the plenoptic function. This fact is a direct consequence of the brightness theorem discussed in the previous section. From this, we can formulate the following theorem:

Theorem (5)**:** The plenoptic function at any point in space from any viewing angle can be put in terms of the plenoptic function at the object with a change of basis of its arguments to reparameterize its coordinates to the coordinates of interest.

Proof: It is clear that an imaging system with one stage satisfies:

$$P_1(\begin{bmatrix} s_1 \\ u_1 \end{bmatrix}) = P_0(\begin{bmatrix} s_0 \\ u_0 \end{bmatrix}) = P_0(T_1\begin{bmatrix} s_1 \\ u_1 \end{bmatrix})$$

where T_1 is the ray transfer matrix between stage 0 and stage 1. Assume it is true that for an N stage system:

$$P_N(\begin{bmatrix} s_N \\ u_N \end{bmatrix}) = P_{N-1}(T_N\begin{bmatrix} s_N \\ u_N \end{bmatrix}) = P_{N-2}(T_{N-1}T_N\begin{bmatrix} s_N \\ u_N \end{bmatrix}) = \ldots = P_0(T_1T_2\ldots T_{N-1}T_N\begin{bmatrix} s_N \\ u_N \end{bmatrix})$$

For stage $N+1$, we can write

$$P_{N+1} = P_N\left(T_{N+1}\begin{bmatrix} s_{N+1} \\ u_{N+1} \end{bmatrix}\right)$$

Combining the equations above yields

$$P_{N+1} = P_0\left(T_1T_2\ldots T_NT_{N+1}\begin{bmatrix} s_{N+1} \\ u_{N+1} \end{bmatrix}\right)$$

which is the proof of the theorem by the induction principle. □

As a simple example, the transformation for a ray traversing a distance d is given by:

$$\begin{bmatrix} y_{out} \\ \theta_{out} \end{bmatrix} = \begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix}\begin{bmatrix} y_{in} \\ \theta_{in} \end{bmatrix}$$

which for a paraxial analysis is equivalent to the two-plane parameterization given by:

$$\begin{bmatrix} s_1 \\ u_1 \end{bmatrix} = \begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix}\begin{bmatrix} s_0 \\ u_0 \end{bmatrix}$$

To find the radiance pattern at the output plane (subscript 1) in terms of the input radiance requires an inverse of the previous equation and yields:

$$\begin{bmatrix} s_0 \\ u_0 \end{bmatrix} = \begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix}^{-1}\begin{bmatrix} s_1 \\ u_1 \end{bmatrix} = \begin{bmatrix} 1 & -d \\ 0 & 1 \end{bmatrix}\begin{bmatrix} s_1 \\ u_1 \end{bmatrix}$$

so that the plenoptic function at the sensor plane is now:

$$P_1\left(\begin{bmatrix} s_1 \\ u_1 \end{bmatrix}\right) = P_0\left(\begin{bmatrix} 1 & -d \\ 0 & 1 \end{bmatrix}\begin{bmatrix} s_1 \\ u_1 \end{bmatrix}\right)$$

Suppose a cosine squared texture pattern is modeled as a Lambertian object (a model used commonly to describe diffuse reflections of *normal* objects) the radiance pattern is defined to be constant from any viewing angle and is given mathematically by:

$$L(\theta) = L_0$$

where θ is measured from the normal of the object is reparameterized in terms of the two plane parameterization as:

$$L(s,t,u,v) \rightarrow L(s,u) = L_0 f(s)$$

which scatters a constant radiance at all angles and is modulated by any function in the spatial domain . In a 2D horizontal slice (s,u) of the 4D hyper plane (s,t,u,v), a 100mW source of this type gives the plenoptic function P_0 shown in Figure 4**.**

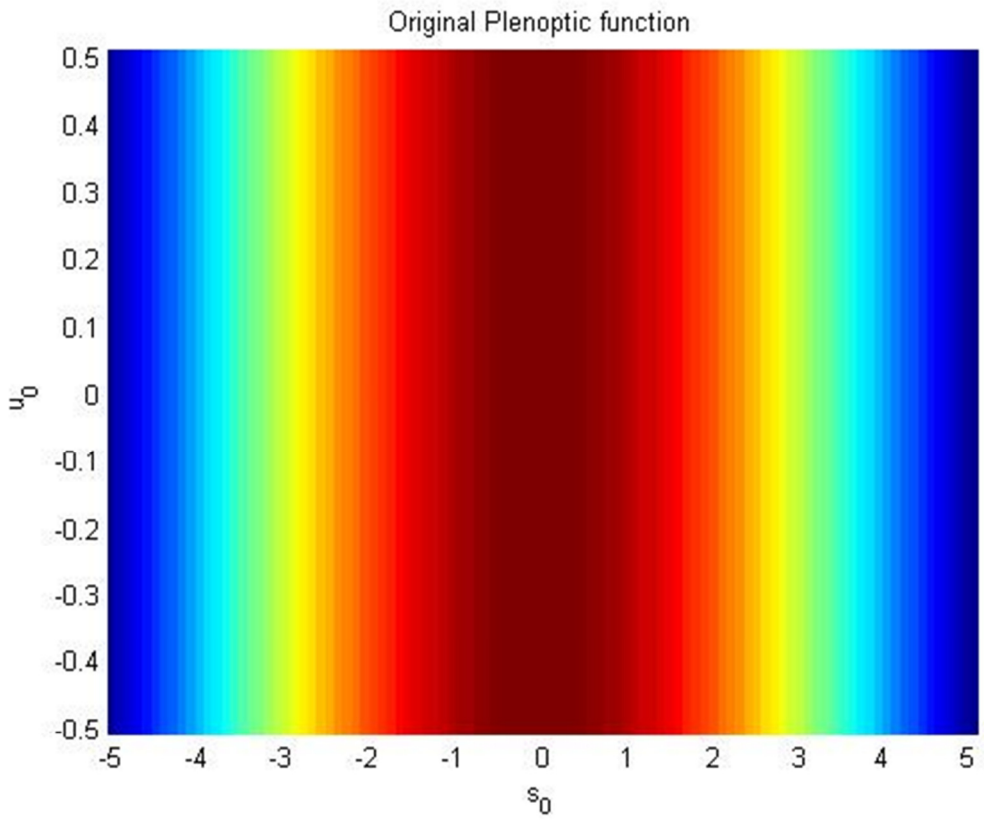


Figure 28: 2D slice of irradiance P_0 of a Lambertian Object.

The cosine squared texture pattern chosen for *f* simply shows that the center of the object is bright and becomes dark as it reaches the edge of the object. The transformed radiance pattern at distance *d* is then given by P_1, which is shown in Figure 29:

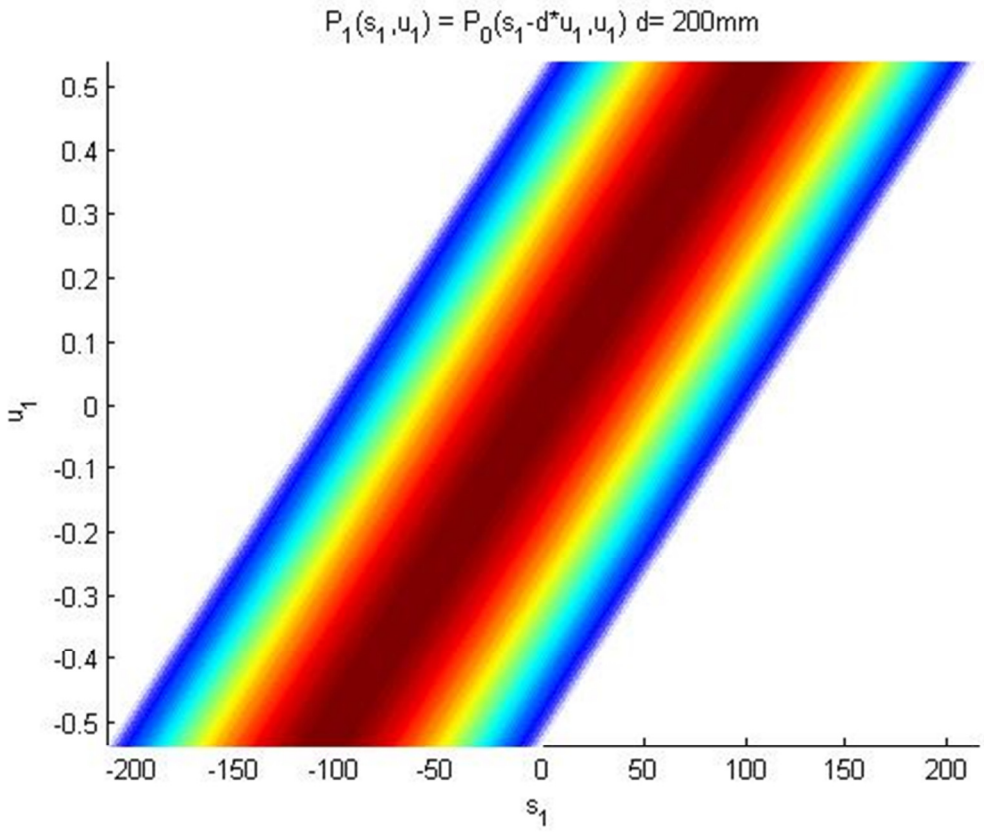


Figure 29: Plenoptic function P_1 has the same overall form as figure 4, but its coordinates are changed to a new basis by the change of basis coordinates derived from the optical matrix transformation

The slope of constant contours is equal to the distance traversed by the light field. As another example, in the two stage system shown in Figure 30, after traversing a distance *d* and hitting a thin lens of focal length *f*,

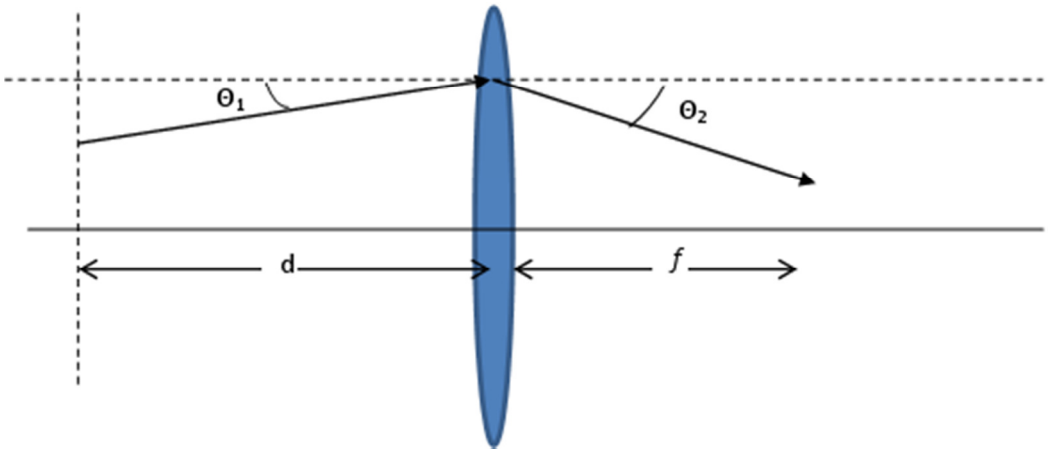


Figure 30: Ray traversing a distance d and striking lens with focal length *f*

the plenoptic function immediately after the thin lens shown in Figure 31 and

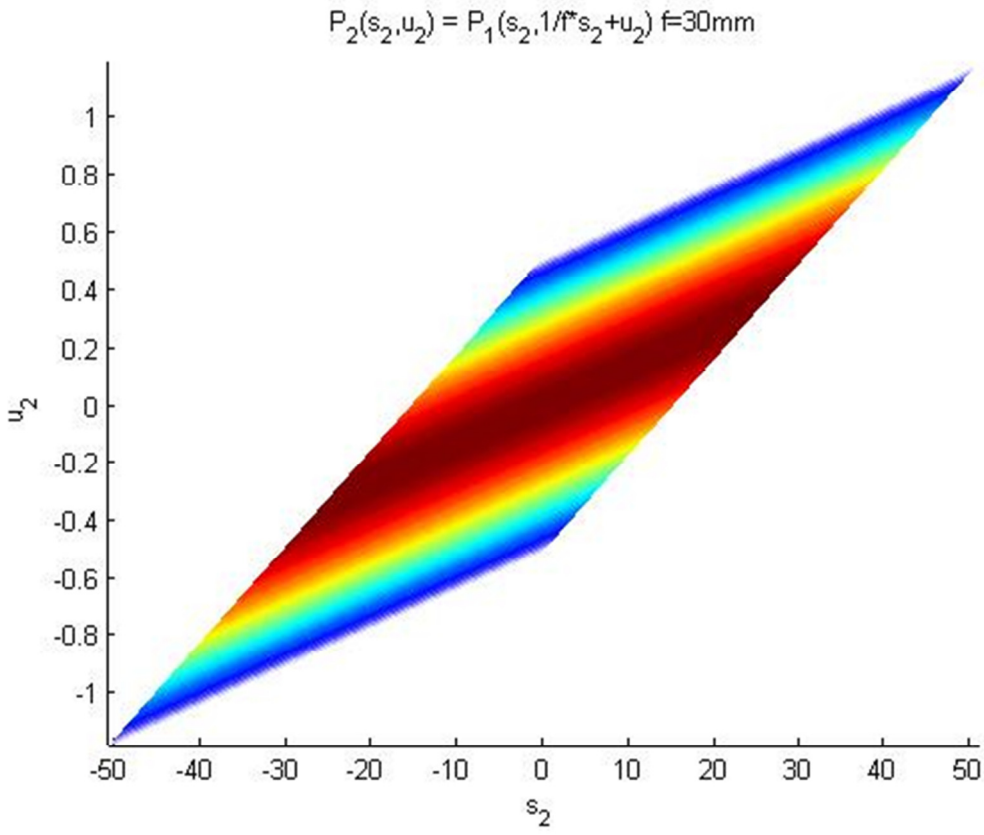


Figure 31: plenoptic function immediately after the thin lens from Figure 6

is given by:

$$P_1\left(\begin{bmatrix} s_1 \\ u_1 \end{bmatrix}\right) = P_0\left(\begin{bmatrix} 1 & -d \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ \frac{1}{f} & 1 \end{bmatrix}\begin{bmatrix} s_1 \\ u_1 \end{bmatrix}\right)$$

Modeling image formation in this manner can now clearly identify the differences between the different models that were outlined in Chapter 2 for imaging systems.

3.2.4 Using Image Formation to Compare Imaging System Models

In the last section, as a consequence of the brightness theorem, a matrix optics formulation was derived for the calculation of the plenoptic function through various imaging systems in the paraxial range. This method can be expanded to provide a clear comparison between the models of the different imaging systems in Chapter 2. Beginning with the pinhole model of the camera, it was shown that it is equivalent to the thin lens model, where the focal length of the thin lens was simply the spacing between the sensor and the pinhole, and an infinitesimal aperture stop is placed in front of the lens only allowing rays of light to travel through the center of the lens. This can be modeled as a modulation of the light field (5) by P_1,

$$P_1\left(\begin{bmatrix} s \\ u \end{bmatrix}\right) = P_0\left(\begin{bmatrix} 1 & -d \\ 0 & 1 \end{bmatrix}\begin{bmatrix} s \\ u \end{bmatrix}\right)\delta(s)$$

where the δ-function represents a pinhole in the horizontal direction at the aperture stop, and d is the distance of the ray traversed from the object. After the thin lens, the light ray travels an extra distance of $d_{sensor,}$, which is the distance traveled after hitting the pinhole to the focal plane. The new plenoptic function is given by:

$$P_2\left(\begin{bmatrix} s \\ u \end{bmatrix}\right) = P_1\left(\begin{bmatrix} 1 & -d_{sensor} \\ 0 & 1 \end{bmatrix}\begin{bmatrix} s \\ u \end{bmatrix}\right) = P_0\left(\begin{bmatrix} 1 & -d \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & -d_{sensor} \\ 0 & 1 \end{bmatrix}\begin{bmatrix} s \\ u \end{bmatrix}\right)\delta(s - d_{sensor}u)$$

This shows that rays only from one point in the reference plane will enter the system and only one image point will be formed from each point in space. The change in the plenoptic function is shown in Figure 32**:**

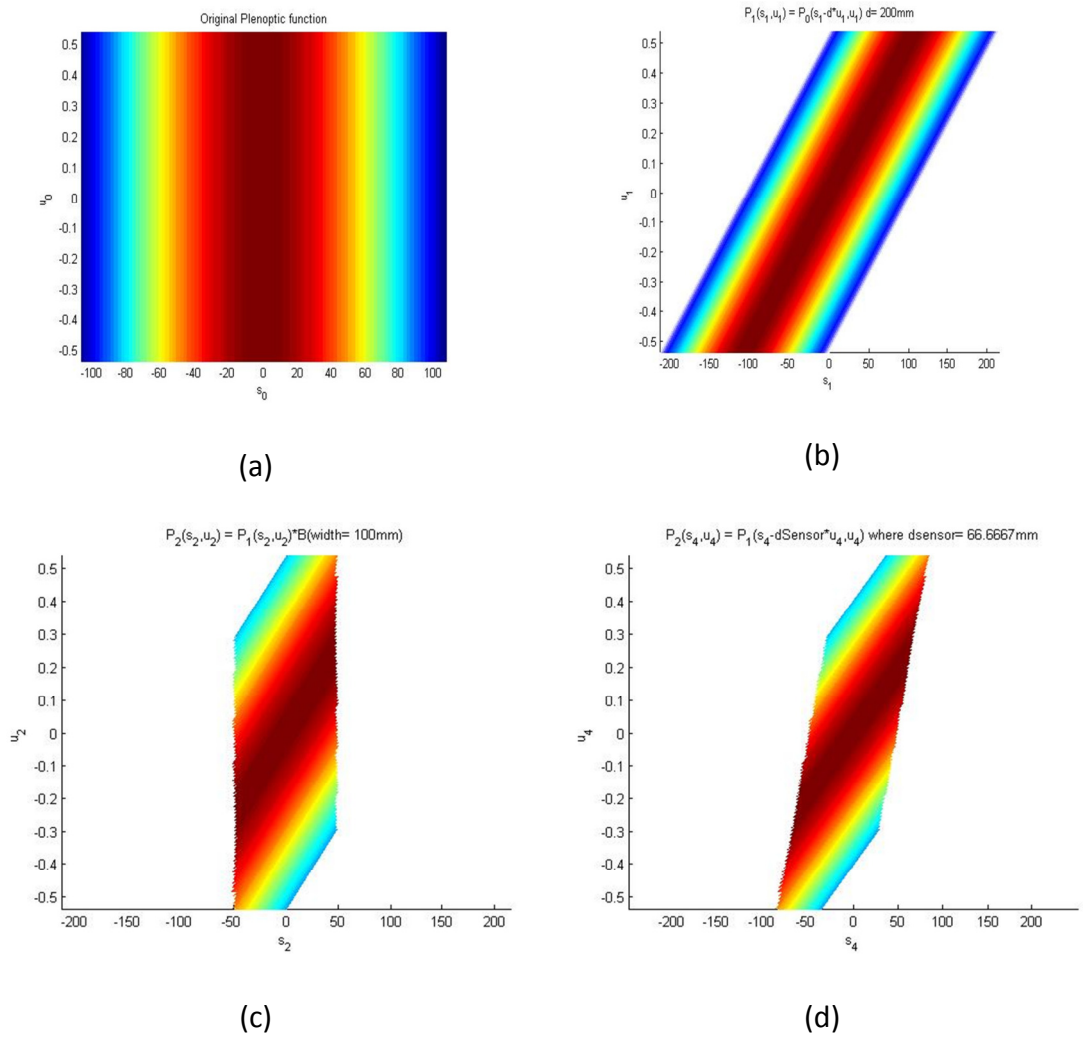


Figure 32: Plenoptic function for (a) original cosine squared texture pattern (b) traveling a distance of d=200mm (c) modulated by a finite aperture width of 100mm and (e) traveling to the sensor plane a distance of f=66.7mm

Notice that the values in Figure 32 for the distance is different from that used in the end of the previous in Figure 29 and Figure 31. The reason for doing this is to illustrate the effect of using the pinhole model as it is compared to the thin lens model. The object in Figure 32 is placed within the focal length of the lens, and from

Chapter 2, this was shown to create a virtual image, namely diverging rays on the image plane but a virtual image on the object side of the lens. The pinhole model does not take that into account as it simply traces the plenoptic function a distance *d=10mm* to the aperture of the pinhole, which was given a width of 5mm to modulate the light field, and lastly traveling the focal plane distance *f=30mm*. Ideally a pinhole should have infinitesimal thickness instead of a modulation width of 5mm, but there are no such practical systems and so a common aperture width for cameras that are modeled as pinholes was used in this simulation.

Removing the modulation aperture from the pinhole model, on the other hand, allows the light rays to hit different parts of the lens, and therefore take into account the refractive properties of the lens. For a perfectly focused image, namely, a point in space that satisfies the thin lens equation:

$$\frac{1}{d} + \frac{1}{d_{sensor}} = \frac{1}{f}$$

The rays undergo three stages as they travel from the object to the image plane: 1) Travel a distance *d* to the lens, 2) refract through a thin lens of focal length *f* and be modulated by its entrance pupil diameter, and 3) travel a distance d_{sensor} to the image plane. From Chapter 2, this was shown to have four scenarios as shown in Figure 14. If the sensor plane is located exactly at the paraxial focus plane, then all of the light fields from the various optical setups shown in Figure 14 would simply reconstruct the light field at the object on the sensor with the appropriate magnification. This is shown for a thin convex lens of unit magnification in Figure 33.

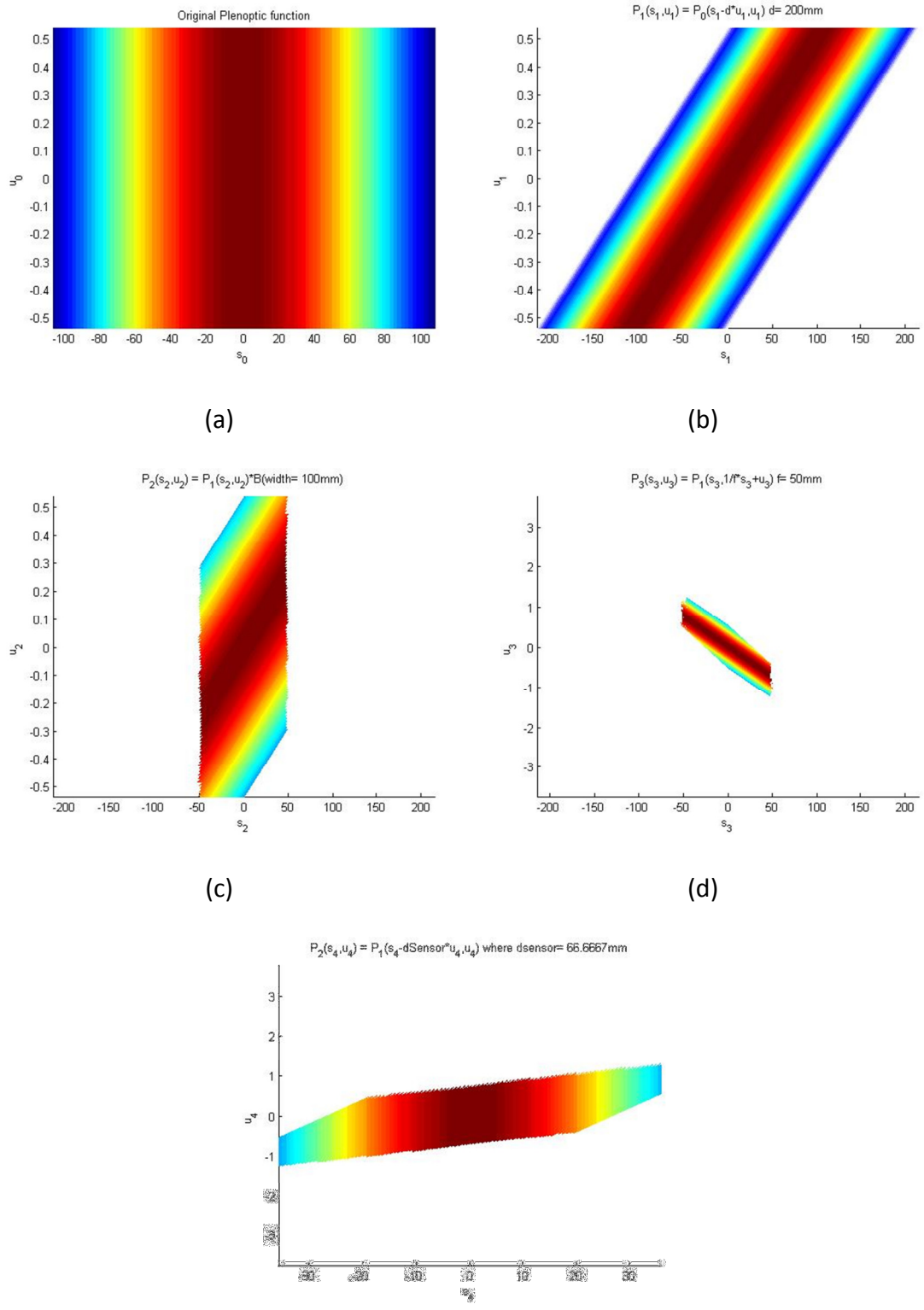


Figure 33: Light fields of (a) cosine squared texture pattern (b) traveling a distance d (c) being modulated by the lens aperture (d) refracting through a CONVEX thin lens and (e) traveling to

the paraxial plane of focus. Note that the *s*-axis limits on the sensor and original planes are the same due to the magnification of *dsensor/d*=1/3.

The thick lens model, the final extension to imaging systems as seen in Chapter 2, extended the thin lens model to account for thickness in the optical prescription of the lens. Recall that the thick lens model had light enter from the entrance pupil, travel a distance through an optical prescription given by a ray transfer matrix that was derived based on a number of optical properties of the imaging system such as principal plane locations, focal lengths, radii of curvature, and indices of refraction of the various lens components present, and finally hit the sensor after leaving the exit pupil of the lens system. It was shown that the input/output relationships of the rays is given by,

$$\begin{bmatrix} y_N \\ \theta_N \end{bmatrix} = M \begin{bmatrix} y_i \\ \theta_i \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} y_i \\ \theta_i \end{bmatrix}$$

where the matrix *M* describes the optical system. In addition, formulas for each of the elements in this matrix were derived based on the various optical parameters. With the availability of software tools like Zemax and CodeV, paraxial ray transfer matrices can be derived for even the most complex of optical systems.

As a simple example, the ray transfer matrix for a single thick lens is given by,

$$M_{sys} = \begin{pmatrix} 1 - dD_1/n_1 & dn_0/n_1 \\ dD_1D_2/n_1n_0 - D_1/n_0 - D_2/n_0 & 1 - dD_2/n_1 \end{pmatrix}$$

where,

$$D_1 = \frac{n_1 - n_0}{R_1} \text{ and } D_2 = \frac{n_0 - n_1}{R_2}$$

The plenoptic function at the entrance pupil is modulated by the entrance pupil diameter and then travels through the optical system with ray transfer matrix given by M_{tot}, the plenoptic function becomes:

$$P_2\left(\begin{bmatrix} s \\ u \end{bmatrix}\right) = P_1\left(M_{tot}^{-1}\begin{bmatrix} s \\ u \end{bmatrix}\right)$$

Lastly, upon leaving the exit pupil, it is again modulated by the exit pupil diameter, which is:

$$P_3\left(\begin{bmatrix} s \\ u \end{bmatrix}\right) = P_2\left(\begin{bmatrix} s \\ u \end{bmatrix}\right)\left(B((s-l)^2 + (t-l)^2 - \left(\frac{ExPD}{2}\right)^2 \leq 0)\right)$$

The following algorithm determines the plenoptic function for all types of optical systems:

Algorithm to determine plenoptic function:

1. Modulation by the entrance pupil.
2. Travel through optical matrix M.
3. Modulation by exit pupil.

As an example, let's implement the algorithm with the telephoto lens from Chapter 2, shown again here in Figure 34.

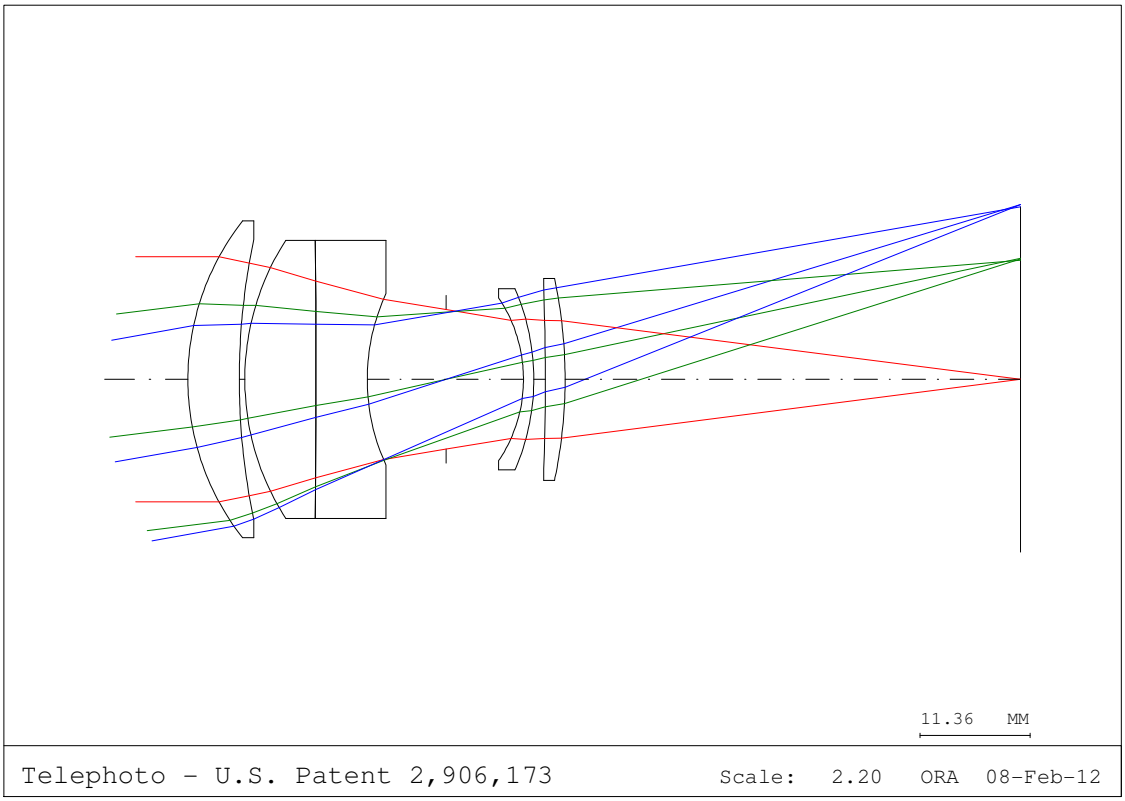


Figure 34: Telephoto zoom lens Patent 2906173 YZ-plane view

Clearly the lens designers put in quite a bit of work to have the complex lens structure of a telephoto zoom lens simply modeled as a pinhole. A light field analysis provides a better appreciation of the optics of such a system and for an image plane placed at the focal plane of the complex lens system, the light field was similar to that of the thin lens.

The question then remains as to when/how an optical setup can be modeled as a pinhole system. Computer vision uses the pinhole model very loosely to provide the transformation from the 3D world to a set of 2D image coordinates, so there must be some inherent error in ignoring the optics. As the distance to the object increases, the light fields show much more shallow slopes than those that were seen in Figure 33.

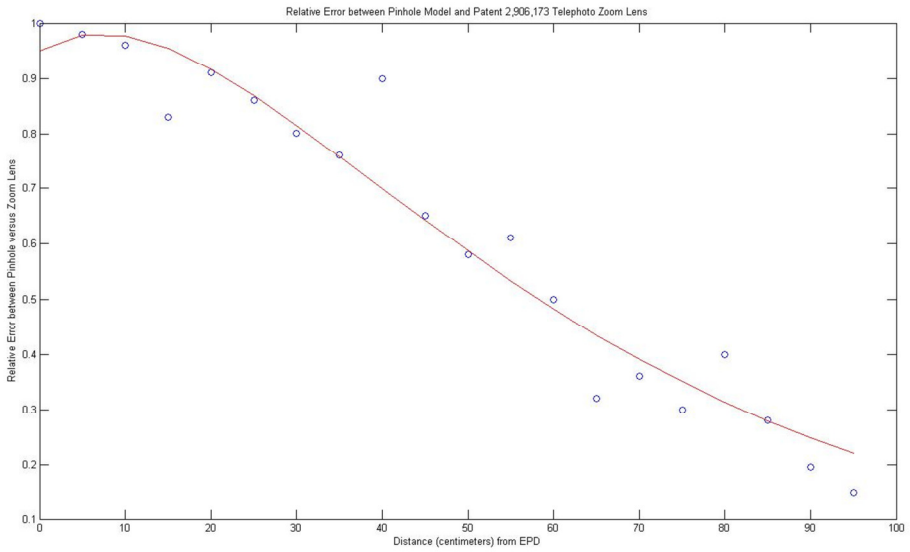


Figure 35: Comparison of light field between complex lens system (Patent 2,906,173) and pinhole system

With the small finite apertures of the optical setup compared to the distance of the objects, Figure 35 shows the errors by assuming the different models. At larger distances, the models are indeed comparable. The small finite aperture compared to both the distance and object size that is being imaged only allows the chief ray to pass through the optical setup. Basically, the angular span of rays that reach the entrance pupil decreases as the object moves to further away from the lens, which is shown in Figure 36.

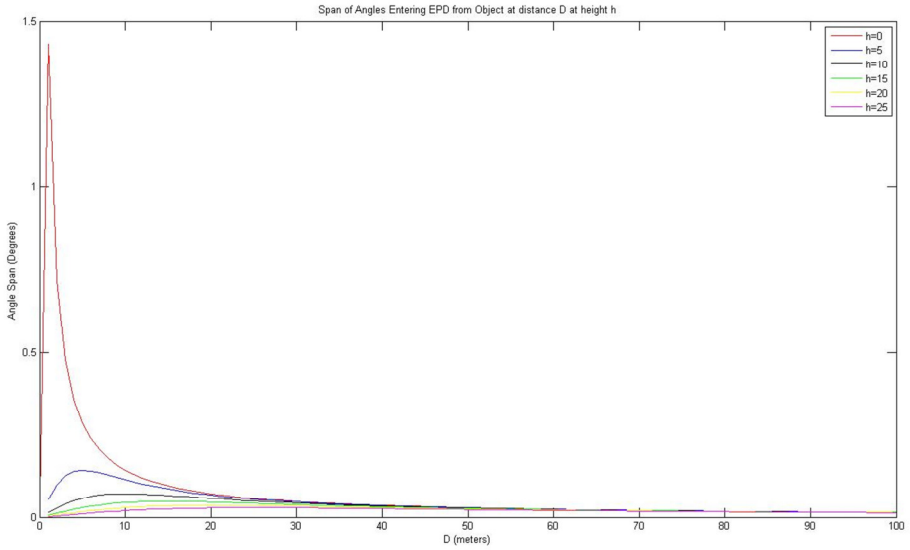


Figure 36: Angular span (in degrees) of rays that reach the entrance pupil diameter for objects at height h (in meters) and distance D (in meters)

By making that assumption, all optical systems preserve the angle between the ray and optical axis before and after the lens model.

3.3 Computational Cameras

Now that different lens systems have been analyzed and methods have been derived to simulate the different light fields for various imaging systems, we must now see how to practically capture the light field. The field of computational cameras is the next step in the evolution from conventional cameras (29).

3.3.1 Camera and Cluster Arrays

Stanford's Computer Graphics Lab built a 16x8 array of tiny little webcams that are networked in a way where they combine images taken by each camera at the same time to form a new image. Their hardware setup is shown in Figure 37.

Figure 37: Stanford Camera Array

The optics, physical spacing, and arrangement are all reconfigurable. This setup theoretically simulates pinholes placed at the (s,t) plane, while their image pixels are at the (u,v) plane. By adding up different pixels from each image to recreate a new image, the focus position and viewing position/angle can be digitally changed from capturing "one" synthetic image, as well as, imaging through occlusions by changing the focus. As it was seen earlier, this setup can be modeled by the setup in Figure 27. The spatial resolution of the array dictates the application that it is used for. If the spacing between the cameras is low, then the cameras act as if they have a single center of projection as shown in Figure 27. If the spacing is high, then the multi-baseline approach can be used to capture light fields and extract more accurate depth of fields. Lastly, if the spacing is not too high or too low, then they function as a

single camera with a large synthetic aperture and can vision through occlusions/digitally refocus images.

To understand the low, middle, and high spatial resolutions of the arrangement of the camera array calls for a study in plenoptic sampling. The most comprehensive framework provided for this area is by Chai, et al in "Plenoptic Sampling" (30). They showed that spectral support of the plenoptic function only depends on the minimum and maximum depths of the object that is being imaged. The minimum sampling curve that is analytically derived for the plenoptic function in this paper shows the minimum number of images and resolution needed for anti-aliased light field rendering. Given the minimum and maximum depths of the maximum camera spacing that is allowed for a given rendering quality was shown to be:

$$\frac{D}{d/f} = \frac{1}{2}\left(\frac{1}{z_{\min}} - \frac{1}{z_{\max}}\right)$$

where D is the rendering quality, namely the distance between two adjacent cameras, d is the size of the smallest resolvable feature on the image plane, f is the focal length of the thin lenses in the micro lens plane, and z are the minimum and maximum distances of the object from the thin lens.

3.3.2 Plenoptic Camera

The idea of placing multiple cameras in an M x N array configuration was to create two points, namely the pixel and the pinhole of the camera, to define a ray for each pixel. Thus, the intensity measured by a single pixel can only be defined by one ray, and unlike a conventional camera that averages over all rays that hit the pixel, the new image preserves the information that is on each ray that leaves the object.

Another way of dedicating each pixel to a specific ray is to refocus the rays that hit the plane of focus to its own pixel, as shown in Figure 38**.**

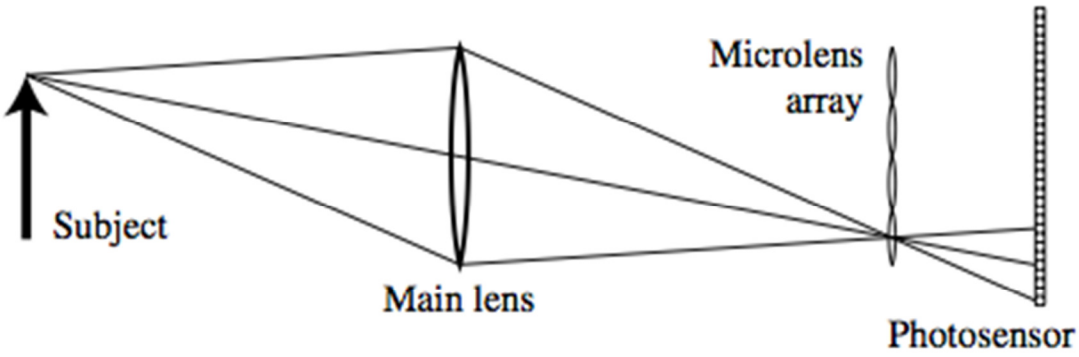


Figure 38: Plenoptic camera model (4)

A conventional camera uses a lens system to focus light onto a sensor plane which averages the radiance from each ray to determine the intensity of the pixel. In other words, by averaging all intensities from all rays that hit that pixel, information that is carried by each ray is lost. The plenoptic camera preserves the information from each ray by placing a lenticular lens array in front of the sensor plane to refocus the light onto different areas of the sensor plane. The intensities now recorded on the sensor plane are the intensities of each ray entering the field of view of the camera.

As the lens array becomes more dense, they capture a number of different views of the object thus approximating the plenoptic function. By choosing a different pixel from each image created, different views of the object can be generated and brought into focus. That is, a single shot of the plenoptic camera produces that which a conventional camera produces with a number of different shots at different focuses. In other words, the lens elements placed at the focal distance

of the main lens forms tiny little images on the sensor plane. The sharpest micro lens image is thus created when it is focused on the principal plane of the main lens.

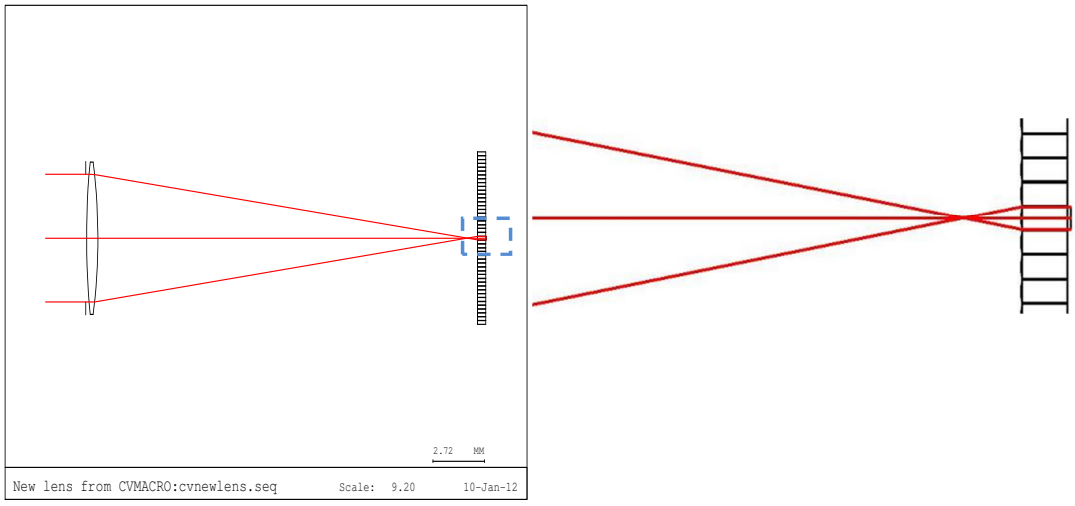


Figure 39: (Left) Planar wave front striking main objective *f/4* lens and being to a micro lens array of the same *f/#*. The focal length of the thin lens is 20mm with a 5mm EPD. (Right) Zoomed in view at the focal point and micro lens.

The camera is designed with cameras that are easily detachable from their sensor arrays. Once detached, a lens array is placed in front of the sensor array whose *f/#* matches that of the main lens to avoid overlap and missing pixels, as shown in Figure 39 (28). These little lenses each create an image of the main lens and thus instead of one image being created, there are MxN images created where, MxN is the size of the lens array.

Using a lens array gives rise to one important tradeoff with the camera/image resolution. Suppose the lens array is of size 10x10 and the camera is a high resolution camera at 1920x1080. This means that the maximum resolution of each image that is created by the lens is 192x108. Different points in these low resolution images are then combined to create what a conventional camera

could only do with multiple snapshots such as, digital refocusing and multiple viewpoints, but in the end, resolutions are those achieved from small, cheap cameras.

As an application to generating a number of images on a single sensor array, one can begin to think of single lens stereo. Adelson and Wang in their paper use the plenoptic camera and the least squares gradient technique from (31) (32) to extract information about both the horizontal and vertical parallax. This avoids the single aperture problem, as well as the correspondence problem that is present in stereovision. Furthermore, with measurements of parallax in both directions, the depth can be computed more accurately. It is like having multiple views of the object with a single camera. The only limitation is the maximum disparity is defined by the entrance pupil of the imaging system. In their paper, Adelson and Wang generated low confidence depth maps of close objects, but unfortunately no ranges/metrics were provided.

3.4 Closure

Light fields and their applications are only currently used for computer graphics models of texture maps and digital refocusing to name a few applications. There is much that can be done if a function can be accurately estimated whose samples are images of a scene from any viewing direction, color, and time. Lytro Technologies Inc. is a new startup company that is bringing the plenoptic camera out of the research laboratory and onto consumer shelves. More importantly, by accurately modeling the lens, the prototype for long range light field measurements can become a reality.

The change in the different light field patterns with respect to the different optical prescriptions shows that simplistic lens models cannot be used to approximate the detail that must be taken into account for imaging a scene in general. High resolution images can be obtained for scenes by advancing such technologies and using better models. Depending on the application, simplifications can be made. This chapter focused on the comparison of these models in a paraxial analysis, which is again a good assumption for objects at far distances compared to the size of the lens. A more precise comparison could be made by using the ray tracing function in CodeV to find exactly where these rays hit the image plane.

Applications of such technologies could then be used in imaging systems to reconstruct the wave fronts that are entering the camera system. Soon Adelson and Wang's vision periodic table may even become a reality, and much of the post processing image feature extraction and classification can be simply identified if certain patterns are found in the light field (33). The remaining chapters will investigate two such applications in astronomy and surveillance systems.

4 Plenoptic cameras and Adaptive Optics

4.1 Introduction

In Chapter 3, the plenoptic camera was described as a system that is able to retrieve the ray information that is lost by a conventional camera. As stated, conventional cameras integrate over all possible rays that strike a particular pixel on the surface and provide the intensity value in usually an 8-bit format. The plenoptic camera, however, has each pixel dedicated to a particular ray that arrives within the system. Then by applying the change of bases described in Chapter 3 for the respective optical systems that are being used, different rays are chosen to create an image. These rays are from light that has come from objects that are in focus at different depths and thus the plenoptic camera is a system that has multiple depths of focus within a single capture of its light field.

In this chapter a closer look will be taken at the technology used by a plenoptic camera to retrieve the directional information of the light ray that strikes a particular pixel. As described at the end of Chapter 3, micro lens arrays are placed before the sensor plane to sample the light rays that strike the sensor plane. These lenslet arrays can be used to provide information of rays that are traveling in a certain direction towards the camera.

Light is a wave, but locally, each point on the wave front can be given a direction by its k-vector. The light ray is defined as a ray that points in the direction of the local k-vector of the wave. Ray transfer matrices and ray traces all translate the Helmholtz wave equation to a geometric problem by analyzing the light rays to determine the properties of the system. CodeV, as will be seen later, primarily focuses its analysis

with the ray definition of light by tracing rays through the optical system and finding where they land on the image sensor plane. Its diffraction analysis, on the other hand, defines an FFT grid to use Fourier transforms to solve the wave equation so that if the user chooses to define differential phases on the wave front the wave properties of light are available to do so. All spot diagrams, which match the FFT Image formation outputs, are found with ray traces, while all image sensor simulations are found by solving the wave equation. In light field analysis, both will be used to show the consistency between the ray and wave nature of light.

Regardless of the definition of light, devices that capture these light fields, like conventional cameras, simply add together the intensities of rays of light traversing from the object to give an irradiance pattern on the image plane given by the equation:

$$I(u,v) = \int_{0}^{\frac{N_y}{2}} \int_{0}^{\frac{N_x}{2}} \rho L(s,t,u,v) ds dt$$

where ρ is the responsitivity of the detectors, $L(x,y)$ is the irradiance of the ray, and the pixel resolution is N_x x N_y. With lenslet arrays in front of the image plane however, we can identify each pixel as the intensity of a single ray, as will be seen.

4.1.1 Micro lens Arrays

Micro lens arrays are, as the name suggests, arrays of micro lens elements that have either been grown on a substrate or cut into thin glass sheets that are epoxied together (34) (35). Use of such devices evolved from the original design of the Hartmann Screen Test which samples apertures of an optical system. A screen with a series of holes in an array was placed in front of an 80cm refracting telescope known as the

"Great Refractor" (36). It was first used to identify the problem in the optics of this telescope by sampling the aperture to only allow certain rays to pass through and reach the pixels behind the hole.

As mentioned, conventional cameras integrate over all ray bundle intensities striking the pixels and thus the ray information is lost. Ideally, far away, a wave front should have rays whose directions are parallel to the optical axis and thus should effectively be sampled by the Hartmann screen to only allow light that travels through the specific hole to strike the screen. If, on the other hand, a wave is distorted, it is no longer a plane wave and will have different k-vectors (rays) traveling at different angles along the wave front, as shown in Figure 40 and Figure 41. The goal of the Hartman screen test is to measure the displacements these points undertake to be able to reconstruct the wave front that was initially incident on the Hartman screen. If that is known, things like deformable mirrors and other optical elements can be used to refocus the distorted ray to its appropriate position.

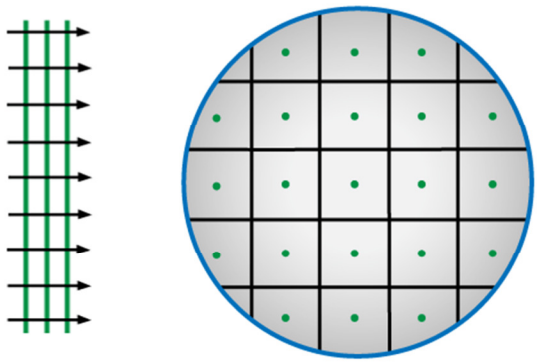


Figure 40: Planar wave front and its image as it passes through a Hartmann Screen (37)

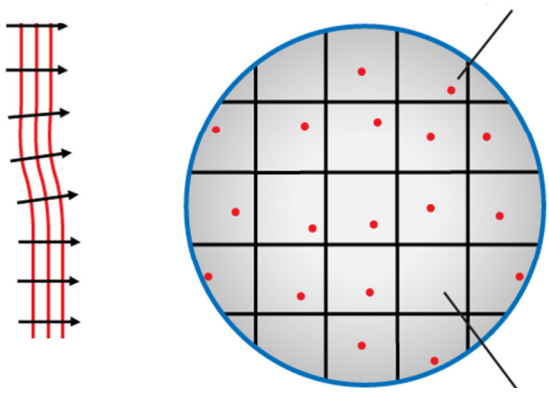


Figure 41: Distorted wave front as it travels through a Hartmann Screen (37)

The issue that arose with the Hartmann screen test was that the image was not bright enough to measure these displacements and thereby reconstruct the wave front. To increase the optical power of the image, micro lenses whose apertures were about the size of the holes with small focal lengths (i.e. large optical powers) were placed in front of an imaging screen inside the holes of the Hartmann screen. These lenses thus focused the small bundle of rays that would pass through the hole to an even finer point, adding their intensities, thus brightening the image as shown in Figure 42. The apertures of the micro lenses are small enough to be considered as pinholes and only allow a small solid angle coverage to pass through them.

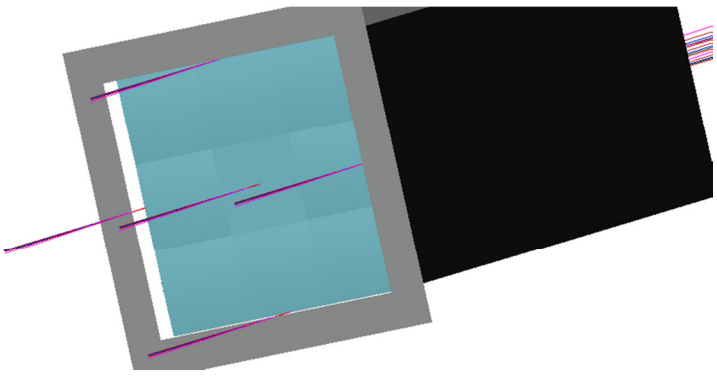


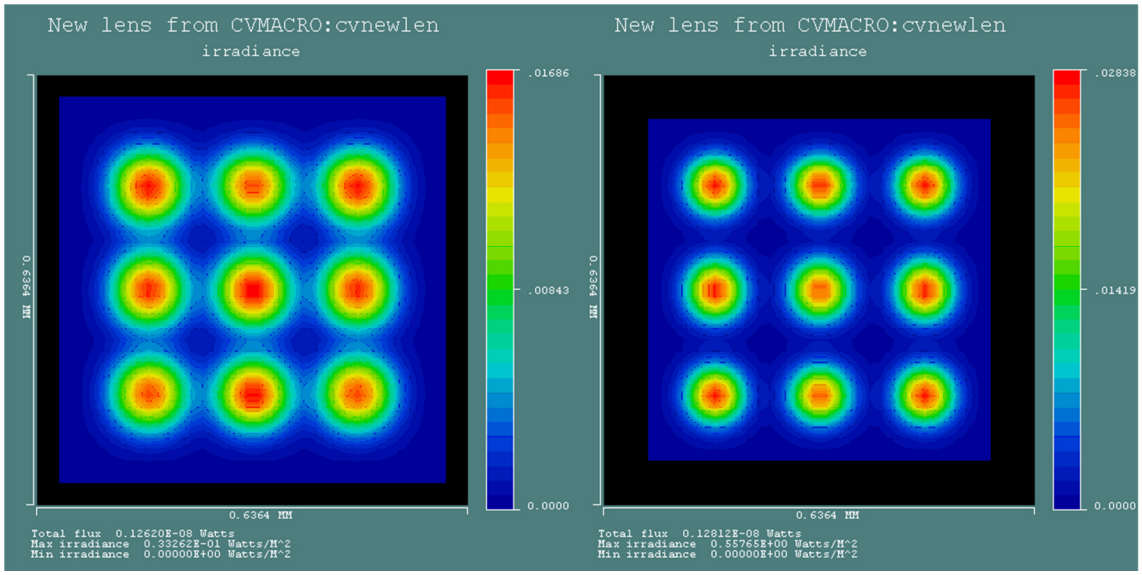
Figure 42: 3D Rendering of 3x3 Rectangular Lenslet Array from CodeV

As a side note, intensities defined along a ray are somewhat of an ill-posed statement. Intensities are defined over areas and the area of a ray is zero. Thus the pixels of a CCD placed behind a Hartman screen are actually measuring bundles of rays whose angles of trajectories are close-enough to be approximated as the same ray.

To increase the amount of light striking each pixel the micro lens elements focus the bundle of rays by adding their intensities. Similarly modeled is the coupling of light to an aperture of the optical fiber shown in Figure 25. Clearly from the brightness theorem the brightness of the pixel cannot exceed the brightness of the source. If the radiant power of the source is *P* watts then the power collected by the aperture is:

$$P' = 2\pi B_1 S_{source}(1 - \cos(\theta_1))$$

where B_1 is the brightness of the source with area S_{source}, and maximum aperture angle of θ_1. For a finite aperture, as seen in Chapter 2, the power collected by the finite aperture would be spread over the image plane located a distance *f* from the finite aperture. By placing a thin lens of small focal length *f* in the aperture, the power of the light is focused to a smaller area on the image plane and therefore produces a much brighter image which is shown in Figure 43**.**



(a) (b)

Figure 43: Source of 100mW with a Lambertian radiance profile placed far away from (a) Hartman Screen (b) Shack Hartmann Sensor

The Hartmann screen gives a maximum irradiance of .0332 Watts/m^2, while inserting thin lenses of the appropriate focal length in the holes of a Hartmann screen increases it by over a factor of 10. From Figure 43 it is very clear that the lenses focus the light that is sampled by the apertures to a much smaller area to achieve the higher brightness. The pinhole of the Hartman screen and the pixel coordinate form a unique line in 3D space which defines the trajectory of the ray, and thus the intensity measured by the pixel of a CCD is the intensity that was traveling in a certain direction of the wave front.

The Shack Hartmann Sensor and Plenoptic (Light field) Camera, are technologies of two different generations that used in completely different applications. The Shack-Hartmann sensor was developed over 30 years ago and the plenoptic camera is only now beginning to enter the market. From two different times, these devices are both used to divide a wave front up into its individual ray components for analysis, only

the plenoptic camera focuses the incoming wave front before it reaches the micro lens array.

4.2 Shack Hartmann Sensor

The Shack Hartmann sensor is used to measure the degradation in a wave front as it travels through space. These degradations are results of either instabilities in the channel, such as turbulence (38) , that perturbs the wave as it makes its way to the optical system, or the optical system itself whose imperfections in manufacturing the lenses or mirrors used cause aberrations within the image (39). The Shack Hartmann sensor was first used in space telescopes where a test was conducted by poking many tiny holes in a surface and placing that mask over a telescope to force the wave front to pass through many different entrance pupils. By placing lenses within each of these holes to create a lenslet array on the Hartmann Screen, the light passing through the apertures is concentrated to a focal spot, which would allow the Hartmann screen to increase the photon density and thus work in places of limited light (36).

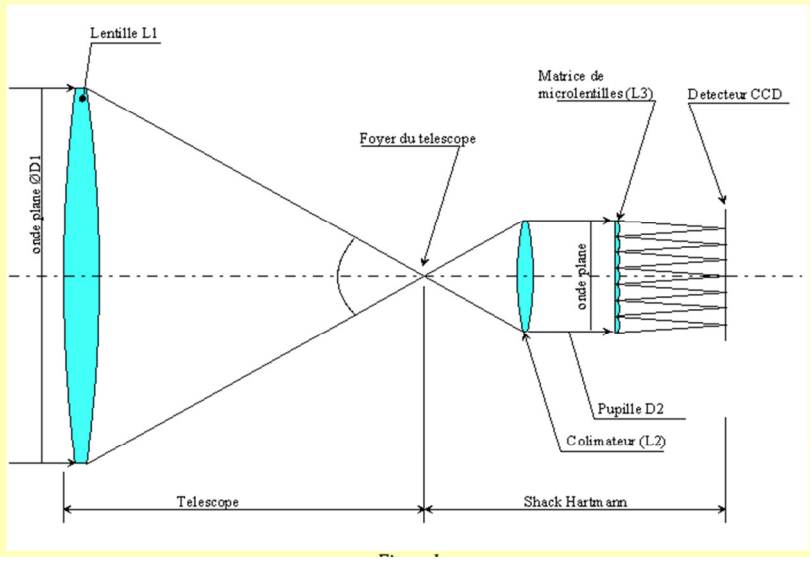


Figure 44: Principle of Shack Hartmann Sensor (40)

An undistorted, collimated laser beam is sent in and used as the reference to measure the displacements of the focused light from the distorted wave as it comes into a focus at a different point depending on the channel and optical imperfections it passes. These displacements are used to measure the imperfections to correct the image for better quality, as shown in Figure 40 and Figure 41.

4.2.1 Principles of Operation

On its website, Thor Labs, provides a nice description of the derivation of the Shack-Hartmann sensor, shown below:

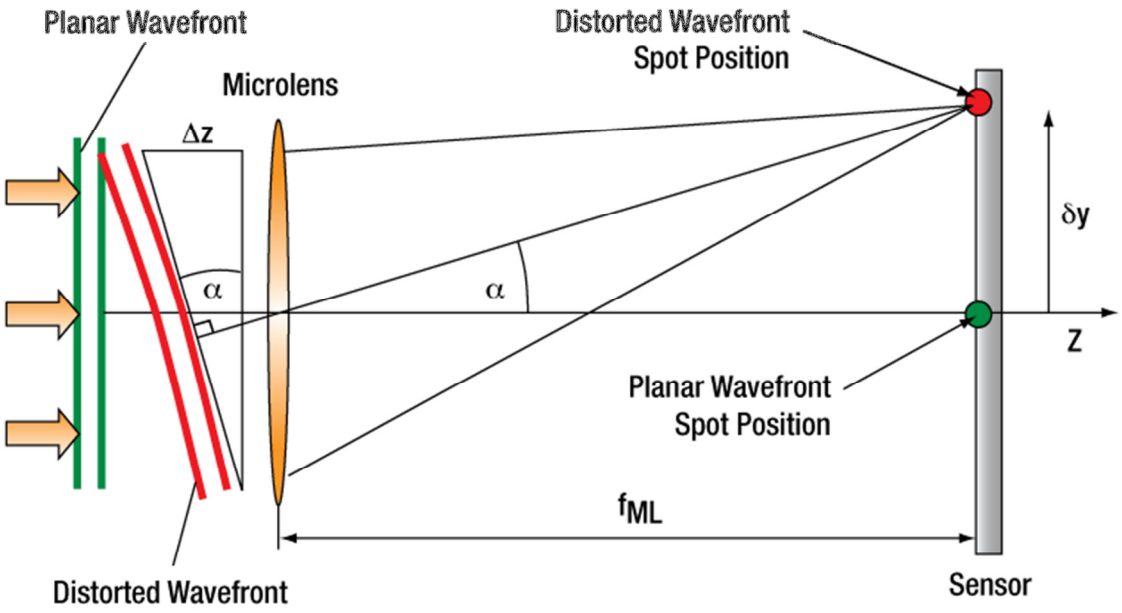


Figure 45: One Lenslet from the array of a Shack-Hartmann Sensor (37)

From paraxial optics, we know that rays coming in parallel to the lens will be focused at the focal point, which in this case is the intersection of the sensor plane and the optic axis. Any wave front travelling at an angle α with respect to the normal of the

lens will be moved δy on the sensor plane. From similar triangles in a 2D domain, we can relate the angle of trajectory to the point displacement by,

$$\tan(\alpha) = \frac{\Delta z}{D} = \frac{\delta y}{f_{ML}}$$

Using this equation gives the angle of trajectory as:

$$\alpha = \tan^{-1}\left(\frac{\delta y}{f_{ML}}\right)$$

The maximum angle that can be detected then depends on the EPD of the lenslet, and is given by:

$$\alpha_{\max} = \tan^{-1}\left(\frac{D/2}{f_{ML}}\right)$$

Angles larger than the maximum will still form patterns on the sensor plane, however, there will be a cross talk between areas in the lens system that correspond to each micro lens element and thus the light from lens 1 for example will fall in the pixel region of lens 2. The minimum measureable angle depends on the pixel pitch of the sensor plane. These deviations are measured through Zernike polynomials whose coefficients represent the order and amount of a specific aberration (tilt, defocus, coma, astigmatism, etc…).

4.2.2 CodeV Simulation Results

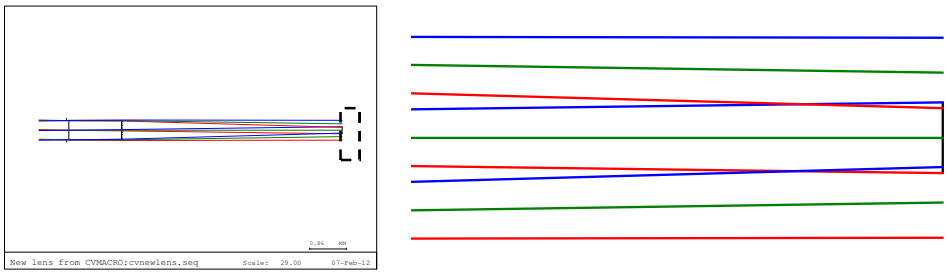
Using CodeV, a 3x3 region of the Shack-Hartman sensor WFS150-5C from Thor Labs was simulated via the Paraxial Ray tracing function. CodeV inputs the angles of rays that are coming from an object placed at infinity. These angles are measured

between the normal of the sensor array, which in this case at the front-end is planar, and the local k-vector of the wave striking the surface.

For this sensor, the maximum angle deviation that can be detected is:

$$\alpha_{\max} = \tan^{-1}\left(\frac{D/2}{f_{ML}}\right) = .8321^{\circ}$$

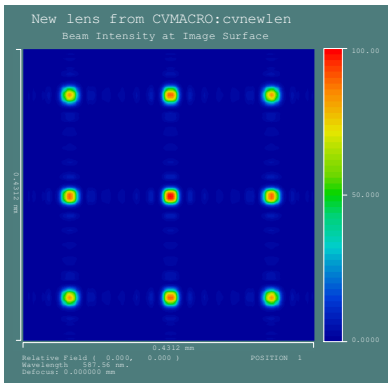
As it can be seen in Figure 46, any angle larger than the maximum differentiable angle, even to the extent of a slight offset to an angle of 0.9 degrees, the rays from separate regions begin to overlap:



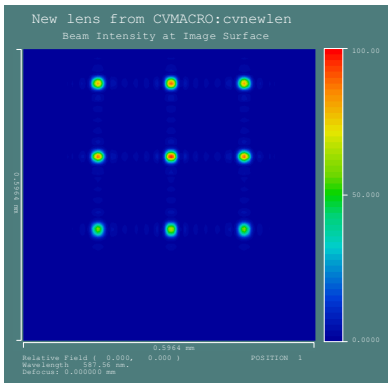
(a) (b)

Figure 46: (a) Shack Hartmann Sensor WFS150-5C from Thor Labs with light rays incident at (red) -0.9, (green) 0, and (blue) 0.9 degrees. (b) The blue and red rays from separate lenses overlap, thus having cross talk in the image plane.

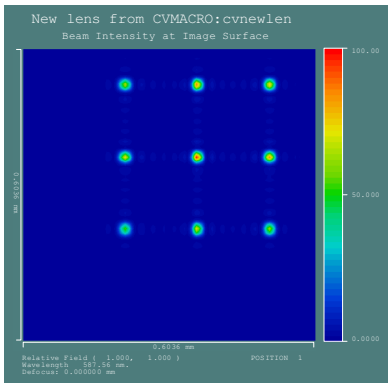
Sampling the range input angles at .1 degrees produced the images in Figure 47. These angles are then compared to the displacements of the centroid of the beam on the image beam.



(a)



(b)



(c)

Figure 47: Image sensor with input as point source hitting lens array at (a) (0,0) deg (b) (0,0.9) deg and (c) (0.6,0.9) deg. Note the shift in (x,y) coordinates of the centroid.

Computing the relative error between the CodeV simulation and its input yields an exponential drop off as the angle reaches the maximum angle. All angles measured in Figure 48 are in degrees:

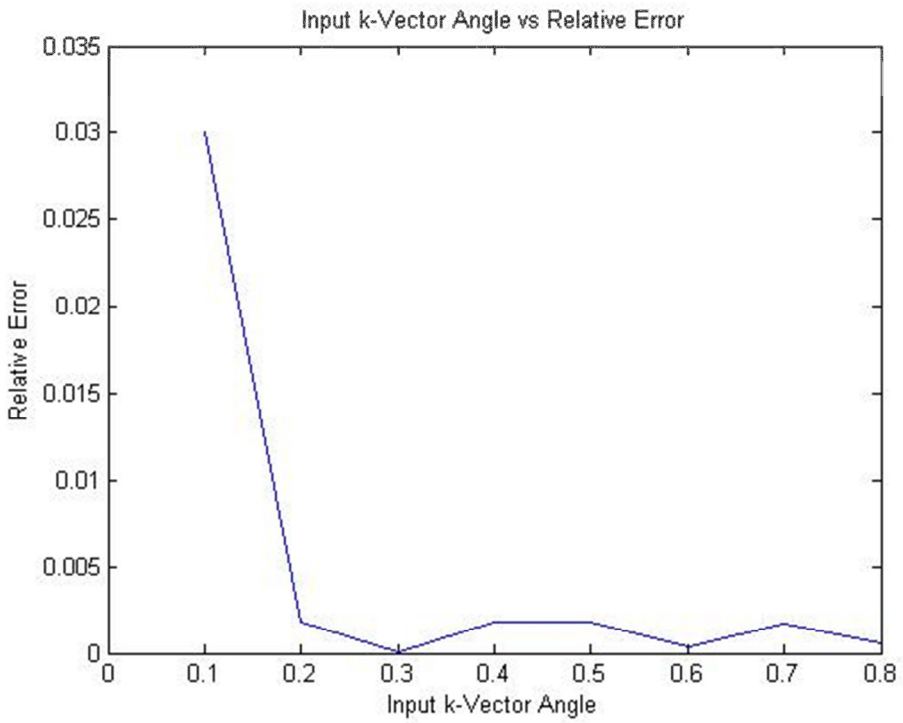


Figure 48: Error in Measured Input Angle versus Input Angle for Shack-Hartmann Sensor

Using adaptive optics with the Zernike coefficients, the aberrations can be corrected to achieve the best focused image.

4.2.3 Zernike Polynomials and Wave Front Corrections

Although not a focus of this thesis, for the sake of completeness, this section describes the mathematical formulation for describing aberrations and their applications in adaptive optics from the inputs of a Shack Hartmann sensor.

The quantitative measure of spread of the ray intercepts on the image plane is known as the aberration (41). In optics there are two types of monochromatic aberrations – Siedel aberrations and Wave front aberrations. Siedel aberrations are those that are produced by the optical system used to image the light coming into the system and

consist of 1) Spherical Aberration 2) Coma 3) Astigmatism 4) Field of curvature 5) Distortion.

The spread of the intercept of rays travelling parallel to the optical axis on the focal plane is known as spherical aberration. It is one of the most important aberrations in a lens (42). This aberration is particularly important when objects are far away from the camera and it is assumed that rays coming from the objects are nearly parallel to the optical axis. Spherical aberrations are analyzed through spot diagrams and can cause focal length shifts of several millimeters for a common singlet lens. Coma is the aberration in which the image of an off-axis point varies for rays passing through different regions of the entrance pupil. In many optical systems, planar objects are projected as curved surfaces on the image plane (43). These effects are due to curvature of field and astigmatism. Distortion can be classified into two categories; 1) Pin cushion distortion: shrinks the image towards the optic axis and 2) Barrel distortion: stretches the image away from the optic axis. Astigmatism arises from the inability of the optics to focus a point object into a focused image on the sensor plane. Wave front aberrations are those that are produced by the channel to distort a planar wave front into one with different k-vectors at a point of common phase.

Zernike polynomials are a group of polynomials that used to describe these aberrations in mathematical form. The aberrations that reach the image plane are given in terms of the Zernike polynomials by:

$$W(\rho,\theta) = \sum_{m,n} C_n^m Z_n^m(\rho,\theta)$$

Where $C_n^m$is the Zernike coefficient for Zernike polynomial Z with radial mode *m* and angular mode *n.* Tables and catalogs of these polynomials are used to describe the

aforementioned distortions (44). For instance, any polynomial with mode with m=0 describes some sort of defocus or spherical aberrations present within the wave since it does not depend on the angle θ. This information is then fed to a deformable member via some control mechanism to force the deformable mirror to assume a shape that is conjugate to the aberration profile.

4.3 Plenoptic Camera

The plenoptic camera is very similar to the Shack Hartmann sensor in its use of micro lens arrays to retrieve ray information with subtle differences.

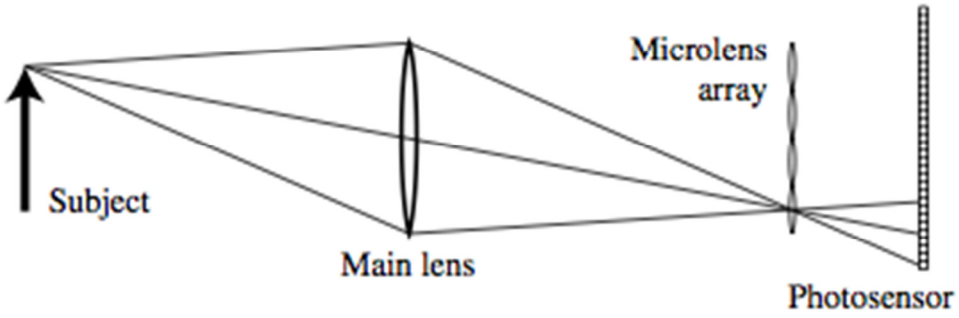


Figure 49: Schematic of plenoptic camera

As it can be seen from the figure above, the lenslet array is placed at the focal point of the main lens and thus light incident onto the array is ideally in focus rather than collimated through a beam expander or eyepiece. In addition, the lenslet array is chosen so that its f/# matches that of the main lens whose focus is at the lenslet array. In the Shack Hartmann sensor the lenslet array is mounted on a C-mount of the lens so that it will take a collimated source rather than a focused one. In addition, most Shack Hartmann sensors are designed for high bandwidth applications (UV-Infrared) while the plenoptic camera is focused in the visible domain. But with a micro lens

array, could the plenoptic camera, which is much cheaper, provide any information about aberrations?

4.3.1 Principles of Operation

Liang and Shih (5) used the principals of matrix optics to evaluate the transfer of the light field through imaging systems as a whole rather than ray-by-ray. The light field analysis was shown in Chapter 3 where the paraxial/ray optics assumptions were compared to the ray traces from CodeV. To analyze the intensities of a plenoptic camera amounts to tracing the intensities along a bundle of rays that strike a micro lens that focus a thin bundle of rays to form a brighter image. The primary

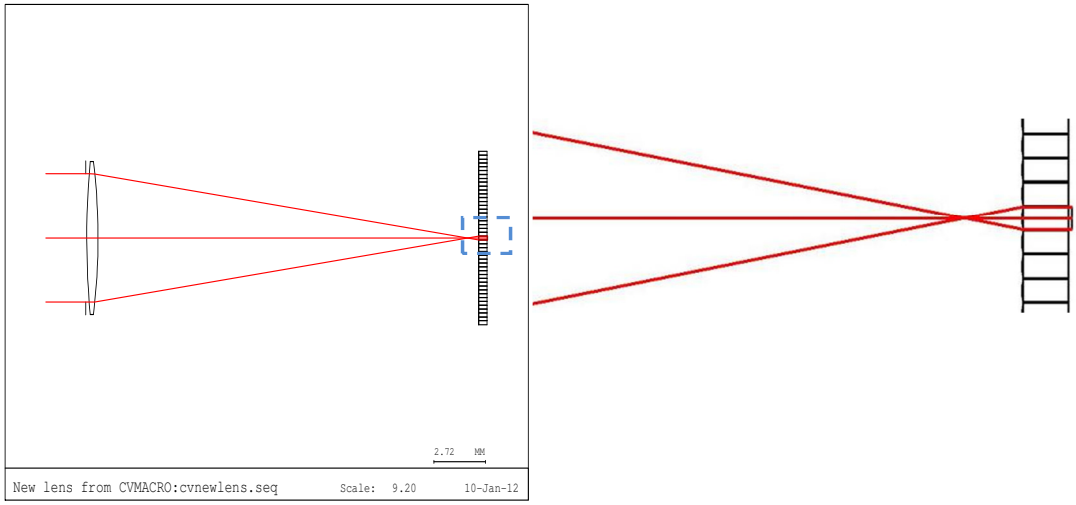


Figure 50: (Left) Planar wave front striking main objective *f/4* lens and being to a micro lens array of the same *f/#*. The focal length of the thin lens is 20mm with a 5mm EPD. (Right) Zoomed in view at the focal point and micro lens.

rule in the choice of micro lens arrays is that the *f/#* of the objective and micro lens array must match. That is, the angle must be preserved so that there will be no samples missed or overlap, as is seen in Figure 50. The ray input-output relationship can be given by the paraxial matrix approximation as:

$$\begin{bmatrix} r_{out} \\ \theta_{out} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -\frac{1}{f_{ML}} & 1 \end{bmatrix} \begin{bmatrix} 1 & f_{ML} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & f_L \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -\frac{1}{f_L} & 1 \end{bmatrix} \begin{bmatrix} \frac{EPD_L}{2} \\ 0 \end{bmatrix}$$

Where f_{ML} (f_L) is the focal length of the micro lens (objective lens) and EPD_L is the entrance pupil diameter of the objective lens. Using the *f/#* equality that:

$$\frac{f_{ML}}{\left(\frac{EPD_{ML}}{2}\right)} = \frac{f_L}{\left(\frac{EPD_L}{2}\right)}$$

the output characteristics of the ray can be found as:

$$\begin{bmatrix} r_{out} \\ \theta_{out} \end{bmatrix} = \begin{bmatrix} -\frac{EPD_{ML}}{2} \\ 0 \end{bmatrix}$$

which is exactly what is seen in the figure. It can be concluded from this that any ray that undergoes aberrations will result in an image that will not be concentrated in the center of the image sensor because it will have components that do not pass through the focal point of the system, and thus pass through other micro lenses within the array as well.

The primary use, thus far, of a plenoptic camera is in digital refocusing. Without loss of generality, using the thin lens model from Chapter 2, it can be seen that objects at different depths are brought into focus simply by moving the image sensor plane. Mathematically, rearranging the thin lens equation at a particular focal length shows that objects at different depths u are focused at different v.

$$u = \frac{fv}{v - f}$$

If each pixel represents a ray, then sampling different pixels will provide different depths of objects in focus with a single image. Recall, that the light field, without any

occlusions can be found from knowing one light field and applying a change of basis to the parameters.

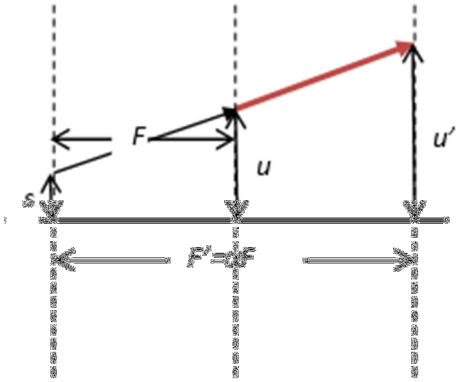


Figure 51: If light field $L_F(u,x)$ is known at, then light field at $L_{F'}(u',x')=L_F(u,x)$. Note that this is a 2D slice of the 4D hyper plane.

That is, if an image plane is placed at *F*, as shown in Figure 51, and then moved to *F'*, the light field at *F'* is given by,

By moving the image plane to *F'*, and recomputing the light field, by a simple change of basis focuses objects at a different distance than those focused at *F*.

To now produce a photograph, an integral projection given by:

$$I(u,v) = \int_0^{\frac{N_y}{2}} \int_0^{\frac{N_x}{2}} \rho L(s,t,u,v)\,ds\,dt$$

where *I(u,v)* is the pixel irradiance at pixel *(u,v)*. Ren Ng generalized the Fourier Slice Theorem (22) to show that it would be easier to compute the Fourier Transform

of the light field and apply the change of basis and integral imaging in the Fourier domain rather than in the spatial domain. The advantage of this method lies in the simplicity of computing an integral image in the Fourier domain. The Fourier Slice Theorem states that a 1D slice of a 2D function's Fourier spectrum is the Fourier Transform of the orthographic integral projection of the 2D function. That is, instead of taking the integral to generate the image above, if the Fourier transform of the light field is found, and the slicing operator defined by,

is applied to the Fourier Transform of the light field, then the Inverse Fourier Transform of that new function will be the image that is computed by the camera. This is schematically shown in Figure 52.

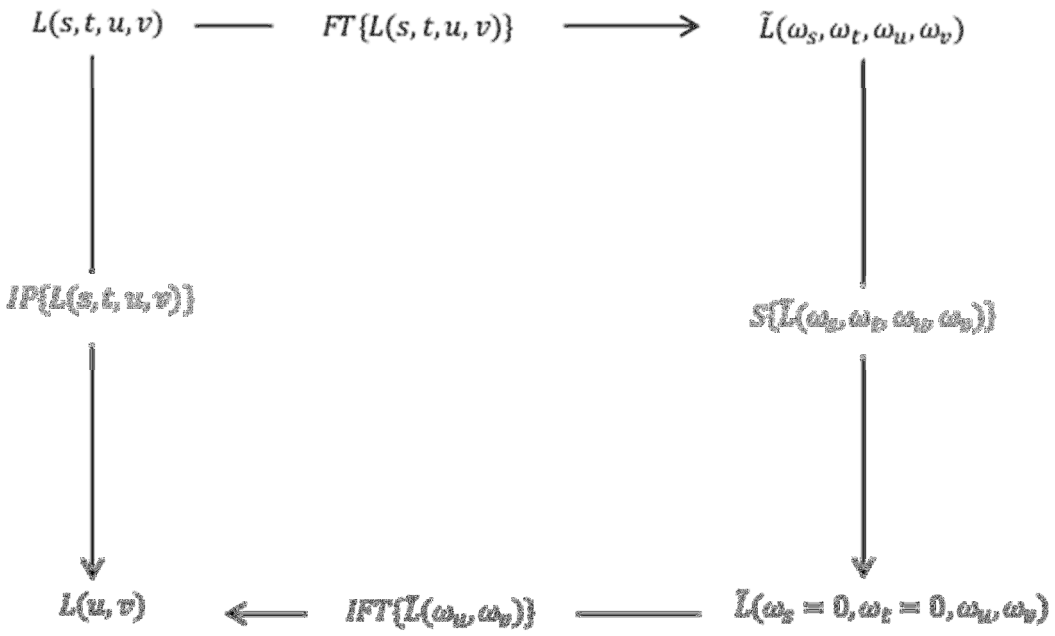


Figure 52: Fourier Slice Photography schematic for light field L. FT = Fourier Transform, S = Slice, IFT = Inverse Fourier Transform, IP = Integral Projection

4.3.2 Simulation Results

As was discussed in the previous section, the light field camera gives the intensity of a ray with two-plane coordinates (s,t,u,v) – (s,t) the coordinates on the first plane, and (u,v) the coordinates on the second, separated by a normalized gap of one.

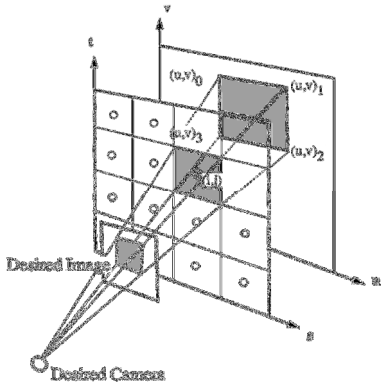
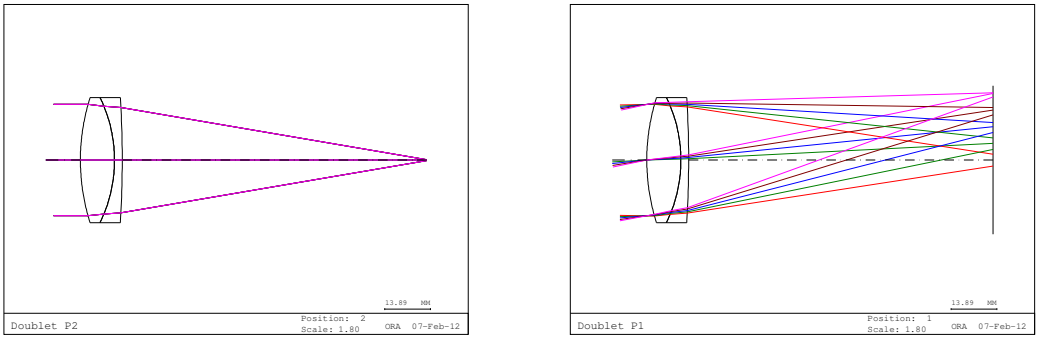


Figure 53: Two Plane Parameterization of Light field (26)

Essentially, the plenoptic camera takes the bundle of rays that a conventional camera adds together at the sensor plane and retrieves the ray information by focusing all angles that hit a particular microlens element to a single point. Objects located at larger distances from the plenoptic camera are always in focus, but as the object moves closer, the lenslets will focus the bundle of rays to the sensor plane so that the ray information from the coordinates (s,t,u,v) can be obtained, where (s,t) is the local coordinate of the lenslet that is used and (u,v) is the pixel coordinate.

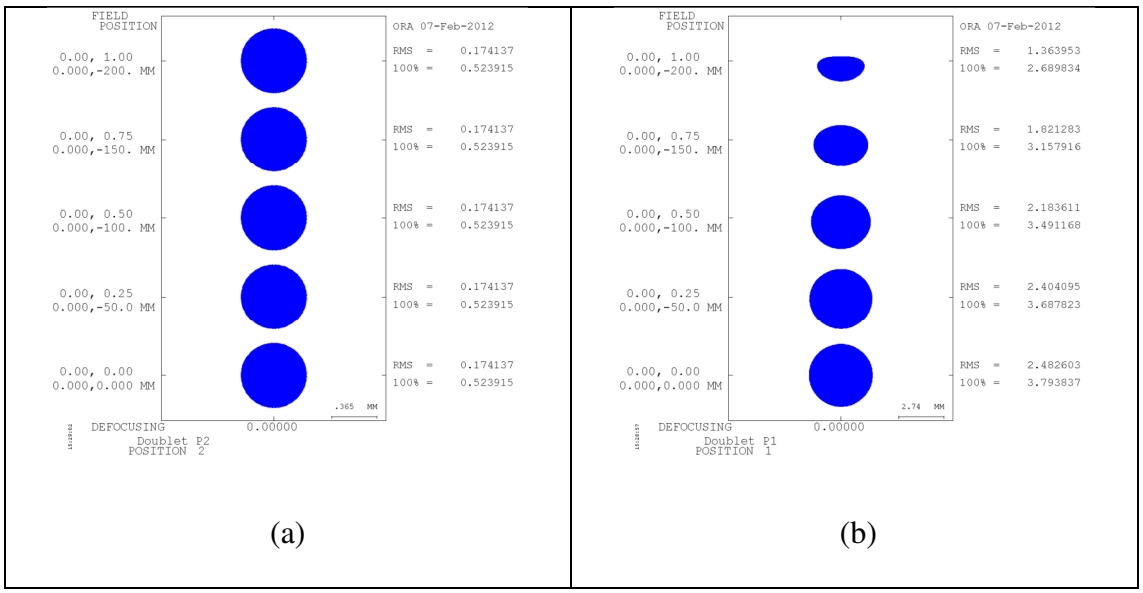
Take the doublet as an example depicted in Figure 54. A doublet is the next simplest lens after the singlet (or regular thin lens) that is designed to reduce aberrations:



(a) (b)

Figure 54: Doublet with height inputs between 0-20cm located at (a) infinity and (b) 1m from the first surface

The spot diagrams in Figure 55 show that as the object approaches the lens, the image gets very blurry.



(a) (b)

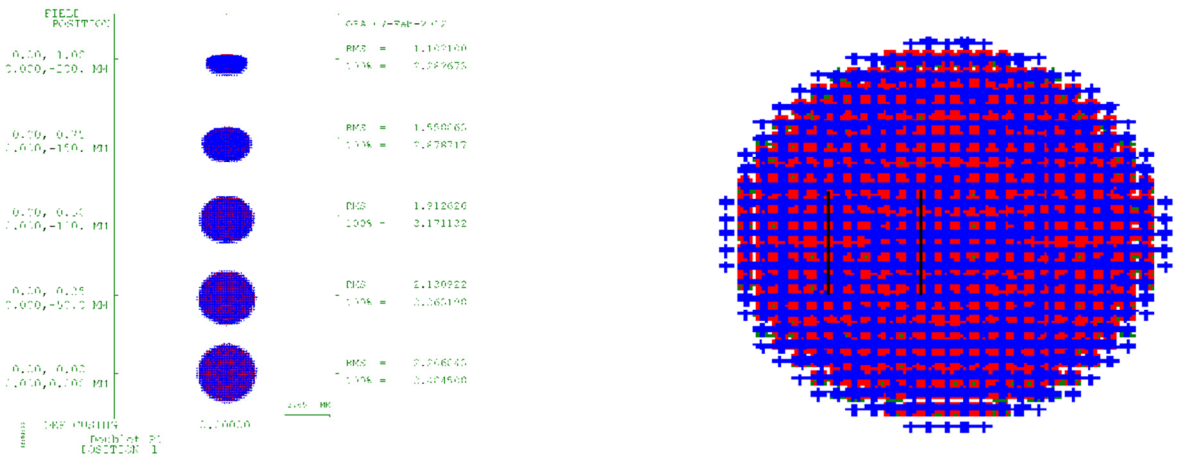
Figure 55: Spot diagrams using doublet lens for object distance of (a) Infinity and (b) 1m from the first surface

Note the 8x scale increase in Figure 55(b) to represent a blurred image. In both Figure 55(a) and Figure 55(b), the sensor plane was brought into focus by placing an object at infinity and using the Paraxial Solve command in CodeV. The object then moves closer to the lens until it is at a distance of one meter from the first surface and

that results in the blurred spot diagram. In addition, the larger object heights of 15 and 20cm have their heights shrunk to add further distortions to the image.

To now model the plenoptic camera for the doublet, a lens array must be placed at the focal plane of the lens. The lens array, as mentioned before, must have its *f/#* match that of the doublet, which is an *f/3* lens. A lens array having an entrance pupil diameter of .15mm and a focal length of .45mm was designed and optimized to meet this requirement. The design followed that of the same micro lens array that was used in the Shack-Hartmann sensor by Thor Labs. To fit the image plane at the simulated object heights, a 296x296 array is used, the same as that used by (4).

The lenslet array samples the spot diagram shown in Figure 14 at the period of .15mm, which is the lenslet pitch.



(a) (b)

Figure 56: (a) Spot diagram of object located at 1m and object heights between 0-20cm (b) Zoomed in view of box to show the periodicity of the spots generated by individual lenslet elements, black lines are 0.8mm apart, yielding a lenslet pitch of 0.16mm

The image formed in the sensor plane shows the micro lens array sampling the wave that is incident on it. The periodicity of these focused rays matches the periodicity of

the lens array and is measured to be 0.16mm, which is close to the pitch of the array that is used for the lenslets.

Note the scale difference of the RMS spot size between using a lens array to focus the light exiting the doublet. The lens array from Figure 13, on the other hand, refocuses the bundle of rays that are incident on it and thus produces more focused images for objects between 10-20cm. In addition, by sampling the pattern, the ray information can be used to digitally refocus objects at different heights that are blurred at different depths as shown in (22)

The focus of this chapter, however, is using the ray information retrieved by the plenoptic camera to identify the input angle of the ray that is coming into the doublet. This will allow the use of a plenoptic camera as a Shack Hartman sensor. Again, as seen in Figure 15, the angle of the ray from the final surface of the doublet and lens array is known by the sampling of the wave front. Inversing the operation shown in the ray equation, yields the input height and angle of the ray that was used to achieve that particular intensity point in the image of the plenoptic camera:

$$\begin{bmatrix} r_{in} \\ \theta_{in} \end{bmatrix} = M^{-1} \begin{bmatrix} r_{out} \\ \theta_{out} \end{bmatrix}$$

where M is the matrix of the system up until the lens array.

Using the inverse equation to determine the input ray angles after going through the aberrations, it is seen that the errors are comparable to those of the Shack-Hartmann sensor in the paraxial approximation.

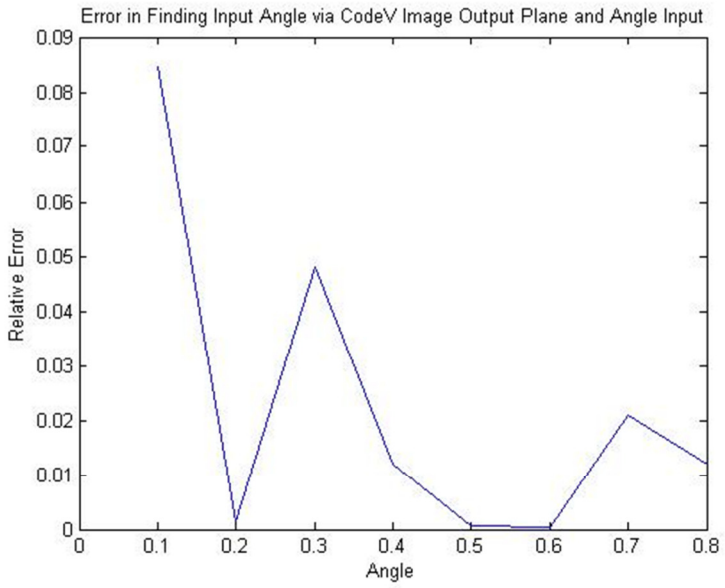


Figure 57: Error in Measured Input Angle versus Input Angle for Plenoptic Camera

The spot diagram of each of the devices gives the degree of measure of blur of a point source whose planar wave fronts are incident at the respective angles with respect to the normal of the first surface.

4.4 Closure

Lenslet arrays are used to retrieve intensity information per ray that is normally lost in a conventional camera. Placing an objective lens in front of the lenslet array gives the plenoptic camera while a collimator placed in front of the lenslet yields the Shack-Hartman sensor. Plenoptic cameras were shown in this thesis to maintain the accuracies of distortion measurement at least as well as the Shack Hartman sensor. Matching the *f/#* of the micro lens to that of the optical objective eliminates the possibility of cross-talk and under-sampling. This concept was used in this thesis to widen the field of view of measurable wave distortions to back track through the optical system to find the wave at the input of the optical objective. The extra

information that can be provided by such an analysis provides the distortions in the object as well as the environment. The algorithm presented in this chapter showed the process of calculating the light field from the source to the sensor, if inverted, it can be used to find the light field at the source knowing the light field at the sensor from the plenoptic camera. This would provide the necessary information needed to send to the deformable mirrors for corrections.

5 Surveillance Systems

5.1 Introduction

Research in vision systems to track a target of interest accurately over a period of time is still expanding at a tremendous rate in the computer vision field. These surveillance systems for localization and mapping can be divided into three parts:

1. 3D-world to 3D-camera coordinate transformations.
2. 3D-camera to 2D sensor transformations.
3. Algorithms performed to identify/classify certain features and/or objects within the 2D image.

Any system is always limited by the quality of its calibration. Camera calibration calls for finding the 11 parameters (6-extrinsic and 5-intrinsic) to identify a function that will take a world coordinate to its image coordinate. Current vision system calibrations (24) (45) (46) primarily focus on using the pinhole model. Nayar in his technical report even says that, “The traditional camera has a detector and a standard lens which captures only those principal rays that pass through its center of projection, or effective pinhole, to produce the familiar linear perspective image. In other words, the traditional camera performs a very simple and restrictive sampling of the complete set of rays, or the light field, that resides in any real scene." (29). As shown in Chapter 3, a real lens can only be modeled as a pinhole with objects placed at far distances from the lens.

The surveillance system presented in this chapter will focus on the model for a PTZ-camera network system whose intrinsic parameters are no longer fixed. This calls for an error optimization over a number of the camera settings to track a specific target

rather than a fixed one. The system architecture introduced in this chapter moves towards dynamic calibrations of the camera in real-time rather than offline video processing, or one time calibrations. These calibrations will be used to track an object at far distances over long periods of time and provide accurate paths traversed by the target.

This chapter will focus on using the pinhole model of a camera for long range surveillance applications. Using a single camera, one can extract a number of features from a room to identify any two dimensional movement, and with a pair of cameras, through stereovision, the depth coordinate can also be obtained. A new novel surveillance technique for tracking of two dynamic cameras tracking a moving object is introduced. This system interpolates the homography matrix between pixels of the master camera and angles on the slave camera for different pan/tilts of the master camera. The master camera will hold the target in a specific region within the image using a homography that will be updated anytime the master camera moves.

5.1.1 Surveillance Systems

The primary use of this system is for tracking a target at far distances as accurately as one can in surveillance applications. Objects that appear to be "suspicious" in a scene should be tracked over long periods of time to determine their position, speed, size, and any number of features that can be obtained from having a series of frames that track the object. These types of problems can be classified into one of three areas:

1. Stationary object tracked by a stationary camera (no fun). This just amounts to finding the 3D position of an object from stereoscopy.

2. Stationary camera tracking a moving object, or, equivalently, a dynamic camera tracking a stationary object. The former is a surveillance system while the latter can be used as an inertial navigation unit.
3. Dynamic camera tracking a dynamic object (this is extremely difficult and is the focus of this research).

The stationary camera solution (47) (48) (49), or the master-slave system architecture with static master camera (50) (51) are well-researched problems and do not satisfy the requirements posed by this thesis. Thus, this chapter focuses on the third method for tracking a target located at far distances. In particular, having such systems operate in real-time is a hurdle within itself (52) (53).

From Chapter 2, it was shown that the MTF of a lens gives a figure of merit as to what object sizes can be resolved at various distances for a certain optical system.

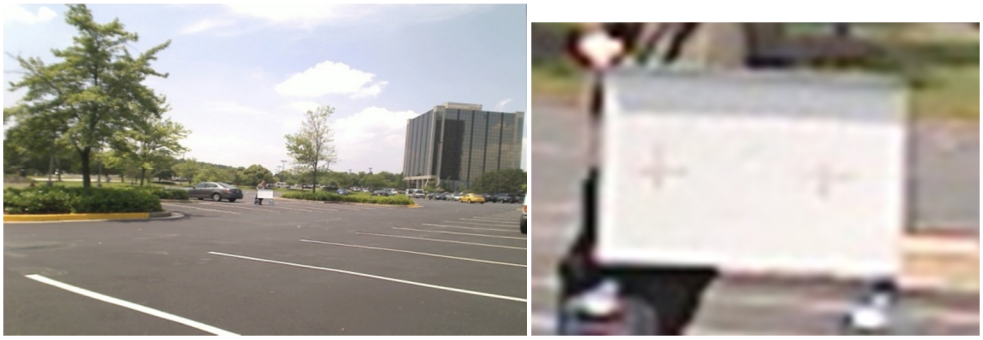


Figure 58: Images taken by Logitech Tessar 2.0/3.7 Webcams at 1600x1200 resolution at a distance of 20m

Low quality cameras do not provide enough data points to carry through the stereo imaging procedure. A camera with a w-pixel resolution imaging an object a length l from the camera EPD dictated by the FOV can resolve a minimum accuracy of w_{obj} given by:

$$w_{obj} = \frac{w}{l\theta_{FOV}}$$

This equation provides the distance w_{obj} that is subtended per pixel, which is the limiting factor in the accuracy. The quality resolution w_{obj} is increased by decreasing the FOV, namely zooming in and operating at a larger focal length. When choosing a camera, resolution should be compared to the physical pixel size of the camera to check that the quantum efficiency is within reasonable limit to provide a good SNR. Details of the quantum efficiency are normally provided by the vendor of the camera and for the purposes of this thesis all were rated to have a level of 25% or above which is common for front-illuminated CCD arrays (54) (55).

Using a 2-megapixel camera (1620x1280) with the tele-photo 22x zoom lens at a field of view of 2x1 degrees when zoomed in provides an accuracy of 2cm at a target range of 100m and height of 10m.

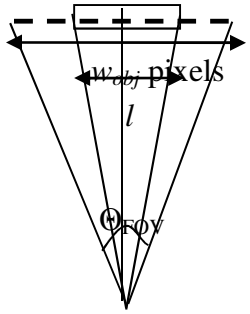


Figure 59: Relationship between object height and pixel coverage

By placing a zoom lens on an Imperx camera with a non-preset 10x 3-Motor Zoom Rainbow CCTV lens the target was able to be resolved at far distances as shown in Figure 60.

Figure 60: Image taken with Imperx 2-Megapixel camera at a distance of 20m

These images were taken at distances of over 20m and the checkerboard is completely visible and markers still detectable. Only 18 images were used in all trials and by increasing the number of images reduces the errors in calibrations.. It is clear that zoom lenses are the solution to the problem of ranging, but the errors need to be minimized, and the system needs to operate in real time. Such lenses would increase the accuracy in any one of the three scenarios by focusing on an object of interest to allow it to take more pixel-space within the image. By increasing the focal length, we are thereby increasing the zoom, and will be able to track objects at far distances, or for that matter, objects that move to far distances, thereby keeping them within the field of view of the camera system over a long period of time.

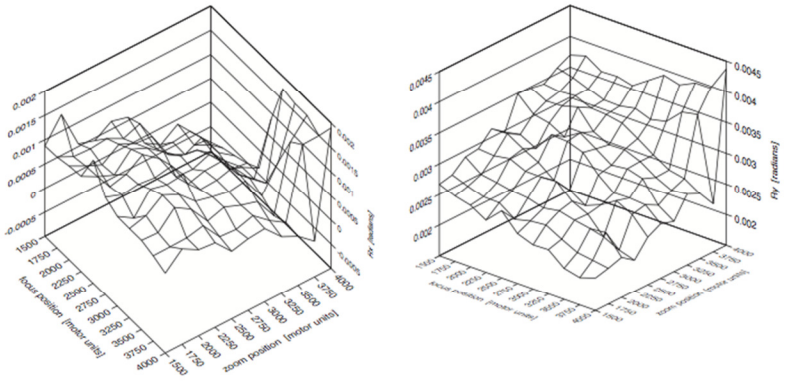
5.1.2 Zoom Lenses

Motorized zoom lenses have great potential in the fields of tracking, vision, and depth mapping (56). These lenses provide images with different magnification that vary with the parameters set by the separation of the lens elements (57). The lenses work together to keep the focus position unchanged as the magnification is varied. The change in the field of view results in a change of the magnification, which changes

the intrinsic parameters of the camera. The goal is now to minimize the variations in aberrations and distortions over a range of settings rather than minimizing them for a single, fixed set of parameters.

Every change in motor settings changes the zoom/focus positions and thus the focal length of the system. With a new set of intrinsic parameters, the extrinsic parameters are also now different. It is clear that rotating 20 degrees at a wide field of view is much different from rotating 20 degrees in a narrower field view, so the extrinsic parameters need to also be recalibrated. For state-of-the art commercial zoom lenses, encoder settings have become very reliable. Chen and Shi (56) claim that a lookup table for zoom lens calibration is not feasible because of the vast number of settings that are available on the camera. They use more of a learned look-up table to form a smaller look-up table. They find the 12 unknown parameters by using a circular pattern rather than a planar one so that they can be symmetrically blurred and the centers of the circles are used as the world coordinates. To adjust for calibration with the varying FOV, larger circles are used when we have a large FOV and zoomed out, while smaller circles are used when FOV is decreased and zoomed. Their results are very similar to those of Wilson's (provided below) and the only major difference is the use of a different pattern for calibration and a bisection method (calibrate endpoint setting, then go half-way) for varying the settings on the zoom lens. It was not at all clear in their paper whether or not they interpolated between the settings that were not within the lookup table or just used the bisection method to insert the new setting within the appropriate location of the look up table if needed (56).

Wilson (45) provides a detailed examination of a methodology for producing accurate camera models as the lens settings are changed. Bivariate polynomials $g_P(S)$, where g is an n^{th} order polynomial, P is a fixed camera model that has the 11 camera parameters, and S is a three-tuple of the lens settings (focus, zoom, and aperture motor settings), are found to interpolate a camera's parameters in real time as the motors are being adjusted. An auto collimated laser was imaged between the various zoom and focus settings, while the aperture was not varied. In other words, only two of the motor settings were adjusted, which allowed him to use a bivariate polynomial. The displacement of the center of the collimated laser was measured to find the planar points for calibration. Zoom and focus settings varied through 11 parameters each, which provided a total of 121 points to measure the 11 camera parameters. To calibrate the camera, at each of these settings, only three images were taken moving the calibration object from 1.5 m, to 2m, then to 2.5m, i.e. only translational movement. The results for some of the parameters are shown below (45):



(a) **(b)**

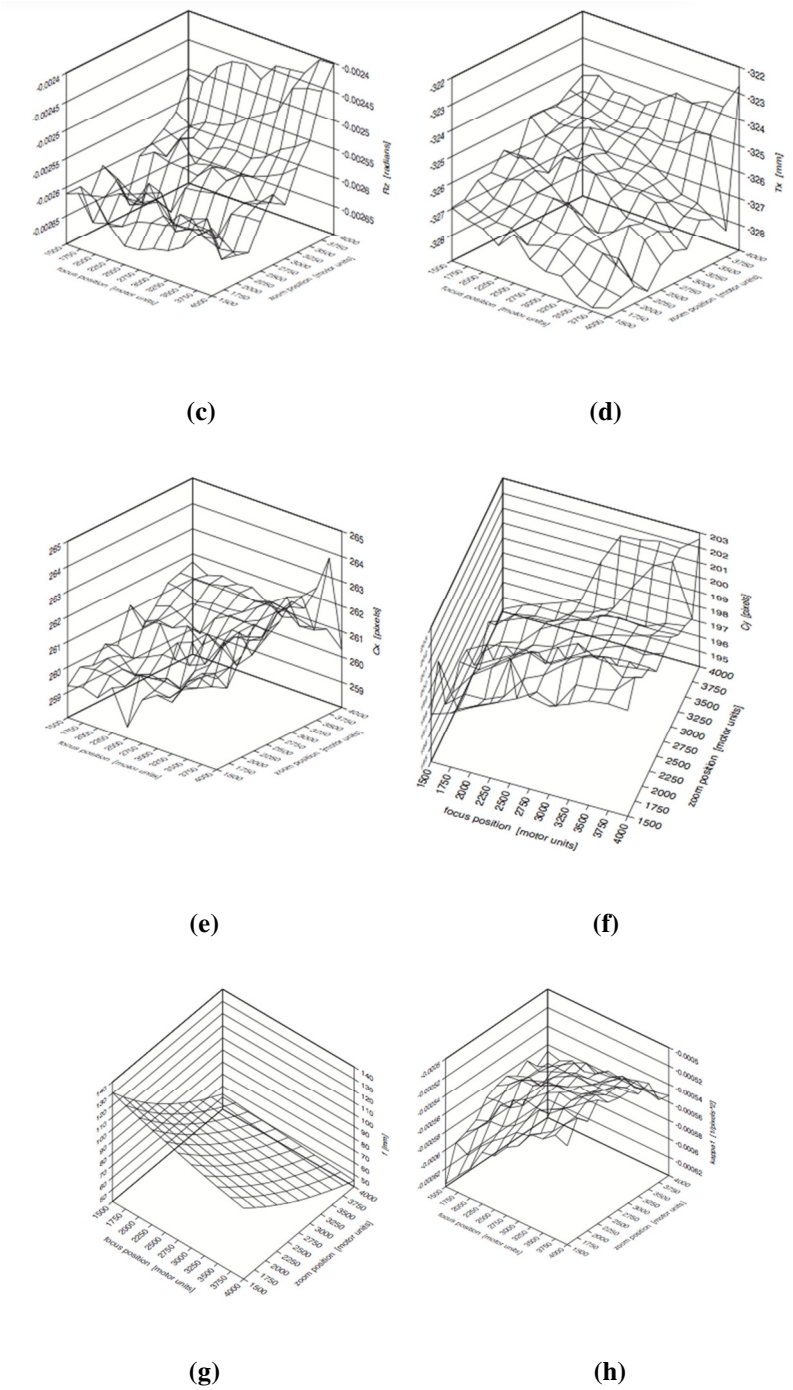


Figure 61: (a) Rx (b) Ry (c) Rz (d) Tx (e) Cx (f) Cy (g) f (h) k1 interpolated versus motor setting (45).

No error plots are provided that explains the accuracy of their parameters, nor are the bivariate polynomials tested to see if providing the input motor parameters would lead to an accurate calibration without the actual calibration procedure. Wilson only

uses three images each for calibration, which from our preliminary tests showed that is not enough data to provide an accurate calibration, particularly with a zoom lens. An automated calibration procedure that makes use of many images for calibration could be used to refine the parameters obtained by Wilson.

Li and Lavest (58) build off of Wilson's work in a more experimental verification of Wilson's theoretical work. They took a series of images over a number of zoom settings and connected the common feature points together in a line which intersected at what they defined to be called the "center of expansion." The major problem with this work is that they assumed the optical axis does not vary "much" as the zoom settings are changed, and by doing so, placed a "hidden" constraint on their system. Using a planar checkerboard on a cart that moved directly backwards and forwards through distances of 0.8-4.5m from the camera, they claim the maximum variation of the optical center is about 0.5 pixels. One of the primary problems we have in our lab with our motorized zoom lens is the shift of the center of the image as the camera zooms in on the object. Lastly, adding to Wilson's work, they concluded that as the aperture is varied, there is also a small change in the focal lengths because brighter objects with darker backgrounds are magnified. This problem can be solved with an auto-iris capability on the lens that keeps the amount of light coming into the camera system at a uniform level throughout the whole image (58).

In a more practical experiment, Lavest, Rives, and Dhome (46) use zoom lenses for 3D reconstruction. Varying the zoom to give the calibration object a "virtual" displacement, the center coordinates of the image are found.

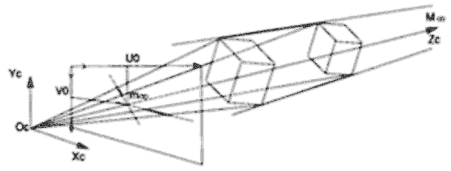


Figure 62: Determining camera center through zooming

Again, this only works if one assumes that the optical axis does not change position within the image as the zoom is varied, and in this paper they claim that the maximum variation is only 0.099 pixels. After finding two of the ten parameters needed for calibration, an interpolation plane is used to extract the other parameters. This is the plane that contains the origin of the camera, image pixel coordinate of interest in (with respect to the camera frame), and *v*, the director of *l*, which is the direction of a line in the image that we have taken (imaged line of a line in the real world). The goal is now to find extrinsic parameters that map the world frame to the image frame and have them lie in this plane. The plane can be computed perfectly because we know each of the points. So really it's back to minimizing:

$$F(\alpha, \beta, \gamma, lx, ly, lz, f_x, f_y, P_w^{ji}) = A_i X_c^{ji} + B_i Y_c^{ji} + C_i Z_c^{ji}.$$

where,

$$\begin{bmatrix} X_c^i \\ Y_c^i \\ Z_c^i \\ 1 \end{bmatrix} = [T_{lx,ly,lz}].[R_{\alpha,\beta,\gamma}].\begin{bmatrix} X_w^i \\ Y_w^i \\ Z_w^i \\ 1 \end{bmatrix}$$

Then the paper goes through some mathematics in finding this through a quadratic criterion and the Taylor polynomial of *F*. These parameters were all found based on the pinhole model of the camera. To model the relationship between the actual zoom lens (a thick lens model), they claim that the effective focal length of the system is

given by using the parameters found in the pin-hole model while sandwiching the principal planes of the lens together (46).

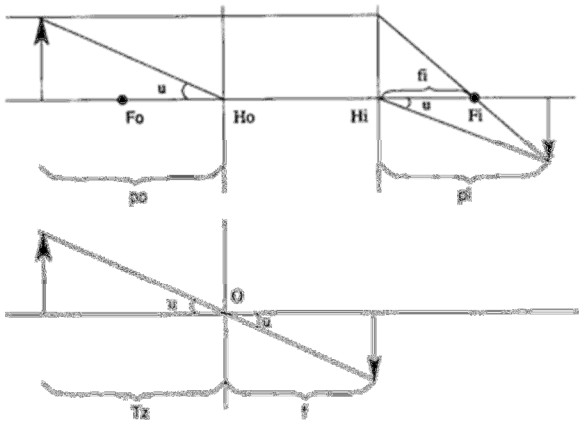


Figure 63 Relationship between the thick-lens model and the pinhole model (46).

$$\frac{1}{f_i} = \frac{1}{p_o} + \frac{1}{p_i} \qquad p_o = T_z \qquad p_i = f.$$

In their experiment, they used a CIDTEC CID2250, 512x512 pixels spaced 15um x 15um camera, and a Angoninex zoom lens T14x9 BIESM, f/1.6, about 13x (from 9 – 124mm). The total size of their checkerboard was 100mm, which is far too small to be able to track objects at far distances. To test the accuracy of their results they reconstructed a cube and had errors from 0.2mm to 8mm., which is rather significant compared to the size of the calibration object they are using.

In the system architecture presented in this chapter, FPAT like concepts in (59) are used to track objects moving on a plane at far distances.

5.2 System Design and Architecture

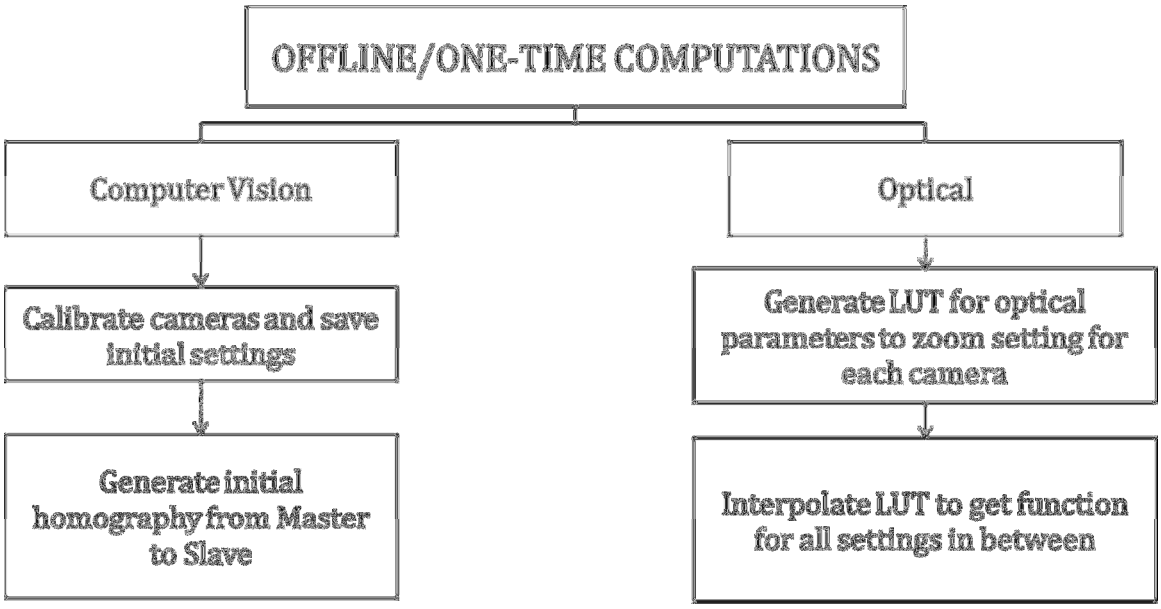
Master-slave relationships between a wide field of view camera and a narrow field of view camera are commonly used in surveillance setups to track an object (53) (60) (61). Rarely, though, is work found using both master camera and slave camera

movement for target tracking. With movement in both cameras, background subtraction is very difficult to use, if not impossible. A solution is to turn off background subtraction while the master camera is moving and since it has a wide field of view, one would not expect it to move much anyway. Thus, a series of homographies can be found between the master and slave cameras for a number of different pan/tilt settings of the master camera. This calls for building a look up table and interpolating that look up table to find the respective homography.

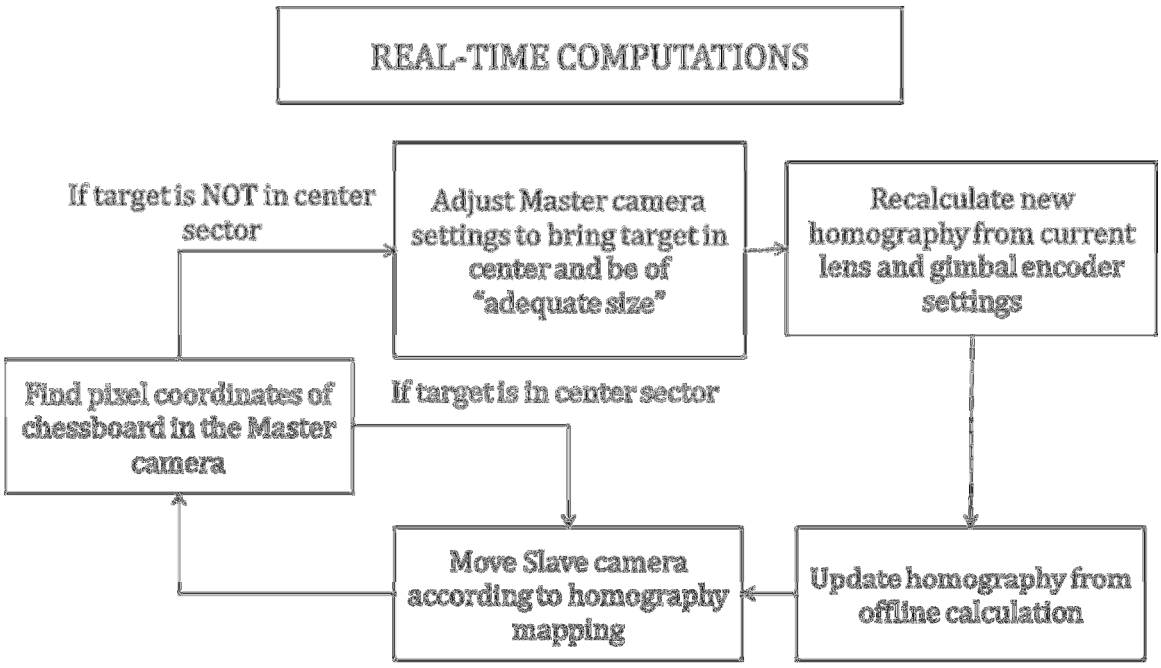
Look up tables (1) and interpolation functions (62), (63) **,** (56) are common tools used to navigate through the different settings to find the optimum setting for target tracking. Figure 61 shows Wilson using an interpolated look up table to find the intrinsic parameters of the camera at a particular motor setting. Reference (10) interpolates a look-up table for a static master camera that is guiding numerous slave cameras. Essentially, a constraint is placed on the target such as the percentage of the image it must cover, or the centering of the target within the image at all times, or a combination of the two and the intrinsic/extrinsic parameters are varied to find the optimum setting that best satisfies these constraints.

Most research nowadays in surveillance systems is often done by post processing on a series of video images instead of in real time because the required computational complexity cannot be achieved in real time. This chapter presents a system that throws most of the processing onto the hardware rather than running complex image processing algorithms. High resolution zoom lenses with presets and gimbals are used to "recalibrate" the system when settings are changed. As the system changes the intrinsic and extrinsic settings, the encoders on the lens and gimbal can be read off

to provide the new "calibration" settings. Thus, the calibration only needs to take place once in the beginning and is then interpolated to find the optimum settings while the target moves through the field of view of the master camera. A system diagram is shown in Figure 64:



(a)



(b)

Figure 64: Tracking procedure for PTZ camera system with (a) offline and (b) real-time computations.

As can be seen from Figure 64, this system is divided into two primary parts: the offline calibrations between master/slave cameras and the real-time tracking of a target.

5.2.1 Off-line/One Time Calibrations

To minimize the amount of image processing needed and thereby reduce the computational complexity of the problem, feature detection should only be done in one camera. Once the target is localized (pixel-wise) in the image of the master camera, a coordinate transformation can be applied to have the slave camera center the target within its image.

Simulation conducted design uses a perfect world with no noise to calculate the homography matrix. Essentially, points in the world are mapped to the image of the

master camera via a projection matrix and rotation matrices are chosen for the slave camera to have that point fall in the center of its image. The setup is shown in Figure 65.

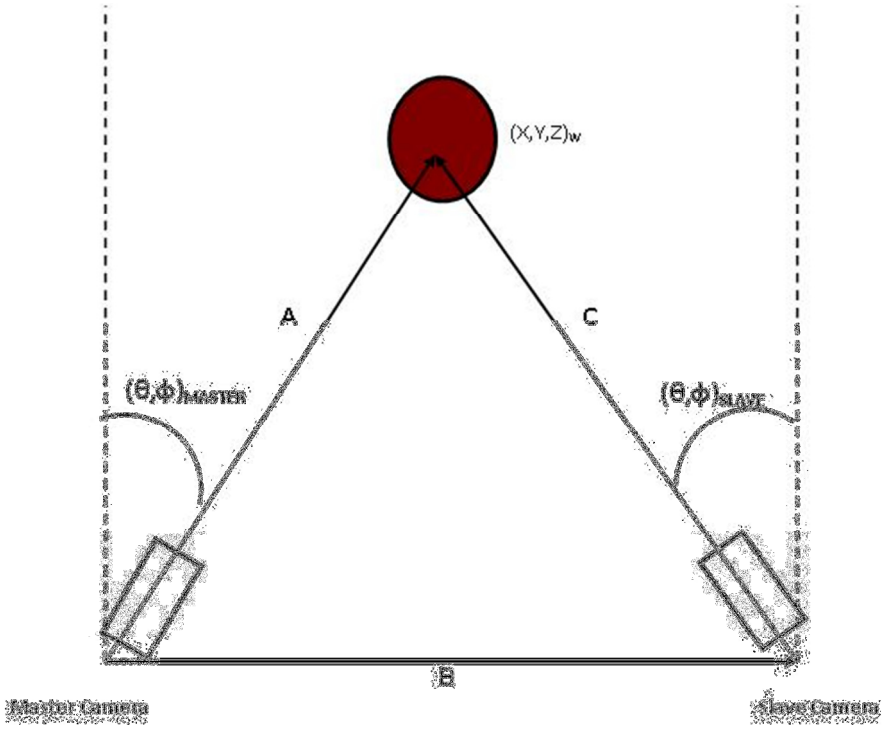


Figure 65: Initial guess to find the slave camera angles from master camera and baseline

An initial guess can be derived for the slave camera to point to the world point by finding the vector C:

$$\vec{C} = \vec{A} - \vec{B}$$

The projection matrix is then optimized to bring that world point into a region of within 10 pixels of the center of the slave image. A maximum of five bounces is allowed if the camera begins to hover around the world point as it tries to bring it within the center of its image. The pair of (x,y) retrieved from the master camera and (p,t) from the slave camera form a calibration point. This is repeated nine times and a homography matrix is found.

$$\begin{bmatrix} p_{slave} \\ t_{slave} \\ 1 \end{bmatrix}_{Zoom\_Slave} = \frac{1}{\lambda} H \begin{bmatrix} x_{master} \\ y_{master} \\ 1 \end{bmatrix}_{Zoom\_Master}$$

At particular zoom settings of each camera, a homography between the pixels of the master camera and the pan/tilt settings of the slave camera to bring the (x,y) calibration point to its center is found. The exact same algorithm to find the camera matrix from Chapter 2 was used to find this homography. Nine points were chosen to have a good spread over the entire image in the master camera and each of those points were centered in the slave camera. Since the lenses that were used in both the simulation and experiment were telecentric, once the slave camera is calibrated for at a specific zoom, the same homography mapping can be used for other zoom settings for targets that are "far enough" away. Errors will arise from the fact that all lenses always have some shift in their optic axis (58) as they zoom in but if the target is far enough away then its image shift will be minimal.

After the homography is found for various pan/tilt settings of the master camera, they must be interpolated for the tracking phase of the system. Figure 66 shows the elements of the homography matrix for various pan/tilt settings of a master/slave camera setup with a baseline of 1.5m, focal lengths of 33mm and 100mm, for an area that is 70x70m at a range of 150m. The blue points show the calibration values chosen while the surface plot is a linear interpolation through the points. The reason for choosing a linear interpolation for the data will be explained in the tracking phase of the system.

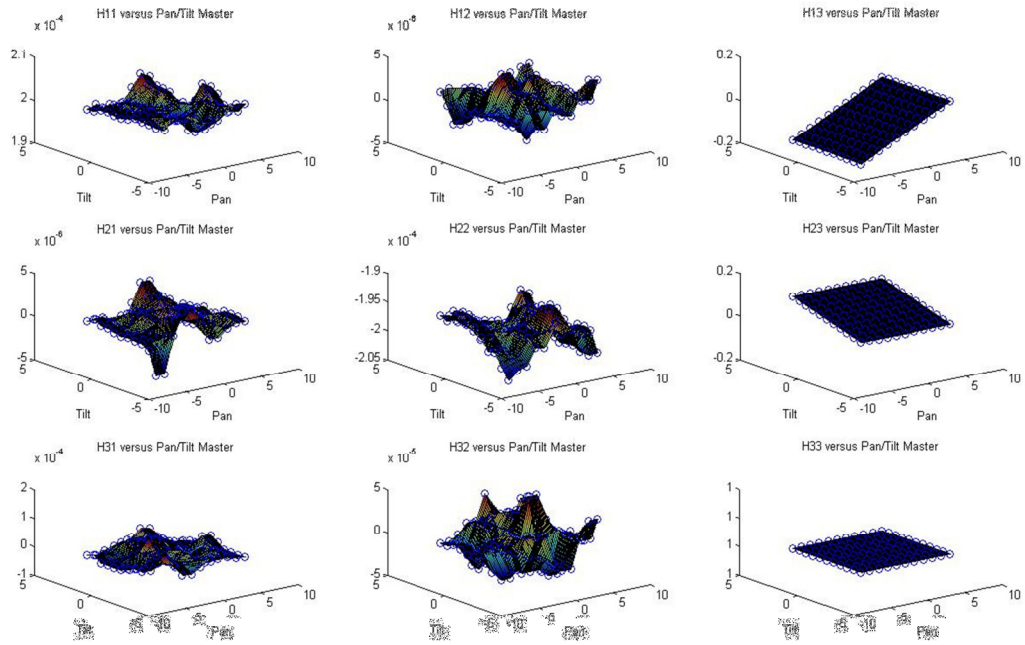


Figure 66: Homography matrix elements for various pan/tilt settings of the master camera

Practically, since the world point is not known, the slave camera is manually moved to bring a point in the world within the center of its image. Once a particular point that falls in the master camera's field of view is chosen, the pixel coordinate in the master camera and the pan/tilt settings from the encoders of the gimbal are saved into an array. This is repeated for nine different points throughout the field of view of the master camera and SVD is used with a linear interpolation to calibrate the homography for various pan/tilt settings of the master camera.

5.2.2 Real-time Tracking and Ranging

A large region of interest about the center defined in the master camera on every frame checks to ensure the target stays within its boundaries. The pan/tilt settings of the master camera are adjusted as the target moves above/below or to the left/right of this region of interest. The increment of adjustment used is the same as that of the

linear calibration to ensure accuracy in the homography being used. Ideally, the master camera should not be moving much since it has a wide field of view and thus using the linear interpolation between these schemes is okay.

Although the cameras are set in a master slave relationship, the gimbal encoders from each camera are independent of one another. This amounts to having two independent, different point of views of the same scene, which is stereovision. Determining range from such a setup can be approximated by a homogeneous linear method of triangulation, which often provides acceptable results. Its advantage over other methods is that it can be easily modified when additional cameras are added, a requirement of this system (64).

Figure 67 uses the interpolations from Figure 66 to extrapolate the world coordinate of a target relative to the master camera as it travels through space and plots the errors in the respective coordinates.

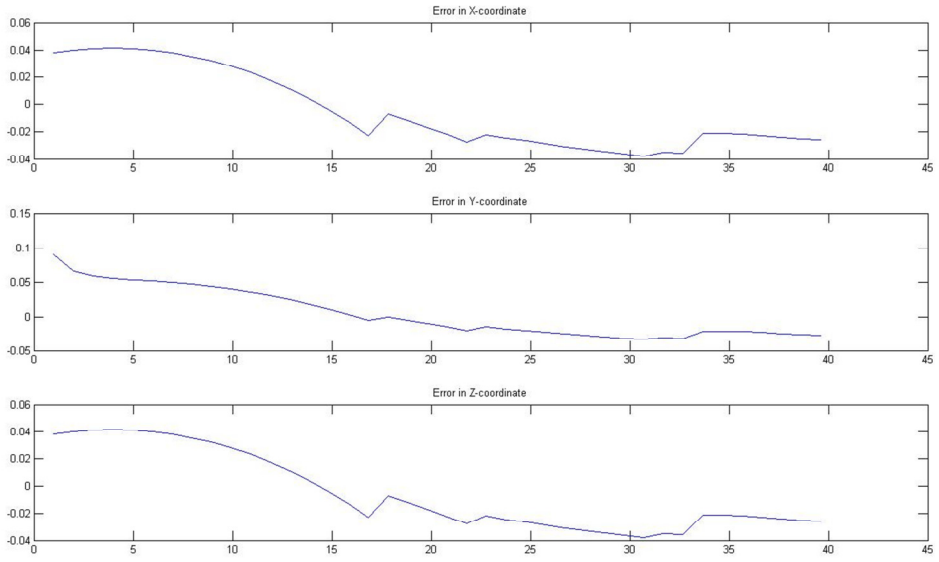


Figure 67: Relative error in the coordinates for a random walk in the calibrated environment

From the simulation it can be seen that the target was a few pixels away from the center of the slave camera (on the order of 8-10 pixels) when it moved away from the plane of calibration. The calibration object used to find the homography between pan/tilt slave settings and pixel master settings for various pan/tilts of the master camera was placed at 150m. It was found that when the target range deviated further from 150m, the errors become worse. As shown in (59) one must ensure that the target remains on a plane so that all of the points of interest are coplanar. The simulation run, however, deviated a bit from this assumption to account for a target that slightly moves off of the plane of calibration 5m on either side.

Figure 68 repeats the same measurements with the addition of white pixel noise to the camera matrix mapping of the slave camera for various baseline measurements.

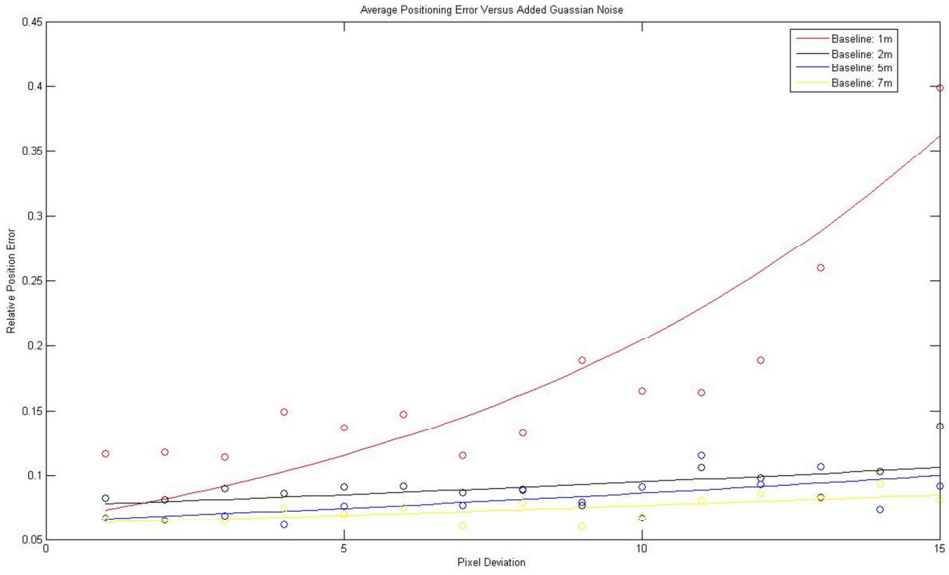


Figure 68: Positioning error with additive noise at different baseline measurements. Larger baselines compensate for the error produced by the noise

To compensate for the noise at longer ranges a larger baseline should be used. Often, increasing the baseline between cameras leads to rather significantly different lighting

conditions and thus corresponding features within two images is difficult. An advantage of this system is that it does not need to worry about corresponding features between cameras since the homographies will all be precalibrated for manually. So long as the target is found in a single camera, the second camera will follow the target.

5.3 Experimental Results

5.3.1 Hardware Setup

The setup in the lab consisted of a Fujinon C22X23R2D-Zp1 Motorized Zoom lens with presets to ensure that our settings operate in a controlled loop in position mode rather than velocity mode. The lenses were equipped with 16-bit encoders to accurately calibrate for the focal length using the MATLAB camera calibration toolbox (12) at a number of zoom settings fitting the model to the commonly used exponential model between zoom/focus settings and focal length. The plots retrieved are similar to those shown earlier by Wilson (45) and other surveillance papers that have optical zoom capabilities (60) (53).

The cameras used are two Imperex GE1050C 1-Megapixel 60fps, Gigabit Ethernet machine vision cameras. It was found that compression added too much noise to the image for the accuracies and ranges the system was designed for thus the raw image format from the video stream of the cameras was used for the measurements/track. The system was operated at 30fps at 810x610 resolution and the Gigabit Ethernet connection allowed the system to operate in real-time.

Figure 69: (Left) Fujinon C22X Optical Zoom Lens 23mm-506mm (Right) GigE Machine Vision Cameras

The gimbals used are designed in-house with two servo motors in the pitch and yaw orientations. The gimbals are driven with two direct-drive brushless servo AC motors with 20bit absolute encoders that readout 0.000343degree resolutions and are interfaced with Gigabit Ethernet connections. They are equipped to hold 50lbs and have a 0.002degree positioning repeatability at that weight. The full system is shown in Figure 70.

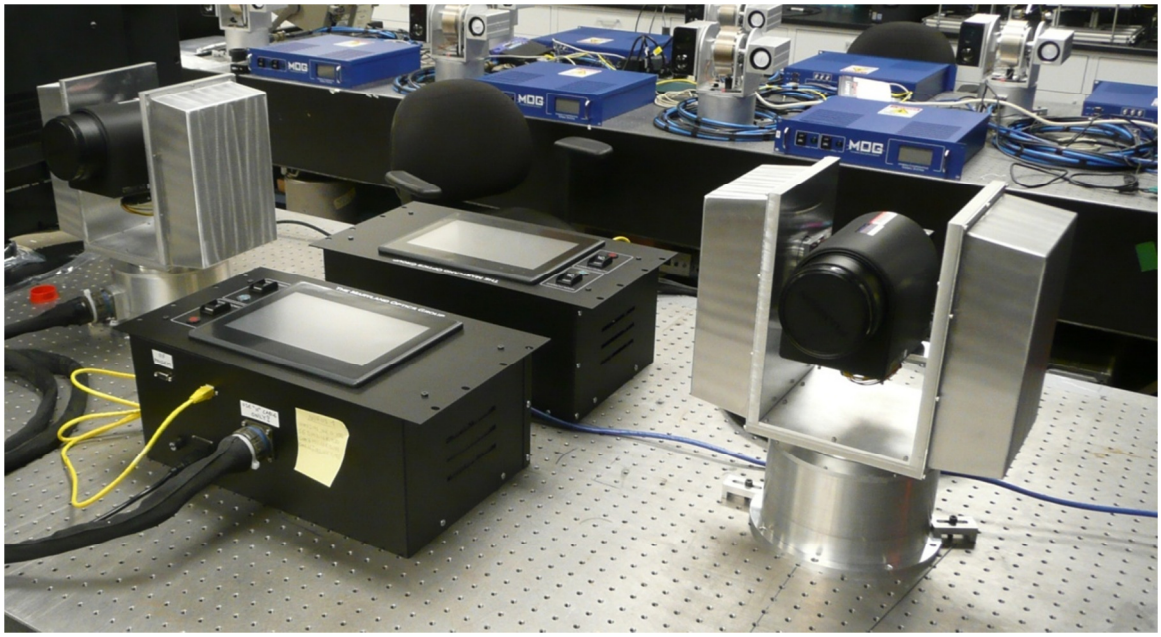
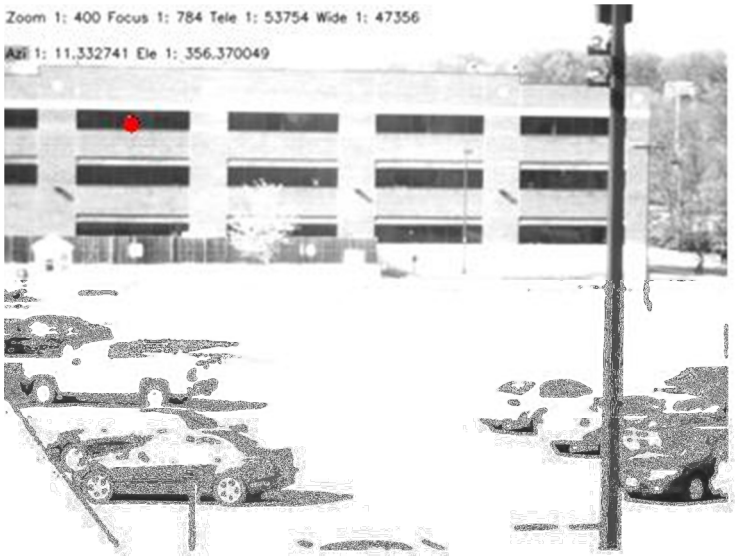


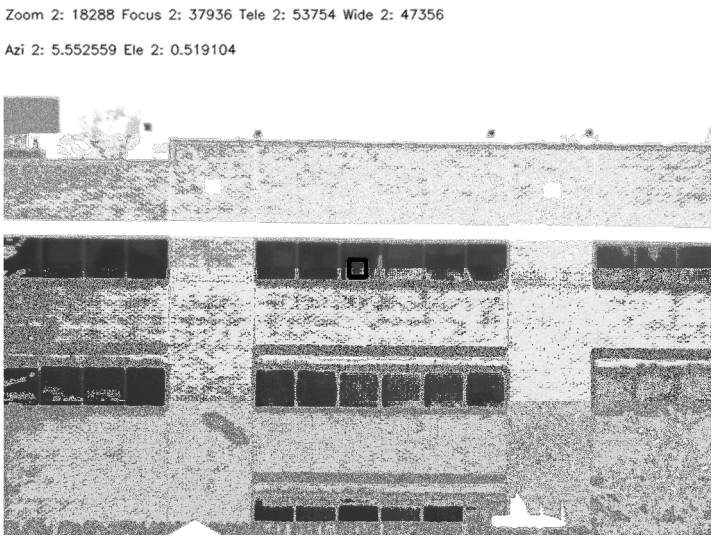
Figure 70: Surveillance system in lab setup

5.3.2 Ranging Results

In the calibration phase, a slave camera is zoomed to a 70mm focal length and manually scans the environment. A 20-pixel wide box is drawn about the center and that feature is found in the master camera. The pair is selected to give the pixel-angle relationship at a particular master camera orientation. In the tracking phase, the pixel coordinates are selected in the master camera at its orientation and the homographies found are used to have the slave camera center the point selected in the master camera within its image. After the slave camera has a fix on the target, before moving to the next point, the world coordinate relative to the master camera is found.



(a)



(b)

Extracted Position Coordinate

X= 15.1078m

Y= 3.8791

Z= 164.5592m

Theoretical Position Coordinate

X=14.2m

Y=4.2m

Z = 170m

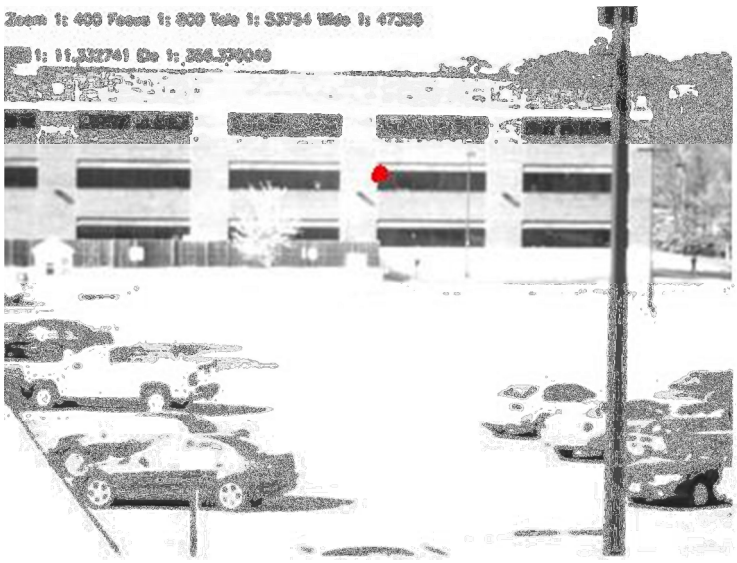
Error In Coordinates

ΔX =0.9078m

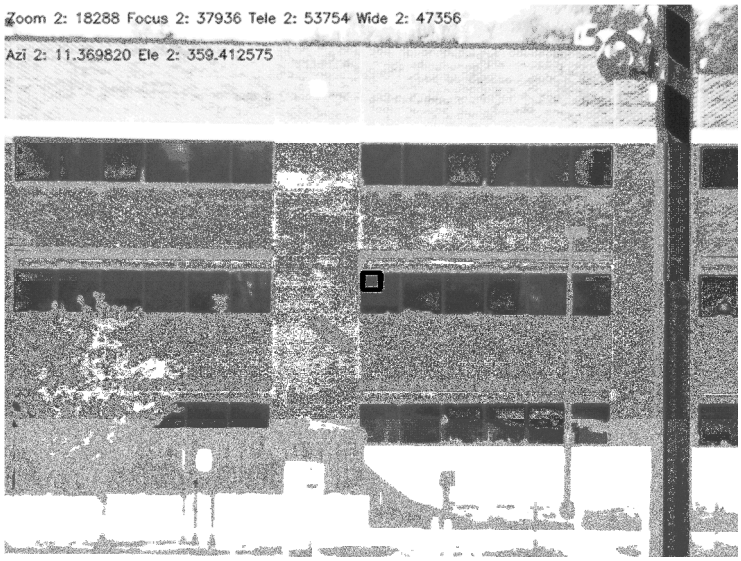
ΔY =-0.3209m

ΔZ =-5.4408m

Figure 71: (a) Master Camera looking at a building with its point selected shown in red (b) Slave camera centering that point within its image and computing the position relative to the master camera.



(a)



(b)

Extracted Position Coordinate

X=35.0589

Y=1.3408m

Z=169.9056

Theoretical Position Coordinate

X=30m

Y=2.8m

Z = 170m

Error In Coordinates

ΔX =5.0589m

ΔY =-1.4592m

ΔZ =-0.0944m

Figure 72: (a) Master Camera looking at a building with its point selected shown in red (b) Slave camera centering that point within its image and computing the position relative to the master camera.

Figure 71 and Figure 72 show experimental results obtained in University of Maryland Kim Building of the system described. The Bimolecular Services Building across from the Kim Building was used as the plane for calibration and points were

selected in the tracking phase to center the slave camera. Google Earth was used to find the range of the building relative to our lab and these were compared to the results given from the camera system. Google Earth's numbers were also tested with a GLR225 Bosch laser range finder and they showed similar results. The (X,Y) positions are roughly estimated based on the size of the windows on the building, which 1.2m wide by 1.3m high. Parking spaces were measured at these distances and yielded an average of about 3m spacing, which compared well to the actual size of the parking space. Various measurements made at different distances from the camera yielded similar results as shown in Figure 73.

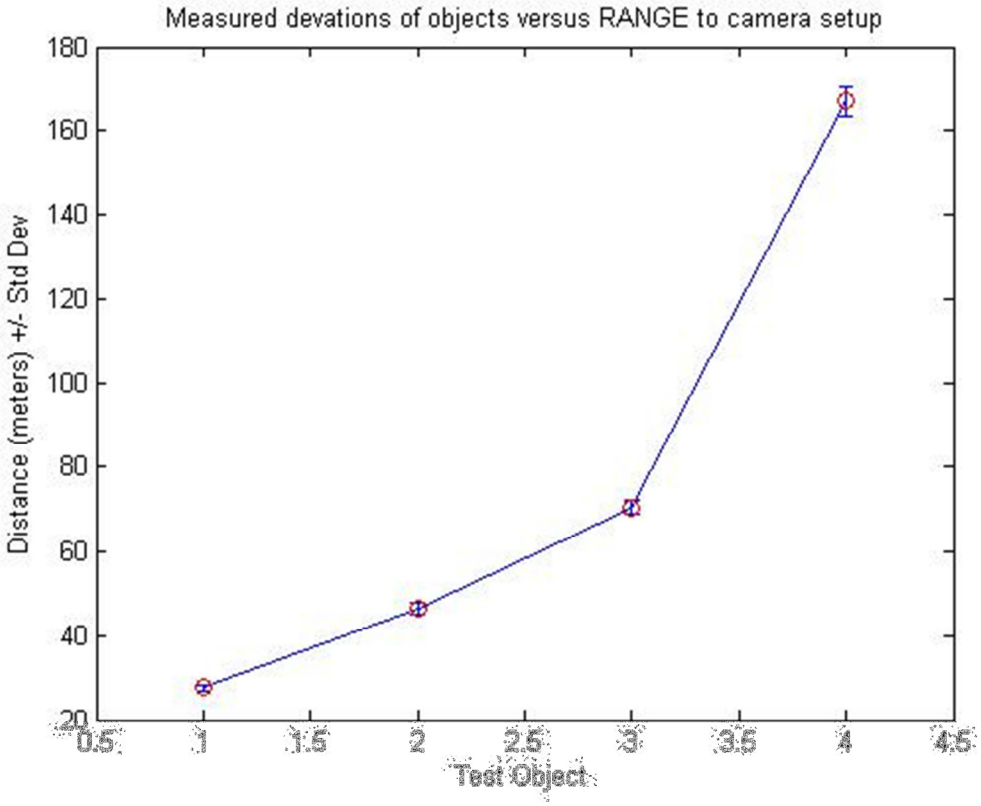
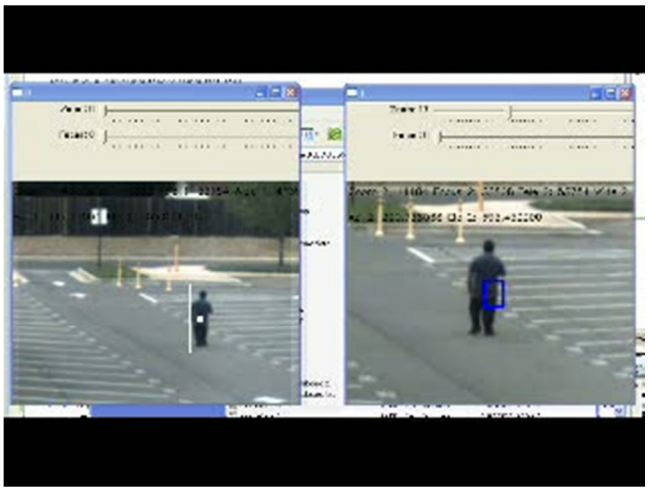


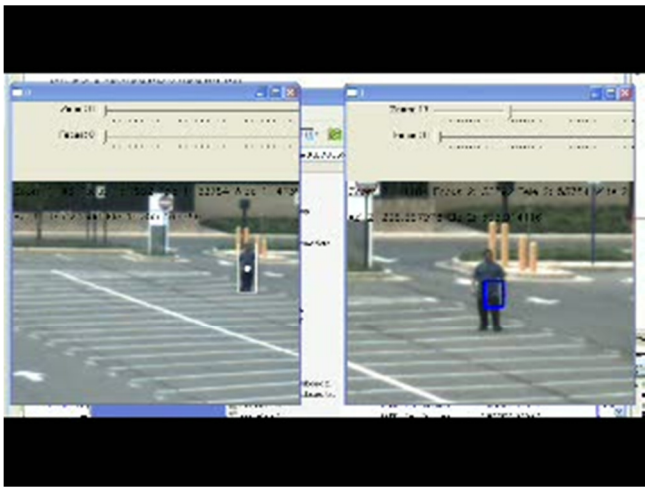
Figure 73: Results found by choosing 10 points around target and finding standard deviation of estimated range

The results in Figure 73 were found by choosing targets at the measured distances shown in the y-axis (specifically a light post, a parking space, and two arrow signs on the road). The error bars shown are the standard deviation of taking 10 points close to the target and comparing it to the actual distance measured by a laser range finder from the feature point to the camera system.

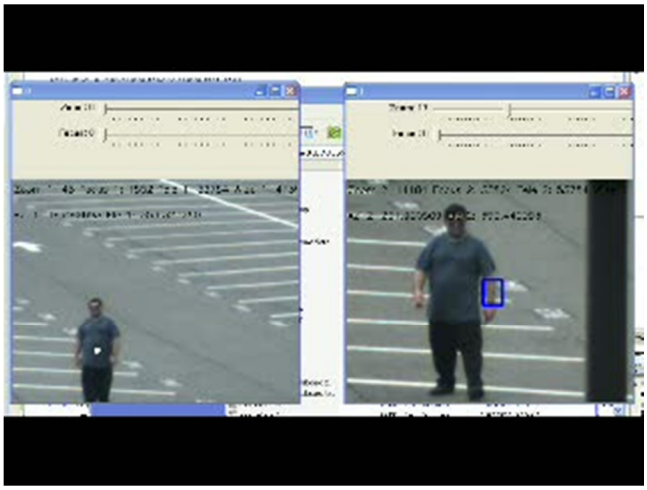
Testing the surveillance setup in real-time (30 fps) at 405x305 resolution to track a target also showed excellent alignment capabilities.



(a)



(b)



(c)

Figure 74: Surveillance system after calibration tracking a target moving from (a) Region 1 to (b) Region 2 (c) Region 3

There were a number of false positives detected (an average of two) when the master camera adjusted its setting to bring the target back within its region of interest. This can be fixed by increasing the number of learning images needed to detect a background so that new objects within the scene are not considered as foreground objects. Increasing the number of images to find a background, however, does increase the latency in tracking the target with the slave camera.

5.4 Closure

This chapter has provided a novel technique for surveillance and ranging of targets in a dual PTZ system at long ranges, which only requires one set of initial calibrations. All processing power is thrown onto the hardware to recomputed camera matrices, and thus the computational expense from complicated image processing algorithms is avoided. Ranges can only be determined as well as their calibrations permit, and the MATLAB toolbox showed significant deviations in focal length and principal point

calculations of the zoom lens. A more sophisticated model would be to place two plenoptic cameras on PT platforms with an objective that has zoom capabilities to have the different optical parameters of a thick lens taken into consideration.

6 Conclusion/Future Work

This thesis uses light field analysis to merge together the fields of computer vision and optics. Its contributions, once again, can be grouped into three general categories:

1. Light field analysis to compare optical systems and provide a figure of merit when models are simplified.
2. Use of light-field cameras in adaptive optics and its simulated comparison to the Shack-Hartmann sensor, in specific.
3. A surveillance system that tracks a target with two dynamic PTZs to extract the world coordinate relative to the camera setup in real time.

This chapter recaps the novelties presented in this thesis for each of the areas mentioned above and concludes each section with a direction for numerous applications that are yet to be investigated. In conclusion, light field analysis and computational photography is the direction that cameras will be taking in the future. Applications of such fields are finding their way into applications in both adaptive optics as well as surveillance systems.

6.1 Comparison Of Optical Models

The optics that make a picture a picture are too often oversimplified. To address these issues, an accurate model for the camera and lens system is needed. Most research nowadays assumes a pinhole model (or perspective imaging model) for the cameras that linearly maps the world coordinates to the image pixel coordinates. Many researchers have worked radial distortion into this model but do not include an analysis of optical effects such as spherical and chromatic aberrations, astigmatism,

and coma. The camera, which primarily has fixed intrinsic parameters, then uses either a model plane like a checkerboard, or other known pattern placed within the scene, near the camera to find orientation, translation, and the optical parameters of the camera (focal length, skew, image center). After all, when purchasing a camera, the bulk of the expense is going towards the lens that accompanies the camera, so finer models are needed.

Lens designer programs such as CodeV provide detailed ray traces to understand what happens to the light as it passes through different optical elements but require the prescription of the lens for an accurate model. These prescriptions are often kept proprietary to the designer/vendor and are not readily available. This thesis derived an extension to the work of Liang and Shih (5) to analyze an imaging system from optical properties such as the back/front focal lengths, and entrance/exit pupil diameters. Matrix optics was used to transform the light field at the object as it reaches the image plane. It was shown that a simple change of basis whose transformation followed the rules of paraxial optics transforms the light field from the object to reach the imaging plane.

Another novel contribution provided by this thesis was to provide a figure of merit between optical models using light field analysis. It was shown that at longer ranges light fields that are formed from complex imaging systems converge to the pinhole model. Essentially, Chapter 2 derived the properties of the different camera models in a unique manner to show how one evolves to the next. The pinhole only allows for a single ray per object point to reach the imaging system and that is the ray through the pinhole. The thin lens model widens the entrance pupil diameter from a delta

function to include the size of the thin lens but limits each ray to bend once at only one location. The complex lens system is basically the input/output model of the full lens system that encapsulates all radii of curvature, spacing, and apertures of the lens. Using the paraxial approximations, the output height/angle is derived from the input height/angle through a linear transformation, namely the ABCD-matrix. But as the distance to the optical setup gets larger, it was shown that the chief ray is a good approximation for the rays that are entering the system and thus using the pinhole model is valid and it greatly simplifies the mathematics for numerous applications. Bearing in mind that consumer-grade plenoptic cameras (Lytro) have been released, an advanced plenoptic camera system would be the next step to add some practical equipment to these simulations.

6.2 Plenoptic Cameras in Adaptive Optics

As shown in Chapter 4, there are many similarities between plenoptic cameras and the design of a Shack Hartman sensor. By placing a micro lens array before the sensor plane to sample and focus the input wave front extrapolates the local distortions experienced at each point in the wave front. These measured distortions are fed to deformable mirrors to correct the offset angle of the input wave front. Shack Hartman sensors measure the displacement from an initial undistorted reference wave in each sector of pixels that are located behind a particular micro lens. The micro lens having a small enough entrance pupil diameter only can have input rays approximated by the chief ray of the thin lens. The sensor plane is located at the focal distance of the micro lens array, which limits the maximum angle of distortion

that can be measured to the ratio of the entrance pupil diameter (which is already small) and the focal length.

Plenoptic cameras were shown in this thesis to maintain the accuracies of distortion measurement at least as well as the Shack Hartman sensor. Matching the *f/#* of the micro lens to that of the optical objective eliminates the possibility of cross-talk and under-sampling. This concept was used in this thesis to widen the field of view of measurable wave distortions to back track through the optical system to find the wave at the input of the optical objective. The extra information that can be provided by such an analysis provides the distortions in the object as well as the environment. The algorithm presented in Chapter 4 showed the process of calculating the light field from the source to the sensor, if inverted, it can be used to find the light field at the source knowing the light field at the sensor from the plenoptic camera. This would provide the necessary information needed to send to the deformable mirrors for corrections.

A next step here would be to actually fabricate an advanced plenoptic camera with an interchangeable imaging array and test its performance for such applications. The pixel width and spacing is obviously limited in the design of the sensor array and it would be interesting to compare that camera to those used in the Shack-Hartman sensor. The wavelength is also limited to the wavelength within the range of silicon based focal plane arrays. but the increase in potential field of view would be a great addition in measuring wave front distortions. To test this setup a collimated laser beam can travel through a beam expander to form a plane wave incident on the plenoptic camera. Before hitting the plenoptic camera though it will travel through a

heat current in the atmosphere created by a convection heater. The convection heater will simulate a controlled form of turbulence to distort the wave which will be measured by the plenoptic camera.

A laser source will go through a beam expander to form a large collimated plane wave that will experience a circular, controlled convection heat steam in the atmosphere before it hits the plenoptic camera. Such experiments are commonly used to control the amount of turbulence a laser beam experiences to analyze the performance of the measurement sensors. Once completed, more astronomical applications will be considered.

6.3 Surveillance/Tracking Systems

PTZ camera networks are the future of surveillance systems. Calibrations of these cameras between one another as they move over time to track targets over long periods of time remain an important area of research in the field. Target localization can only be as good as the calibrations and this calls for both the aforementioned focus on optical models to better relate world to image coordinates, as well as advanced cooperation schemes between the cameras.

In this thesis a novel system with multiple dynamic cameras to track a target's world point in a master-slave relationship was investigated. As the target moved out of the region of interest in the master camera, the master camera moved to bring the target back into a certain predefined window. Calibrations between pan/tilt settings of the slave camera and pixel settings of the master camera are then updated based on the moves of the master camera to ensure the slave camera keeps the target within its center. The absolute encoders available on the optical system and the gimbals were

then used to treat the problem as two independent cameras to find the world coordinate of the target relative to the master camera in real-time. To improve accuracies, it was shown that the baseline of the system should be increased which is relatively easy to incorporate with the system that is described.

To expand on the system currently running in real-time in our lab would call for an incorporation of image features to be used for correspondence. These vision algorithms are computationally expensive if they are to be run on the whole image, particularly when the video stream is in the form of uncompressed megapixel imagery data coming from machine vision cameras. Finding corresponding features between two cameras in real-time is unfeasible with current computational power. By limiting the search region to a window of interest within the master camera and another window of interested centered in the slave camera would alleviate such computational complexities. In addition, with the larger baselines discussed for improved accuracies, lighting becomes a very important factor in image correspondence. If the baseline is too large, the lighting coming into one camera could show a completely different image of the same scene between the two cameras and correspondence would fail. However, by minimizing the search regions for image correspondence this problem is also minimized.

Numerous applications from target tracking to optical transceiver alignment can be tested with the accuracies measured in our system. Our system was implemented with two cameras and a target. Surely a next step would be to use this setup as a single node within a much larger surveillance network. The network could communication through a secondary control network to pass the target ID, namely the

measured world coordinate and velocities from optic flow measurements, to other nodes for a longer track period. A lower data rate secondary channel transferring small portions of data would allow the network to be implemented in real-time for accurate target tracks. It would also solve numerous handoff issues of multiple objects occluding each other for a while as their paths cross in space since they would have different world coordinates.

7 Bibliography

1. **Badri, Julie, et al., et al.** Camera-to-Camera Mapping for Hybrid Pan-Tilt-Zoom Sensors Calibration. *Lecture Notes in Computer Science.* Berlin : Springer-Verlag, 2007, Vol. 4522, pp. 132-141.

2. *Cooperative Mapping of Multiple PTZ Cameras in Automated Systems.* **Chen, Chung-Chen, et al., et al.** s.l. : IEEE Computer Society, 2009. IEEE Conference on Computer Vision and Pattern Recognition. pp. 1078-1084.

3. **Adelson, Edward H and Bergen, James R.** The Plenoptic Function and Elements of Early Vision. *Comoputational Models of Visual Processing.* 1991, pp. 3-20.

4. **Ng, Ren.** Digital Light Field Photography. [Thesis]. July 2006.

5. *Light Field Analysis for Modeling Image Formation.* **Liang, Chia-Kai, Shih, Yi-Chang and Chec, Homer H.** 2, February 2011, IEEE Transcations on Image Processing, Vol. 20, pp. 446-460.

6. *Modified Shack-Hartmann wave front sensor usin an array of superresolution pupil filters.* **Rios, Susana and Lopez, David.** 12, June 8, 2009, Optics Express, Vol. 17, pp. 9669-9679.

7. *Shack-Hartmann sensor based on a cylindrical microlens array.* **Ares, M, S, Royo and J, Caum.** [ed.] Anthony J Campillo. 7, April 1, 2007, Opitcs letter, Vol. 32, pp. 769-771.

8. *Sorting Method to Extend the Dynamic Range of a Shack-Hartmann wave-front sensor.* **Lee, Junown, Shack, Roland and Descour, Michael.** 23, Aug 10, 2005, Aplied Optics, Vol. 44, pp. 4838-4845.

9. *The CAFADIS camera: a new tomographic wave front sensor for adaptive optics.* **Rodrigues-Ramos, Jose Manuel.** 2009, Proc. Adaptive Optics Extremely Large Telescope, p. 05011.

10. *Heterogeneous Fusion of Omnidirectional and PTZ Cameras for Multiple Object Tracking.* **Chen, Chung-Hao, et al., et al.** 8, 2008 йил August, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 18, pp. 1052-1063.

11. http://opencv.willowgarage.com/wiki/. [Online]

12. **Bouget, J.** http://www.vision.caltech.edu/bouguetj/calib_doc/. [Online]

13. **Zhang, Z.** *A Flexible New Technique for Camera Calibration.* Redmond : Microsoft Corporation, 1998.

14. **Hartley, Richard and Zisserman, Andrew.** *Multiple View Geometry in Computer Vision.* 2nd Edition. s.l. : Cambridge University Press, 2003.

15. *Depth from Defocus: A spatial domain approach.* **Subbaro, M., Surya G.** s.l. : International Journal of Computer Vision, 1994. Vol. 13, pp. 271-294.

16. *Image Based Calibration of Spatial Domain Depth from Defocus and Appliction to Automatic Focus Tracking.* **Park, S.Y., Moon, J.** s.l. : Proceedings of ACCV, 2006. pp. 754-763.

17. *Depth from Defocus Estimation in Spatial Domain.* **Zious, D., Deschenes, F.** 2, s.l. : Computer Vision and Image Understanding, February 2001, Vol. 81, pp. 143-165.

18. *A variational approach to recovering depth from defocused images.* **Rajagopalan, A.N. and Chaudhuri, S.** s.l. : IEEE Transactions on Pattern Analysis and Machine Intelligence, October 1997, pp. 1158-1164.

19. *Single-Image Shape from Defocus.* **Torreao, J.R.A and Fernandes, J.L.** s.l. : SIGGRAPH 2005, October 2005, Computer Graphics and Image Processing, pp. 240-246.

20. *Optical 3D Sensing by a single camera with a lens of large spherical aberration.* **Seiko, Y and Miyake, H.** s.l. : IEEE, 2004. Intelligent Sensors, Sensor Networks and Information Processing Conference.

21. *Light FIelds and Computational Imaging.* **Levoy, M.** August 2006, IEEE Computer Society, pp. 46-55.

22. *Fourier Slice Photography.* **Ng, Ren.** [ed.] Markus Gross. New York, NY : ACM, 2005. ACM SIGGRAPH. pp. 735-744.

23. *A Realistic Camera Model for Computer Graphics.* **Kolb, C, Mitchell, D and Hanrahan, P.** s.l. : ACM SIGGRAPH, 1995. pp. 317-324.

24. **Tordoff, B.** Active Control of Zoom for Computer Vision. [Thesis]. 2002.

25. **Davis, CC., Moore, J., Coplan, M.** *Building a Scientific Apparatus.* 4. s.l. : Cambridge, 2008.

26. *The Lumigraph.* **Gortler, Steven J, et al., et al.** New York : ACM SIGGRAPH, 1996. pp. 43-52.

27. *Light field Rendering.* **Levoy, M. and Hanrahan, P.** s.l. : ACM SIGGRAPH, 1996. pp. 31-42.

28. **Ng, Ren, et al., et al.** *Light Field Photography with a Hand-held Plenoptic Camera.* s.l. : Stanford Tech Report, 2005.

29. **Nayar, Shree K.** *Computational Cameras: Approaches, Benefits, and Limits.* Computer Science , Columbia University. New York : s.n., 2011.

30. **Chai, Jin-Xiang, et al., et al.** *Plenoptic Sampling.* s.l. : Microsoft, 2000.

31. *An Iterative Image Registration Technique With an Application to Stereo Vision.* **Lucas, B.D. and Kanade, T.** s.l. : Proc. Image Understanding Workshop. pp. 121-130.

32. *Methods for measuring small displacements of television images.* **Cafforio, C. and Rocca, F.** s.l. : IEEE, 1976, Transactions of Information Theory, pp. 573-479.

33. *Feature-specific imaging.* **Neifeld, Mark A. and Shankar, Premchandra.** 17, June 10, 2003, Applied Optics, Vol. 42, pp. 3379-3389.

34. **Adaptive Optics Associates.** Monolithic Lenslet Modules. *Micro-Optics - Lenslet Array.* [Online] [Cited: February 28, 2012.] www.as.norhtropgrumman.com/businessventures/aoa/mlm/assets/MLM_Data_Sheet.pdf.

35. *A Tunable Shack-Hartmann Wave front Sensor Based on Liquid-filled Microlens Array.* **Hongbin, yu and Guangya, Zhou, Siong, Chau Fook, et al.** 10, s.l. : Journal of Micromechanics and Microengineering, 2008, Vol. 18.

36. *Historical Development of the Shack-Hartmann wave front sensor.* **Schwiegerling, J and Neal, D.** [ed.] J. E. Harvey and R. B. Hooker. Bellingham : SPIE, 2005. Robert Shannon and Roland Shack: Legnes in Applied Optics. pp. 132-139.

37. **Thor Labs.** SH Tutorial. *Shack Hartman Wave front Sensor.* [Online] December 19, 2011. [Cited: February 28, 2012.] http://www.thorlabs.com/NewGroupPage9.cfm?ObjectGroup_ID=2946.

38. *Estimation of optical turbulence characteristics from Shack Hartmann wave front sensor measurements.* **Brennan, Terry J. and Mann, D.C.** s.l. : Proc. SPIE 7816, 2010. p. 781602.

39. *Spherical aberration correcdtion system using an adaptive optics deformable mirror.* **Schwertner, M. and Booth, M.J, Tanaka T., Kawata, S.** 2, July 15, 2006, Optics Communications, Vol. 263, pp. 147-151.

40. **Cavadore, C.** Shack Hartmann. [Online] December 2003. [Cited: February 28, 2012.] http://astrosurf.com/cavadore/optique/shackHartmann/Shack-Hartmann.htm.

41. **Davis, C.C.** *Lasers and Electro-Optics.* s.l. : Cambridge University Press, 1996.

42. **Shannon, R.R.** *The Art and Science of Optical Design.* s.l. : Cambridge University Press, 1997.

43. *Firefly Capturing Method: Motion caputring by Camera with Large Spherical Aberration of Lens and Hough Transform based Image Processing.* **Seiko, Y. and H., Miyake.** 2006, Intelligent Sensors, Sensor Networks, and Information Processing.

44. *The Zernike Polynomials.* **Tripoly, N.** s.l. : Optikon, 2002.

45. *Modeling and calibration of automated zoom lenses.* **Wilson, Reg K.** s.l. : SPIE, 1994 йил October, Proc. SPIE, pp. 170-186.

46. *Three-Dimensional Reconstruction by Zooming* . **Lavest, Lean-Marc, Rives, Gerard and Dhome, Michel.** 2, s.l. : IEEE, 1993 йил April, IEEE Transactions on Robotics and Automation, Vol. 9, pp. 196-207.

47. *Camera cooperation for achieving visual attention.* **Horaud, R., Knossow, D., Michaelis, M.** 6, s.l. : Machine Vision Applications, 2006, Vol. 16, pp. 331-342.

48. *Human Tracking in Multiple Cameras.* **Khan, S., Javid O., Rasheed Z.** s.l. : Proceedings of IEEE International Conference of Computer Vision, 2001. pp. 331-336.

49. *Real-tiem wide area multi=camera stereo tracking.* **Zhao, T., Aggarwal, M., Kumar, R., Sawhney, H.** s.l. : Proceedings of IEEE Conference of Computer Vision and Pattern Recognition, 2005. Vol. 1, pp. 976-983.

50. *Towards Calibrating a Pan-Tilt-Zoom Camera Network.* **Sinha, S., Pollefeys, M.** s.l. : EECV Conference Workshop, 2004.

51. *Acquiring mult-scale images by pan-tilt-zoom ontrola nd automatic multi-camera calibration.* **Senior, A., Hampapur, A., Lu, M.** s.l. : IEEE Workshop on Applications on Computer Vision, 2005. pp. 433-438.

52. *Dual Camera Zoom Control: A study of zoom tracking stability.* **Nelson, E.D, Cockburn, J.C.** s.l. : Proceedings of IEEE International Conference of Acoustics, Speech and SIgnal Processing, 2007. pp. 941-944.

53. *The Chameleon Like Vision System.* **Zhou, J., Wan, D., Ying,W.** s.l. : IEEE Signal Processing Magazine, September 2010, pp. 91-101.

54. *Design and Characteristics Measurement of High-speed CMOS Imaging System.* **Shang, Yuanyuan, Guan, Yong and Zhao, Xiaouxu.** Shenzhen : International Workshop on Imaging Systems and Techniques, 2009.

55. **Allied Vision Technologies.** *AVT Prosilica GC Technical Manual.* s.l. : Allied VIsion Technologies, 2011.

56. *Camera Calibration with a Motorized Zoom Lens.* **Chen, YS, et al., et al.** Barcelona : IEEE, 2000. International Conference on Pattern Recognition. Vol. 4, pp. 495-498.

57. **Shannon, R. R.** *The Art and Science of Optical Design.* 1st Edition. s.l. : Cambridge University Press, 1997.

58. *Some Aspects of Zoom Lens Camera Calibration.* **Li, Mengxiang and Lavest, Jean-marc.** 11, s.l. : IEEE, 1996 йил November, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 18, pp. 1105-1110.

59. **Ho, Tzung-Hsien.** *Pointing, Acquisition, and Tracking Systems for Free-Space Optical Communication Links.* 2007.

60. *Stereo vision using two PTZ cameras.* **Zhou, J., Wan, D.** 2, s.l. : Computer Vision and Image Understanding, November 2008, Vol. 112, pp. 184-194.

61. *Multi-resolution and wide scop depth estimation using a dual PTZ camera system.* **Wan, D., Zhou, J.** 3, s.l. : IEEE Transactions of Image Processing, 2009, Vol. 18, pp. 677-682.

62. *UAV Attitude Estimation by Vanishing Points in Catadioptric Images.* **Bazin, Jean-Charles and Demonceaux, Cedric.** Pasadena : IEEE, 2008. International Conference on Robotics and Automation. pp. 2743-2749.

63. *Using Vanishing Points for Camera Calibration.* **Caprile, B and Torre, V.** 2March, 1990 йил, International Journal of Computer Vision, Vol. 4, pp. 127-140.

64. **Hartley, R., Zisserman, H.** *Multiple View Geometry in Computer Vision.* s.l. : Cambridge University Press, 2004.

65. **Quimby, Richard S.** *Photonics and Lasers: An Introduction.* Hoboken : John Wiley & Sons Inc., 1996.

66. **Moore, John H, Davis, Christopher C and Coplan, Michael A.** *Building Scientific Apparatus.* s.l. : Cambridge, 2008.