2002

An Analysis of Camera Calibration for Voxel Coloring Including the Effect of Calibration on Voxelization Errors

Elwood Talmadge Jr. Waddell
University of Kentucky, elwood.waddell@eglin.af.mil

ABSTRACT OF THESIS

An Analysis of Camera Calibration for Voxel Coloring Including the Effect of Calibration on Voxelization Errors

This thesis characterizes the problem of relative camera calibration in the context of three-dimensional volumetric reconstruction. The general effects of camera calibration errors on different parameters of the projection matrix are well understood. In addition, calibration error and Euclidean world errors for a single camera can be related via the inverse perspective projection. However, there has been little analysis of camera calibration for a large number of views and how those errors directly influence the accuracy of recovered three-dimensional models. A specific analysis of how camera calibration error is propagated to reconstruction errors using traditional voxel coloring algorithms is discussed. A review of the Voxel coloring algorithm is included and the general methods applied in the coloring algorithm are related to camera error. In addition, a specific, but common, experimental setup used to acquire real-world objects through voxel coloring is introduced. Methods for relative calibration for this specific setup are discussed as well as a method to measure calibration error. An analysis of effect of these errors on voxel coloring is presented, as well as a discussion concerning the effects of the resulting world-space error.

Keywords: Voxel Coloring, Camera Calibration, Calibration Error, Reconstruction, Voxelized Space

Elwood Talmadge Waddell, Jr.
23 July 2002

An Analysis of Camera Calibration for Voxel Coloring Including the Effect of Calibration on Voxelization Errors

by

Elwood Talmadge Waddell, Jr.

Dr. Christopher O. Jaynes
Director of Thesis

Dr. Grzegorz Wasilkowski
Director of Graduate Studies

July 22, 2002

RULES FOR THE USE OF THE THESIS

Unpublished theses submitted for the Master's degree and deposited in the University of Kentucky Library are as a rule open for inspection, but are to be used only with due regard to the rights of the authors. Bibliographical references may be noted, but quotations or summaries of parts may be published only with the permission of the author, and with the usual scholarly acknowledgements.

Extensive copying or publication of the thesis in whole or in part requires also the consent of the Dean of the Graduate School of the University of Kentucky.

THESIS

Elwood Talmadge Waddell, Jr.

The Graduate School

University of Kentucky

2002

An Analysis of Camera Calibration for Voxel Coloring Including the Effect of Calibration on Voxelization Errors

---

THESIS

A thesis submitted in partial fulfillment of the
requirements for the degree of Master of Science
in the College of Engineering
at the University of Kentucky

By

Elwood Talmadge Waddell, Jr.

Director: Dr. Christopher Jaynes, Dept. of Computer Science

Lexington, Kentucky

2002

To Gladys Hughes Chaffin & Jay Tilden Chaffin, the most loving, supportive, noble industrious, interesting, and wise grandparents that a grandson could ever have.

Acknowledgements

I would first like to acknowledge and thank my parents, Mary M. & Elwood T. Waddell, Sr., for their love and support. Without them, neither my life nor my academic career thus far, could have been nearly as enjoyable or beneficial. I would also like to thank my maternal and paternal grandparents, Gladys & Jay Chaffin, and Ruth & Ross Porter, respectively, for their loving support, assistance, and encouragement.

The completion of this thesis has been possible due to the direction and advisement of Dr. Christopher Jaynes. His inputs, advise, supervision, and direction in the past two years have been invaluable. I owe nearly all I know about research in the academic environment to him. He has been a tremendous example, and I cannot thank him enough. Additionally, I would like to thank the members of my Committee, Dr. Fuhua (Frank) Cheng and Dr. Brent Seales, for their comments, revisions, advice and support through this process.

During his tenure as Professor of Aerospace Studies, Col Steve Parker, USAF, was instrumental to allowing me to pursue this work. His example of compassionate and effective leadership is one to which I aspire as an Air Force Officer.

I would also like to thank the many friends throughout the University of Kentucky, the Air Force, and the City of Lexington for their moral support and friendship through the completion of this work. Especially supportive have been many students of the College of Engineering, the Networking and Vision Labs, the 290th AFROTC Detachment and Cadet Wing, the members of the First Free Will Baptist Church of Lexington, the staff of Camp Caleb Christian Camp and Conference Center, and the members and associates of the 36th Electronic Warfare Squadron, 53rd Wing, Eglin Air Force Base, Fl.

Table of Contents

Table of Figures

Table of Files

ETThesis.pdf 570K

Chapter 1
General Problem Overview and Related Work

A primary goal of computer vision is the accurate three-dimensional description of scenes observed by one or more cameras. This goal has been addressed both by research efforts that address the general scene interpretation problem directly [23,24,38,40,41] as well as those that focus on specific, but related, subproblems. A significant subproblem related to this goal is the automatic acquisition of a three-dimensional geometric model from multiple viewpoints. Geometrically reconstructed scenes are capable of supporting other computer vision tasks such as object recognition, segmentation, attention selection, event recognition, and a number of others. Automatically acquired three-dimensional models have utility in themselves, and three-dimensional systems have been developed for a variety of applications ranging from automatic acquisition of three-dimensional city models [6,20,30] to reconstruction of existing artifacts for archival and study [8].

A common element among these reconstruction algorithms is the need for some form of camera calibration. In particular, camera-based systems typically must recover the relative positioning of cameras with respect to one another [15], to the scene [14], or to other devices used in the acquisition process [4,9,11,13,19]. There is significant research related to relaxing the camera calibration required to perform scene reconstruction, e.g. self-calibration [3,12,10,42]. These approaches are promising. Oftentimes the resulting cameras are calibrated up to an unknown projective or affine transformation. This is useful if the reconstructed scene is to be used for visualization purposes or is to be used to measure features that are invariant under the unknown transformations. However, using these methods, Euclidean reconstruction of the scene under observation is not possible. For a

number of applications, such as close-range photogrammetry [18,32], manufacturing, and storage of three-dimensional artifacts for future scholarly study a Euclidean model that supports metric analysis is important.

Given the importance of camera calibration to the three-dimensional reconstruction problem, there have been serious efforts related to accurate acquisition of camera parameters in a variety of contexts [1,31,39,44]. In addition, analytical and empirical studies relating errors in the camera models to expected errors in the world have taken place [2,5,16,17]. This thesis undertakes an analysis of camera calibration in the context of three-dimensional Euclidean reconstruction. In particular, we focus on how camera calibration errors are related to errors in a reconstructed scene that has been acquired using a volumetric reconstruction approach [7,33,34,35,36,37,45,46]. We characterize how error is propagated from camera to voxelization error in the scene for a specific, but common, reconstruction setup. In addition, methods in which camera calibration can be constrained using specific surfaces are introduced. Conclusions about calibration error and expected volumetric reconstruction accuracy are the discussed.

1.1. A Brief History of Volumetric Reconstruction Methods

Volumetric reconstruction refers to the automatic acquisition of a three-dimensional volume that corresponds to an object or scene under observation from multiple cameras. Because these techniques produce one or more closed volumes that describe the geometric shape (and potentially relative position) of one or more regions, they are often applied to acquire a coherent description of an object rather than a scene. That is, volumetric techniques take several views of an object, segment it from the background using several

different approaches, and ultimately construct a closed volume that is likely to have given rise to the set of observed views. One of the more successful approaches to passive, volumetric reconstruction from more than one viewpoint is referred to as the *voxel coloring algorithm*.

The idea of voxel coloring was first put forward by Seitz et. Al [35], in 1998. From its inception, voxel coloring methods have attempted to achieve photo-realistic reconstruction without explicit point matching. This has made it very attractive in modeling and computer graphics, where special measurements are not as important as an accurate reprojection.

The voxel coloring algorithm is an *object-space* approach to reconstruction. Given accurate calibration, traditional stereo reconstruction matches pixels in different views through an image-based matching scheme. For each matching pair of pixels, a corresponding point can be added to or reconstructed in the scene by intersecting the back projected the rays defined by the cameras and image pixels. In this way, a scene is slowly constructed as processing proceeds in *image-space*. Voxel coloring assumes a particular volume divided into finite sized three-dimensional volume elements, called voxels, which are then forward projected into images for processing. In this way, processing takes place in the Euclidean space of voxels rather than directly in the images. Although this difference may seem, at first glance, to be unimportant, processing directly on the volume in object-space has been shown to have several advantages for particular contexts [7,33,34,35,36,37].

In general, the voxel reconstruction algorithm is straightforward. Voxels are iteratively processed to determine which belong in the volume as opposed to those that do not and should be eliminated. Starting from the outermost voxels (those closest to the camera centers), each voxel is projected into all images, regardless of visibility constraints implied

by the current set of voxels in the volume. Once projected into each image, analysis of the corresponding pixels in each view determines the voxel's consistency. This measure is typically related to the variance of the color for each projected region in all of the available views. High variance implies that the voxel is either not present in the data, or is occluded by other voxels, in either case, the voxel is removed from the scene. This carving procedure removes voxels until the remaining volume is consistent in all the images. Once a set of consistent voxels remain in the dataset, a final color assignment for each voxel can be assigned using the available images and camera calibration information through direct interpolation. The general setup of the algorithm is shown in Figure 1.1. For the complexities related to viewing conditions, accurate consistency measures, and logistics of the algorithm , the reader is referred to [7,33,34,35,36,37].

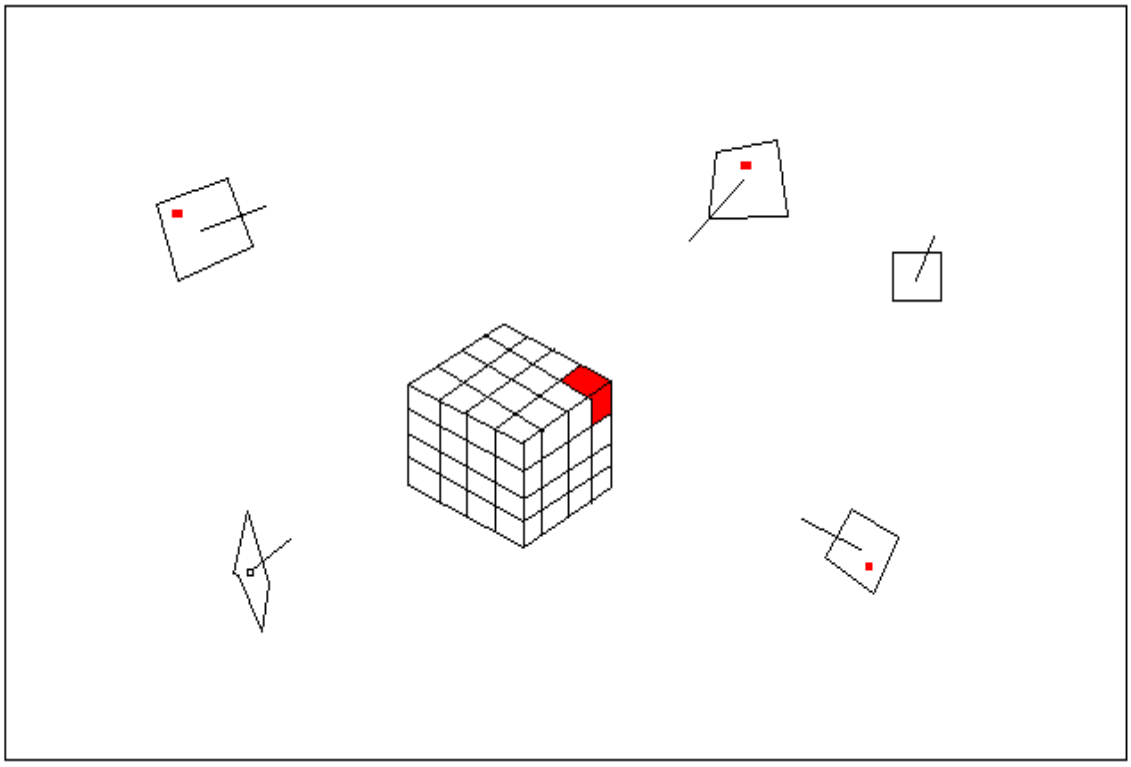

Figure 1.1: Typical setup used for voxel coloring. Multiple views observe a volume of voxels representing the potential object shape prior to the algorithm proceeds. Voxels are projected into each camera using corresponding projection matrices to determine voxel consistency and either retain or remove voxels from the scene. Note that the chosen voxel does not, project into one of the cameras(far right) and projects into an uncolored region in another camera (far left).

Recent modifications of the algorithm have focused on relaxing constraints concerning camera viewpoints [7], new consistency measures [7], and hardware acceleration techniques as well as algorithm changes to address the problem of complexity related to the traditional coloring algorithm.

Other researchers have focused on describing the initial object volume as a projective space, governed by the known relative (Epipolar) geometry between the available views. That is, the voxel space is no longer Euclidean and voxels are not uniform in size and shape. The space is directly defined by the relative geometry of the cameras. An analysis of how

camera calibration error influences volumetric reconstruction under these conditions seems important and appropriate but is the focus of future work (see Chapter 7).

Both the traditional algorithm and the more recent extensions to the technique rely on accurate camera calibration information. An in-depth analysis of camera calibration and its influence on the behavior of the object-space voxel coloring algorithm is important to characterizing the behavior of these existing approaches. In addition, results presented here have implications for new research in the area of volumetric reconstruction.

1.2. Thesis Statement and Contribution

This thesis focuses on an understanding of how camera calibration errors are related to three-dimensional, object-space, errors in a reconstructed volume. In addition, a few methods to improve camera calibration for conditions typically found in multi-view object reconstruction are suggested based on the analysis of error detailed here. The work described here is based on a specific thesis. This three-part thesis is:

> *Accuracy of volumetric reconstruction can be directly related to the magnitude and character of error in the camera models used by the reconstruction technique. An understanding of this relationship is important to the prediction and characterization of voxel coloring errors. Once this relationship has been understood, this knowledge can be used in a number of ways to reduce the errors likely to occur when reconstructing objects and environments.*

This thesis statement is expanded in the next several chapters to describe the details of how different camera parameters are related to expected reconstruction error. The depth of the thesis is demonstrated by conducting an explicit analysis of error for both the general case of multiple camera, volumetric reconstruction as well as a specific case, encountered commonly by automatic multi-view reconstruction systems.

1.3. Outline of the Thesis

The work described here was inspired, in part, by ongoing research to extend and improve the base voxel coloring algorithm. When possible, analysis of calibration effects on reconstruction accuracy is generalized to arbitrary volumetric methods. However, a large portion of the thesis is specific to methods that have been developed by particular researchers [7,33,34,35,36,37] that are in wide use, or are currently under development. Chapter 2 describes the motivation behind this work and the related research efforts that provide the impetus for this analysis.

Following that, a discussion of the issues related to calibrating multiple cameras for volumetric reconstruction is discussed in Chapter 3. Chapter 4 describes how weak topological and geometric constraints about the relative positioning of cameras can be used to improve camera calibration. The techniques suggested in this chapter are designed for common experimental setups encountered applying voxel coloring algorithms.

Chapter 5 describes how we model error in the multi-camera calibration process. Introduction of error into each of the camera models is important in understanding how overall error is propagated through all cameras involved in the reconstruction process.

Chapter 6 describes how this error models is related to error in the reconstructed object when applying volumetric reconstruction methods. The specific case of voxel coloring under a known camera noise model is discussed.

The thesis concludes in Chapter 7 with several observations related to camera errors and propagation of this error into world space. Finally, future directions for related research are suggested.

Chapter 2
Problem Motivation

The analysis carried out in this thesis is specifically directed toward voxel coloring algorithms using multiple, calibrated cameras. There are a number of examples where voxel coloring algorithms are being used to support other vision tasks [21] as well as commercial applications. The work here is directed at these general approaches but was carried out in the context of supporting rapid volumetric model acquisition for model acquisition systems that must acquire a base geometry and accurately align the supporting images to the model. In this case, accurate camera calibration is important for both the model acquisition phase and the subsequent coloring/texturing of the model during rendering.

An example commercial application of accurate voxel coloring is the Archvision, Inc, Rich Photorealistic Content (RPC) objects that can be purchased and placed into scenes for visualization, simulation, gaming, and other applications. The RPC object is a mixed geometric and image-based model that is textured according to viewpoint at rendering time. At its core, the RPC object contains a base geometry that is acquired by capturing several controlled viewpoints of a real-world object to which voxel coloring is applied.

The RPC object is created by first placing an object onto a turntable such as the one in the setup described by figure 2.1.

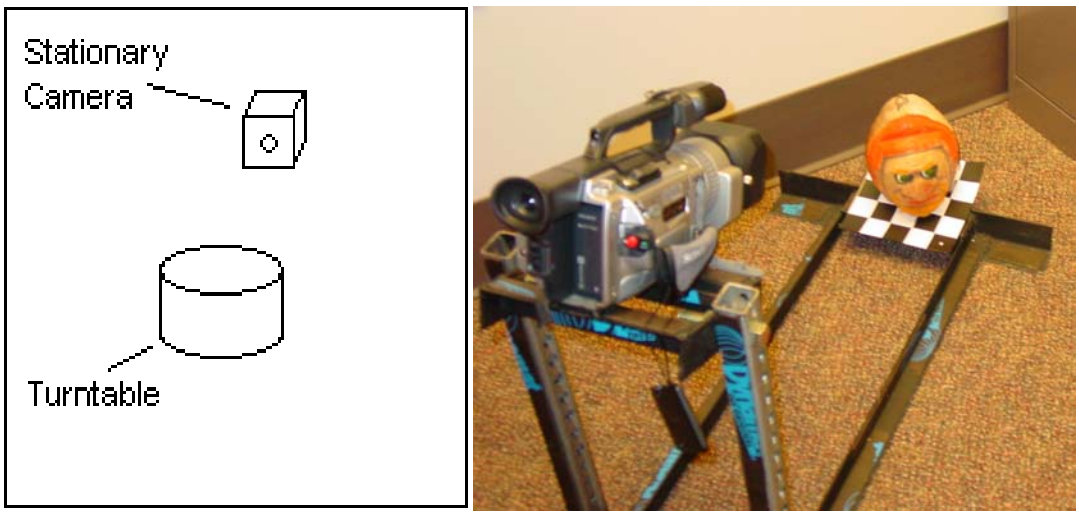

Figure 2.1: Experimental setup to capture a volumetric model from a real-world object. (a) Multiple viewpoints of turntable are calibrated to a world coordinate system. (b) Actual setup used to collect data used in this thesis.

The turntable (in Figure 2b) is printed with a calibration target that can be seen by the camera. The turntable is rotated through a predefined sequence of positions and at each, the camera is calibrated to the target using the well-known eight point algorithm [43]. Next the object is secured to the turntable and it is again rotated through the predefined set of viewpoints. The stationary camera records a large number of images, which correspond to consecutive views of the object that an observer would see were they to walk full circle around the object while looking directly at it.

In a direct image-based rendering approach, the RPC objectss can be placed into a virtual environment allowing appropriate images to be rendered based on relative viewpoint for an observer in the environment. As a result, a perfectly accurate view of the object is always available in the environment, assuming that the object is viewed from any of the same angles from which the camera that created the original images viewed the object.

Although this pure image-based approach works quite well for views taken in the same plane as the object, there are two drawbacks. If a view of the object is required from

an angle from which the camera did not record an image (say from above the object, or otherwise significantly out of the plane), significant distortion could possibly occur. Also, the number of images which must be stored can grow very large, and if a significant resolution is used, the files can become quite cumbersome, and infeasible for use in such applications as web-commerce.

2.1. Application of Voxel Coloring

Voxel coloring presents an elegant approach to extending the limitations inherent in pure image-based rendering approached. Given that a set of calibrated viewpoints of the object in question are already available, voxel coloring applied to these images is a natural approach to take.

A voxel model, which is 3-dimensional by nature, has the capacity to project correctly even if the object is viewed outside the plane from which the original images were taken. Voxel coloring is particularly suited to this application due to its focus on reprojection accuracy, which is the primary concern addressed by the RPC format.

The production of a voxel model from an RPC file is very straightforward. First, it is assumed that the object is stationary, and the camera has moved in a circle around the object. Secondly, so long as the camera model used to create a 2-D projection of the voxels into images in the virtual environments is the same as the camera model used to create the voxel model, then the true value of the intrinsic parameters of the camera are unimportant, and can be assumed to be any reasonable value. After these assumptions are made, the only task remaining is to calibrate the cameras (relative to each other) and proceed with the generalized voxel coloring algorithm.

The remainder of this thesis will deal with the problem of camera calibration. Although much work has been done in the area for both stereo and generalized voxel coloring, this particular image collection setup present new challenges and sensitivity to error which are unique. The chapter describes the specific geometric setup and how error from one pair of cameras is related to other cameras in the setup.

Chapter 3

General Calibration for N Cameras in Arbitrary Space

As mentioned in Chapter 1, an important first step in the voxel coloring algorithm is camera calibration. Camera calibration information is used to "carve" voxels by projecting each voxel into available images and measure a consistency score. In addition, calibration information (in the form of the projection matrix for each camera) influences the colors that will ultimately be assigned to the voxels. Poor calibration then, may result in an inaccurate geometric model as well a one that is far from photorealistic. In this chapter we discuss how calibration accuracy for *N* cameras, positioned somewhat arbitrarily will influence the resulting acquired object model.

Figure 3.1 illustrates the effective positions of the cameras in the world described in Chapter 2.

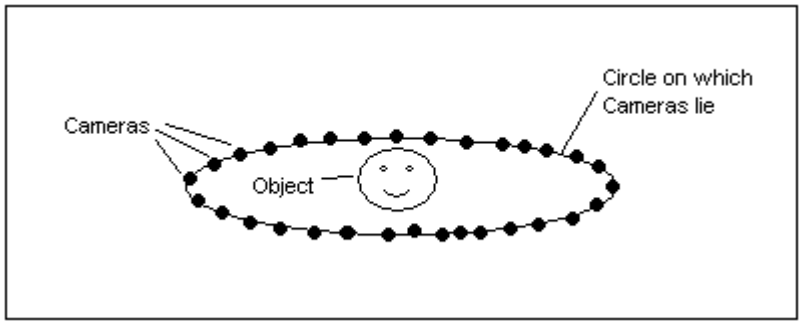

Figure 3.1: The effective positions of cameras in space. When a solid object is rotated so that a single stationary camera can photograph it multiple viewpoints, the resulting system can be modeled as multiple cameras rotated in space around the object. Note that each camera lies on a circle.

Figure 3.1 also illustrates the fact that all the cameras may lie on a particular circle in the world. So long as the turntable is leveled, this circle is defined in a plane parallel to the turntable on which the object was rotated, which passes through the camera center. The radius is defined by the distance from the camera center to the intersection of the plane and

the axis of rotation of the turntable. Note that, although we can make these observations, they are not exploited until Chapter 4. Here we focus on the more general relationship between camera position accuracy and the resulting object model.

A standard approach to calibrating a set of cameras is to discover a set of points in the world whose 3D positions are known and whose corresponding pixel positions can also be discovered. This method allows both the intrinsic and extrinsic parameters for all cameras to be found, but requires known world coordinate points. The process of finding these points is often not trivial.

As noted in Chapter 1, if the actual intrinsic parameters of the camera are not to be used to make measurements of the reconstructed object, then calibration for the set of cameras can be relative to one camera in the system rather than direct calibration to the world coordinate system. A list of match points is input to the Eight Point Algorithm [43] which has been used to calibrate pairs of cameras, but without world truth, or intrinsic camera parameters, the calibration will be correct up to a projective transform. As the same camera will be used to reproject the voxels into images, it is possible to refer to the calibration as correct up to a scale factor, as the "world truth" for this case concerns the reprojection, and not ground truth.

Although this would work well for two cameras, there are a host of cameras around the object. For this reason, it is not possible simply to calibrate each pair of cameras with one camera selected as the "world reference camera." The Eight Point Algorithm always returns a unit translation vector, not all cameras lie 1 unit away from the same camera. Even if it were possible to get world coordinate match points, these points would need to be selected on all sides of the object, as it is unlikely that any given point on the object will be visible to

all cameras. As a result of these considerations, it can be concluded that a method is needed to calibrate all of the cameras relative to each other.

It is possible to begin by assuming 3 cameras $C_1,C_2,C_3,$ and extending from there. Assume that either the intrinsic parameters for all three cameras are know, or that the intrinsic parameters for all three cameras are the same.

Given estimates of the fundamental matrices $F_1,F_2,F_3,$ for pairs of cameras C_1 to C_2, C_2 to C_3, and C_1 *to* C_3 respectively, it is possible to determine the relative (extrinsic) calibration of the cameras up to a scaling factor. From the Fundamental matrices, the true essential matrices $E_1,E_2,E_3,$ can be obtained if the intrinsic parameters of each camera are known. If the parameters are not known, but assumed to be the same, $E_1,E_2,E_3,$ can be obtained up to a projective transform. As mentioned above, in the case of voxel reconstruction from RPC files, this projective transform is simply a scaling factor. From the Essential matrices, the rotation matrices R_1,R_2,R_3, and unit vectors T_1,T_2,T_3 can be obtained in a very straight forward manner. *(reference from trucco)* Figure 3.2 depicts the relation between the cameras and the translation vectors.

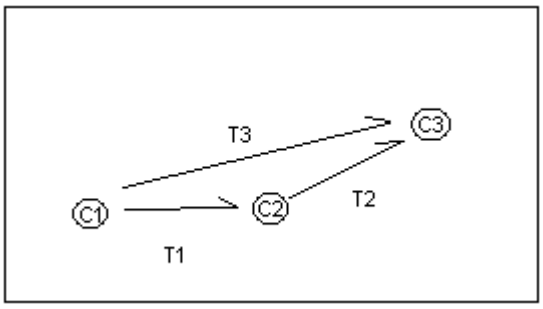

Figure 3.2: For three cameras $C_1,C_2,$ and C_3 the translation vectors from C_1 to C_2, C_2 to C_3, and C_1 *to* C_3 can be labeled $T_1,T_2,$ and T_3 respectively.

If the estimates F_1,F_2,F_3 contain no error, and R_1,R_2,R_3,T_1,T_2,T_3 are all obtained with no numerical instability. Then the following constraints will apply, and a relative

reconstruction can be obtained. In this case, the constraints involving relative rotations and translations can be written as:

$$(R_1)(R_2)=R_3 \qquad \text{(Equation 3.1)}$$

$$s_1T_1 + s_2(R_1)(T_2)= s_3T_3 \qquad \text{(Equation 3.2)}$$

where each s_i is scalar.

The rotation constraint is fully defined by the rotation matrices, and no further effort need be expended to this constraint when reconstructing the relative calibration; however, before relative calibration can be computed, each scaling factor in the translation constraint must be determined. Assuming no error or instability, the translation constraint yields a system of three equation and three unknown variables, $Ax=b$ where A is a 3x3 matrix and x and b are both 3 element vectors as follows

$$\begin{bmatrix} T_1 & T_2 & T_3 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ -s_3 \end{bmatrix} = 0 \qquad \text{(Equation 3.3)}$$

$$\begin{bmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ -s_3 \end{bmatrix} = 0 \qquad \text{(Equation 3.4)}$$

The system, has been formed by three vectors connecting three points in space, hence all three vectors are co-planar. The system is obviously rank deficient, as the third column of A is a linear combination of the first two columns. In order to obtain any one of the infinite number of solutions to the system, any one of s_i must be set. The values of the other s_i are then determined by the system uniquely. After solving for all s_i the configuration of the system is known exactly up to a scaling factor.

A system of N cameras can be considered consisting of $C_1,\ldots,C_N$. It can be assumed that fundamental matrices are found such that:

For points in pixel coordinates p_1 in C_1's image and p_2 in C_2's image, let F_1 be the fundamental matrix such that $p_2^t F_1 p_1 = 0$

For points in pixel coordinates p_n in C_n's image and p_{n+1} in C_{n+1}'s image, let $F_{2(n-1)}$ be the fundamental matrix such that $p_{n+1}^t F_{2(n-1)} p_n = 0$

For for points in pixel coordinates p_n in C_n's image and p_{n+2} in C_{n+2}'s image, let F_{2n+1} be the fundamental matrix such that $p_{n+2}^t F_{2n+2} p_n = 0$

This numbering convention is illustrated using translation vectors in Figure 3.3

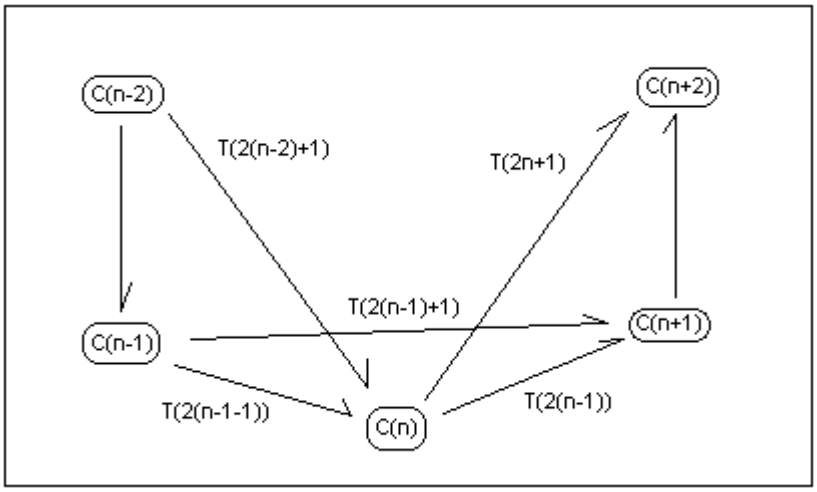

Figure 3.3: Relational numbering system. This illustrates, using translation vectors, the relationships between cameras and vectors under the numbering convention adopted in this chapter. Vectors not directly connected to the systems involving camera n have not been labeled.

Adapting this numbering convention to the Rotation constraint, it can be seen that the rotation constraint for N cameras is as follows:

$$(R_i)(R_{i+2}) = R_{i+3} \text{ for } 1 < i < 2(N-2)+1 \qquad \text{(Equation 2.5)}$$

The translation vector requires a bit more work to obtain. Given the above, the three camera process can be repeated *N-2* times with an extra step being taken after the first three cameras have been considered. Intuitively, if for the first three cameras, the value of a single s_i has been set to any arbitrary value other than zero, and if the other s_i have then been determined, the process can be repeated treating C_2 as C_1 was previously treated, and C_3 as C_2 and C_4 as C_3. At this point, rather than setting an arbitrary s_i in the system as was done with the first three cameras, the scaling factor associated with the one overlapping vector previously determined should be set to the value determined in the previous step. This must be done so that the configuration of the entire system will match. This process will yield the following system, which is a generalized translation constraint for the system of cameras described.

$$\begin{bmatrix} T_1 & T_2 & \dots & T_{2n} & \dots & T_{2((N-1)-1)} & T_3 & T_5 & \dots & T_{2n+1} & \dots & T_{2(N-3)+1} \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ \dots \\ s_{2n} \\ \dots \\ s_{2((N-1)-1)} \\ -s_3 \\ -s_5 \\ \dots \\ -s_{2n+1} \\ \dots \\ -s_{2(N-3)+1} \end{bmatrix} = 0$$

(Equation 3.6)

After solving for all s_i, the relative position between any camera in any other camera's reference frame can be calculated. For the location of C_{finish} in C_{start}'s coordinate system (where *finish*>*start*), we need only to evaluate

$$s_{2(start-1)}T_{2(start-1)} + \sum_{i=start+1}^{finish-1} s_{2(i-1)}\left(\prod_{j=start}^{i} R_{2(j-1)}\right)T_{2(i-1)} \qquad \text{(Equation 3.7)}$$

The special case occurs if C_{start} is the first camera. In that case the following equation must be used.

$$s_1T_1 + \sum_{i=start+1}^{finish-1} s_{2(i-1)}\left(\prod_{j=start}^{i} R_{2(j-1)}\right)T_{2(i-1)} \qquad \text{(Equation 3.8)}$$

The cases where *start>finish* are seen to be similar.

Note that the reconstruction relies only on the vectors relating consecutive cameras, and not those relating every other camera. As long as the translation constraint is enforced, it does not matter which set of vectors is used as both sets must return the same positions.

In summary, given only match points between pairs of adjacent cameras, and cameras separated by one other camera, it is possible acquire rotation matrices and unit translation vectors, and then attain the global calibration, up to a scaling factor.

Chapter 4
Calibration for N Cameras on Specially Defined Surfaces

As mentioned at the beginning of Chapter 3, if certain geometric constraints exist for a set of cameras, they can be used to aid in camera calibration. This chapter will focus on calibrating cameras, which lie on a surface with special, known properties.

Assume N cameras $C_1, C_2, \ldots C_N$ and assume that either the intrinsic parameters for all cameras are known, or that the intrinsic parameters for all N cameras are the same. Given that the fundamental matrices $F_1, F_2, \ldots, F_{N-1}$ are obtained for pairs C_1 to C_2, C_2 to $C_3, \ldots, C_{N-1}$ to C_N, then $R_1, \ldots, R_{N-1}, T_1, \ldots, T_{N-1}$ can also be obtained in straightforward manner. Figure 3.1 illustrates the relationship between this numbering scheme and the resulting translation vectors.

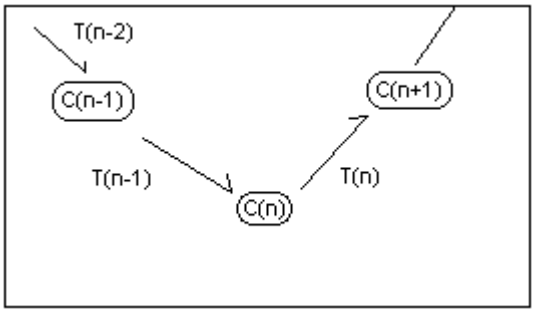

Figure 4.1: Numbering scheme illustrated with translation vectors used in this chapter.

Suppose that the cameras lie on a surface and that the optical axes of the cameras are aligned with a constant function of the gradient of the surface at the point where the camera is located, and that the rotation from each camera to another around the optical axis is zero. Further suppose that the surface, or any constrained subset (given that the constraint is known) of the surface on which the cameras are located has the following property.

For any point p and on the surface, and the function $f_{dir}(p) = \frac{\Delta f(p)}{|\Delta f(p)|}$ *then* $f_{dir}(p) = f_{dir}(p')$ if and only if $p=p'$, that is the inverse of f_{dir} is defined (by the definition of an inverse function).

Given this, if the rotation of one camera's coordinate system is known along with the constant function of the gradient expressed as a rotation R_{cfnr}, and the equation for the surface and the inverse of f_{dir} then all cameras can be placed onto the surface corresponding to their true world coordinate using only rotation information.

Algorithm:

For the camera C_i, apply $R_{i'}$ which maps the world coordinate system to the camera coordinate system, to the vector [0,0,1] (the optical axis). Further, apply the inverse of the constant function of the gradient to this vector. This vector now uniquely determines the position of C_i on the surface so that the position p is:

$$p = f_{dir}^{-1}\left(R_{cfnr}^{-1} R_{i'} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right) \qquad \text{(Equation 4.1)}$$

The position of every other camera can be determined according to the following equations. For camera j, where $j>i$, the position of i determined according to Equation 4.1, the position p is given by equation 3.2

$$p = f_{dir}^{-1}\left(R_{cfnr}^{-1} R_{i'} \prod_{k=i}^{j-1} R_k \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right) \qquad \text{(Equation 4.2)}$$

4.1. Calibration with Shift Invariance

If the surface described in the previous section also has a property of rotational shift invariance, then it is not necessary to determine the rotation of any single cameras coordinate system from the world coordinate system, as any camera can be picked and set to the "world coordinate system" camera, and placed arbitrarily.

Shift invariance implies that for any valid rotations R and any valid vectors V that Equation 4.3 must hold.

$$R_{avr2}^{-1}\left(f_{dir}^{-1}\left(R_{avr2}V_{avv1}\right)-f_{dir}^{-1}\left(R_{avr2}V_{avv2}\right)\right)=R_{avr1}^{-1}\left(f_{dir}^{-1}\left(R_{avr1}V_{avv1}\right)-f_{dir}^{-1}\left(R_{avr1}V_{avv2}\right)\right)$$

(Equation 4.3)

The shift invariance property assures that the resulting camera placement will be identical to any other camera placement, relative to all other cameras. The most common surface, and possibly the only surface with this property is the sphere.

Chapter 5
Introduction of Error into the General Calibration System

In this chapter, we discuss the direct effects of calibration error on volumetric reconstruction accuracy. In particular, the general calibration approaches introduced in Chapters 3 and 4 are re-evaluated in the context of error in the relative mapping between cameras.

5.1. Introduction and effects of Error

Assuming again three cameras, if F_1,F_2,F_3 are obtained empirically, through the eight point algorithm [43] or any other method, then they are certain to contain some noise. This noise will occur in both the rotation matrices and the translation vectors.

As a result it is very unlikely that either the rotation constraint or the translation constraint will continue to hold. Specifically, the equality in the rotation constraint will fail while the translation constraint will most probably yield a system for which there will exist only the trivial solution x=[0,0,0]. This will occur whenever the net error, causes the three vectors to be no longer co-planar. More formally, this occurs when the dot product of the error expressed as a vector, and a cross product of any two of the translation vectors is not zero.

$T_{error} \bullet \left(T_i \times T_j\right) \neq 0 \Rightarrow$ the system has only a trivial solution. (i and j not equal)

This failure of the constraint has two implications for the calibration process. Firstly, adaptations must be made to the process in order to enforce the constraints and allow for the

calibration described in Chapter 3. Secondly, the induced error can be measured, analyzed, and possibly corrected.

5.2. A note concerning assumptions used when dealing with Error

In the following sections, methods of measuring error and mitigating the effects of measured error will be discussed. It will be assumed that the errors introduced are independent, and that all errors are generated according to the same distributions. In reality, there may be any number of sources for the error in the rotation matrices and translation vectors, and it is possible that the errors may not be independent, but this case will not be discussed.

It should also be noted that if certain things about how the matrices and vectors were obtained are known, then it might be possible to efficiently determine which vectors are prone to more noise and which are not, and use this information when dealing with the noise. For example, if the vector T_1 was obtained from the eight point algorithm using 108 well chosen points, while T_2 and T_3 were each constructed using the same algorithm with only 8 well chosen points, then any measured error in the system has most likely resulted from noise in T_2 and T_3, all things being otherwise equal. For the following sections, it will be assumed that the same methods are used across all camera pairs, so that no one vector should be trusted any more than any other vector.

5.2.1 Error in the rotation matrix

If the errors in the Rotation matrices are assumed to be independent, and the distribution of the noise is assumed to be identical in each case, then the Error Rotation Matrix can be formed as follows from the Rotation Constraint.

$$R_{error} = (R_3^{-1})(R_1)(R_2) \quad \text{(Equation 5.1)}$$

To enforce this constraint, it would be desirable to spread the rotation error out across all three matrices evenly, such that equation 5.2 holds.

$$(R1_{correct})(R_1)(R2_{correct})\,(R_2) = (R3_{correct})R_3 \quad \text{(Equation 5.2)}$$

Although this is desirable, it is not intuitively obvious as to how to break the error up evenly. Although each matrix could be decomposed to its Euler angles, and the error around each axis distributed to each matrix, this process incurs the cost of dealing with singularities and ambiguities inherent with Euler angles. Although quaternion representation of the rotations would allow analysis without the difficulty of singularities, quaternion representations still do not commute, leaves a difficulty in equally distributing error back across the constituent rotations.

5.2.2 Error in Translation Vectors

As indicated above, the introduction of noise into the translation vectors creates a system for which there exists only a trivial exact solution. We choose a value for any of s_i, and attempt to solve the system exactly, no solution will be found. If, however, a least squares solution is sought, then approximations for the true relative values of s_i will be found, up to a scaling factor. This is analogous to the method suggested by Trucco and Verri for resolving ambiguities in stereo reconstruction. That is, given a baseline (or a set s_i), locate the point p which is closest and equidistant to the rays of interest. This is illustrated in Figure 5.1. Note that the point is found by constructing a perpendicular line between the rays and extracting the midpoint.

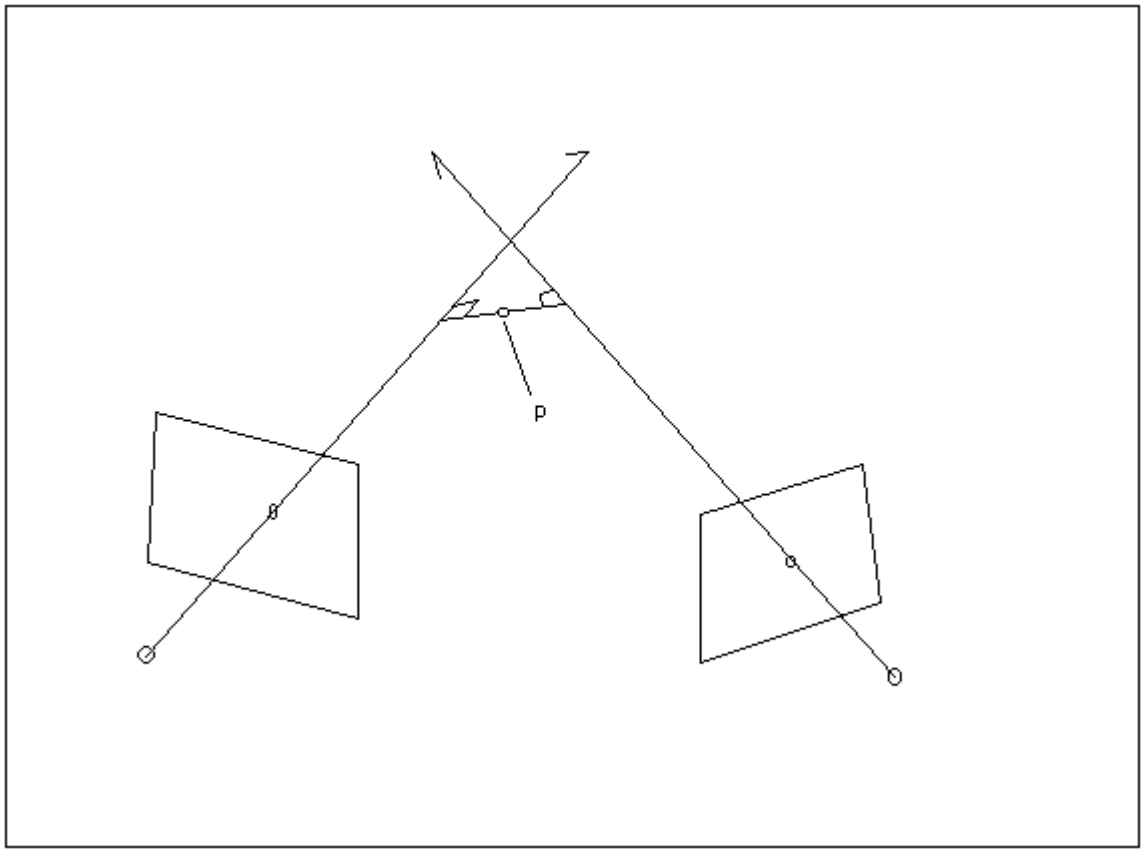

Figure 5.1: Point *p* is the point in space which projects most closely (measured in least-squares) into the indicated points in the left and right images.

Given the best approximations for all s_i, an error vector can now be formed from the Translation Constraint as follows

$$T_{error} = s_1T_1 + s_2 (R_1)(T_2) - s_3T_3 \qquad \text{(Equation 5.3)}$$

If it is assumed that the error in each *T* vector is independent, and the distribution in each case is identical, and it is further assumed that the noise in the rotation matrices will not have a significant effect on the system formed by the Translation Constraint, then it can be seen that the total error is the sum of the errors contributed by each T vector multiplied by the appropriate scaling factor s_i and that the sum is T_{error}.

Given that the error in each T_i was assumed to be independent, then the error contributed to the vector T_{error} will be proportional to the scaling factors s_i. It is possible to distribute the

T_{error} vector to each T_i , thereby forcing a system for which an exact solution. In order to equally distribute correction for the error across the system, the normalized scaling factors can be formed as follows.

$$se_1 = \frac{s_1}{s_1 + s_2 + s_3}$$
$$se_2 = \frac{s_2}{s_1 + s_2 + s_3} \qquad \text{(Equation(s) 5.4)}$$
$$se_3 = \frac{s_3}{s_1 + s_2 + s_3}$$

These can now be used to scale T_{error} for application to each of the translation vectors, to enforce the translation constraint as follows.

$$T_{error} - T_{error} = s_1T_1 + s_2(R_1)(T_2) - s_3T_3 - T_{error} \qquad \text{(Equation 5.5)}$$

$$T_{error} - T_{error} = s_1T_1 + s_2(R_1)(T_2) - s_3T_3 - se_1T_{error} - se_2T_{error} - se_3T_{error} \qquad \text{(Equation 5.6)}$$

$$0 = s_1\left(T_1 - \frac{T_{error}}{s_1 + s_2 + s_3}\right) + s_2\left((R_1)(T_2) - \frac{T_{error}}{s_1 + s_2 + s_3}\right) - s_3\left(T_3 + \frac{T_{error}}{s_1 + s_2 + s_3}\right) \qquad \text{(Equation 5.7)}$$

5.3. Error across N cameras

If N cameras are considered, then the approach to solving for the scaling factors is no different than dealing with three cameras—the first (or any arbitrary s_i)can be set, and the system can be solved minimizing the error across all cameras.

Once each s_i has been determined, it is possible to collect *N-2* error vectors. At this point, several analysis options are possible. The distribution of the raw error vectors can provide interesting information concerning the overall error. This should give a good

indication of absolute error across the system of cameras. This analysis can be performed on the *X, Y,* and *Z* elements of the vectors to test independence, and test to see if the noise is zero mean. Similarly, the normalized error can be analyzed so that cameras which have greater spacing—and a greater error as a result, will not affect the analysis.

This analysis can be useful in that, if the noise in the system is expected to be of zero mean, but a T-test or boot-strap analysis indicates that this is very unlikely, then it might be possible to take steps to mitigate the error. Specifically, if there is a bias vector which distributes noise with non-zero mean to the translation vectors, and the vector of mean error is $[eps_x, eps_y, eps_z]$, then the resultant error vectors will, all other things being equal, tend to $[eps_x, eps_y, eps_z]$. Once detected, the bias can be removed from each translation vector.

The remaining unanswered question is "what to do with the residual error vectors in an *N* camera system?" Even after bias (if it was detected) is removed, the system will still require some adjustment to enforce the translation constraint. Possible promising methods include optimization toward minimum change to all translation vectors, and also calculating the fixes for each of the sets of three cameras independently, then scaling and rotating each set of three vectors as appropriate at each camera to bring the system into alignment.

Chapter 6
Effects of Error as Applied to Voxel Coloring

As mentioned in the thesis statement, once camera calibration errors have been modeled and analyzed, this information can be used to predict and rectify errors in the voxel coloring algorithm. This chapter will discuss the sources of error in the voxel coloring algorithm and briefly discuss the associated geometry, which can give rise to a set of errors even in the presence of perfect calibration. After this discussion, the propagation of calibration error through the system will be discussed, and the effects of this error on the algorithm will be analyzed in detail, specifically as it relates to the rejection of consistent voxels.

6.1. Error Sources in Voxel Coloring

Several sources of reconstruction errors in both voxel coloring and other volumetric approaches have been studied by [25,26,27,28,29,34,45,46]. Reconstruction errors can be categorized by cause into three general areas; errors due to voxelization, errors due to calibration and mis-projection, and errors due to violations of color-constancy to included ill chosen rejection criteria for voxels.

Significant work has been done in the area of rejection criteria [7], and some work has been done in the area of voxelization error, [25,26,27,28,29,34,35,45,46]. This chaper will deal primarily with the errors introduced due to calibration errors and the resulting mis-projections. This will be preceeded by a brief discussion of errors due to voxelation, which expands upon current work.

6.1.1 Voxelation Errors and the Perfect Case

Given any set of cameras, it can be seen that there exists a perfect voxelization of the space where each voxel projects into at most one pixel in any image. This voxelization

consists of the voxels defined by the volumes defined by the intersections of the projections of the rectangular pixels into the environment. For a set of *N* cameras, a voxel can be uniquely identified by an N-tuple indicating the single pixels in each of the *N* images into which the voxel projects. Although there is a large number of possible N-tulpes, as described by Equation 6.1, a potential majority of the N-tuples are invalid, as the volumes associated with an arbitrary set of pixels from different images are unlikely to form volumes in space, which intersect.

$$\prod_{i=1}^{N} PixelCount(Camera_i) \qquad \text{(Equation 6.1)}$$

Work done by [33] makes use of voxels which are defined in projective space by a pair of basis views—for a pair of cameras, this space is perfect. For more cameras, the voxels are no longer perfect, but it can be seen that the number of total voxels cannot continue to grow increasingly faster, as the size of the voxels defined in *N* images will most probably project into a an increasingly smaller area of each additional image. This is obviously not valid when the additional cameras begin to voxelize a new section of space.

The perfect voxelation of the space contains three types of voxels, which shall be called types *I, II,* and *III*. These voxels are illustrated in figure 6.1, which depicts two single dimensional cameras which each have four pixels. The type *I* voxels are those which are adjacent to the optic center or "pinhole" of the camera. If voxel coloring is to be performed in the perfect voxel space, type *I* voxels must be ignored, otherwise, these voxels have the potential to be the first in each image to be colored, and will then occlude the entire scene, producing a reconstruction which is both trivial and perfect. Type *III* voxels are those, are unbounded on one side and not adjacent to any camera center—These are the voxels approximated by warped voxels in a semi-infinite or infinite space in [37]. Type *II* voxels

are finite, and not adjacent to the optic center of any camera. These should form the majority of voxels in a voxel space useful for voxel coloring. As far as relative abundance, there can be no more Type *I* voxels than there are pixels in all images. There are at least as many type *III* voxels as there are pixels, assuming none of the Type *I* voxels are infinite. Although it is possible to arrange cameras so that there are no type *II* voxels, this case is not easily achieved, nor particularly desirable for those applications for which voxel coloring is used.

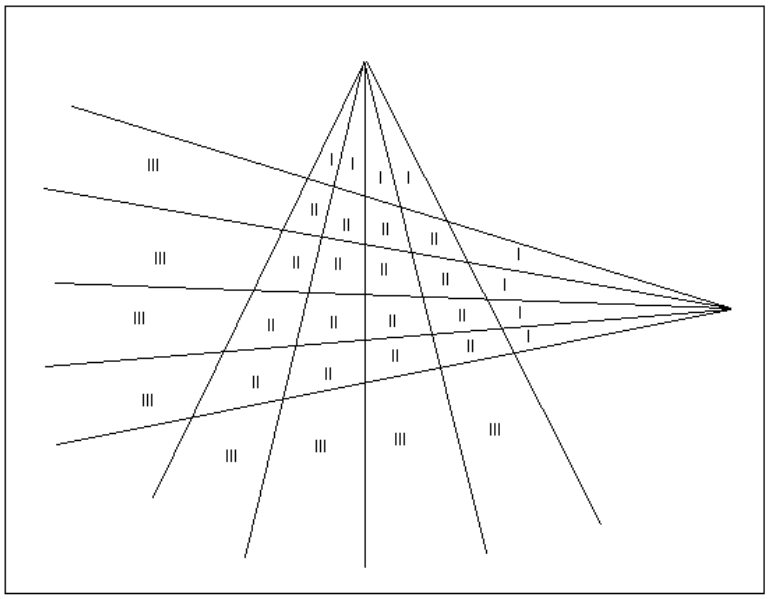

Figure 6.1: This illustrates the perfect voxelizaition of a space in 2-Dimensions defined by 2 cameras each having 4 pixels. Voxels of the first, second, and third kind are labeled.

Any voxelization of the space which utilizes voxels larger than, or which themselves could be divided by, the voxels which form the perfect voxelization, has the potential to introduce errors. This is especially true when any divided voxel contains a depth discontinuity described in [36]. Equivalently, it can be said that any voxel which projects into more than one pixel in any image has the potential to introduce error, which would otherwise not occur, even in the absence of projection errors.

To fully make use of this perfect voxel space, occlusions must be correctly handled as it is likely that a single voxel will not cover an entire pixel when it is declared consistent, or otherwise considered solid. A potential solution is to note the exact reprojection of consistent voxels into each pixel, subdivide each pixel, and allow reprojection to occur to the remaining portions. Practically, this solution is not currently feasible, nor is the use of the perfect voxel space for many cameras, but this space and associated reprojection rules define the best possible theoretical case, upon which no improvement can be achieved.

It should be noted that in the presence of error, the actual voxelization of the space can be drastically affected by any error.

6.1.2 Projection Error

Traditional voxel coloring algorithms test of the consistency of a voxel by comparing the variation over all pixels (in all cameras) into which the voxel projects to a standard variance. It is assumed that calibration error is negligible, and that variation in the pixels will be caused by lighting, or other external factors, which for consistent voxels will result in a significantly lower variation than for inconsistent voxels [35,36]. Statistically measured, this significantly lower variation is used to set the criteria for rejecting inconsistent voxels, and coloring consistent voxels.

When a single voxel is projected into a set of M images by M cameras whose collective calibration has been affected by noise, the potential for rejecting consistent voxels will increase according to several factors, which will be discussed in this section.

If the error in the Rotation and Translation vectors across all M cameras has been modeled, then the projection of a single voxel into any camera can be modeled in the presence of the noise. Normally, the projection matrices of the cameras are assumed to be

noiseless and the voxel is projected into a specific set of pixel(s) in each image. Of a given set of pixels P_i in an image into which a voxel might project in the presence of error, the pixels into which a specific voxel should project can be labeled P_i'. In the presence of error, the voxel may project into a larger area in the image as defined by the possible error. It is easily observed that the size of P_i will vary directly with both the rotational error and the relative distance of the voxel from the camera. The size of P_i will also be affected by the Translation error, but possibly to a lesser degree. As the translation error increases or decreases the distance of the voxel from the camera, it may increase the size of $P_{i,}$ directly. But this effect will vary inversely with the distance of the voxel from the camera. Translational error in the plane perpendicular to the ray passing through the center of the voxel will similarly affect size and location of P_i , but again, this effect will also vary inversely as the distance from the voxel. If it is assumed that the translation errors are small compared to the distance of the camera from the voxel, then the effect becomes nearly trivial, and the size of P_i becomes related primarily to the rotation error. Similarly, if the distance to the voxel is small as compared to the translation error, the translation error will become a key player in the size of P_i.

If the projection equations are used to project a point into a camera in the presence of noise, the previous intuitive statements can be shown mathematically.

It can be assumed that the pixel-scale factor (f) is the same in the X and Y direction. This yields projection equations from the camera coordinate system to the pixel coordinate system for a single camera

$$x_{pixel} = o_{x-imag-center} - f_{pixel-scale}\left(\frac{x_{camera}}{z_{camera}}\right) \qquad \text{(Equation 6.2)}$$

$$y_{pixel} = o_{y-imag-center} - f_{pixel-scale}\left(\frac{y_{camera}}{z_{camera}}\right) \qquad \text{(Equation 6.3)}$$

Translation from the world coordinate system to the camera coordinate system is given by the following equations.

$$\begin{bmatrix} x_{camera} \\ y_{camera} \\ z_{camera} \end{bmatrix} = R\begin{bmatrix} x_{world} \\ y_{world} \\ z_{world} \end{bmatrix} + T \qquad \text{(Equation 6.4)}$$

Equation 6.4 can be expanded to a system of three equations, which contain the terms of *R* and *T*, however this will not be done here. The terms of the rotation matrix are composed of trigonometric functions of the actual angle errors along each axis . An analysis of the error in the rotations propagated through to the image projections is non-linear in nature, and thus beyond the scope of this thesis. It will suffice to say that the magnitude of error due to the rotation matrix will increase with an increase in magnitude of the world coordinate vector—as was mentioned in the preceding paragraph.

Error in Translation and rotation can be added to Equation 6.4 to produce Equations 6.5,6.6, and 6.7, from which the previous statements concerning error propagation can be seen.

$$\begin{bmatrix} x'_{camera} + x_{error} \\ y'_{camera} + y_{error} \\ z'_{camera} + z_{error} \end{bmatrix} = RR_{error}\begin{bmatrix} x_{world} \\ y_{world} \\ z_{world} \end{bmatrix} + T + T_{error} \qquad \text{(Equation 6.5)}$$

$$x'_{pixel} = o_{x-imag-center} - f_{pixel-scale}\left(\frac{x'_{camera} + x_{error}}{z'_{camera} + z_{error}}\right) \qquad \text{(Equation 6.6)}$$

$$y_{pixel} = o_{y-imag-center} - f_{pixel-scale}\left(\frac{y'_{camera} + y_{error}}{z'_{camera} + z_{error}}\right) \qquad \text{(Equation 6.7)}$$

As inferred from experimentation in [34] the error introduced by misprojecting a voxel is dependant on where in the image the voxel is misprojected, and how the image changes in that area. For example, in the extreme best case, a set of cameras will be looking at a mono-chromatic wall—regardless of where a voxel might reproject due to noise, the color of the pixel(s) into which it will project will always be effectively the same or at least within the bounds of the properly set rejection criteria set for the algorithm. As a result, in the best case, even the worst noise will not cause a consistent voxel to be rejected. In an extreme worst case, it will be assumed that a voxel projects into exactly one pixel (the resolution of the space is identical to the actual change in the space, or as stated by Seitz, the effective Nyquist rate is achieved), and that each voxel is a substantially different color than its neighbor—this will imply that each pixel will be a substantially different color than its neighbor. In this case, as little as one pixel in projection error will result in substantial inflation of the rejection criteria statistic, and lead to false rejection of any voxels, which are mis-projected by only a small amount. In practice few images if any fall into either extreme, but individual neighborhoods of pixels will do so. Voxels projecting to the middle of large mono-chromatic areas (those likely to allow the production of cusps), will be quite insensitive to calibration noise. Voxels projecting to areas near edges, or other areas with where the autocovariance in the region has a fast fall off (or is otherwise similar to the Dirac function) will be very sensitive to noise which would allow the voxel to project across that edge, or onto a region of a different color.

This observation indicates that it might be possible to weigh the contributions of different images to the overall consistency/rejection statistic based on any projected error models and the image area into which any given voxel will project. Specifically, the expected value of

the variance of the all pixels in the image (P_i) convolved with the probability distribution of the error in that region of the image compared to the variance of the pixels into which the voxel would otherwise be determined to project (P_i') gives an estimate of reliability which could be used which determining the rejection statistic. This measure of reliability is expressed in equation 6.12.

$$E[P_i'] = \frac{1}{Count(P_i')} \sum_{P_i'} color(p_i) \qquad \text{(Equation 6.8)}$$

$$E[P_i] = \sum_{P_i} pdf(p_i) color(p_i) \qquad \text{(Equation 6.9)}$$

$$Var[P_i'] = \frac{1}{Count(P_i')} \sum_{P_i'} (color(p_i) - E[P_i'])^2 \qquad \text{(Equation 6.10)}$$

$$Var[P_i] = \sum_{P_i} pdf(p_i)(color(p_i) - E[P_i])^2 \qquad \text{(Equation 6.11)}$$

$$\frac{\sqrt{Var[P_i']}}{\sqrt{Var[P_i]}} \qquad \text{(Equation 6.12)}$$

Certainly, the calculation of this reliability coefficient would add overhead to the voxel coloring algorithm at its deepest level. As a result, it is possible that this adaptation would add an undesirable increase in run time for the algorithm, and may be infeasible at this time. However, it does provide a measure of confidence which has until now has not been considered for voxel coloring.

Chapter 7

This thesis has shown it is possible to calibrate a large number of cameras in the presence of error and has proceeded to show methods of error measurement. Chapters 3 through 5 specifically illustrate how a few constraints can allow a full relative system calibration in the presence of noise, with Chapter 5 indicating noise measurement techniques. Chapter 6 has shown how the calibration error propagates through the system directly into the image, and how this error can affect voxel coloring. Specifically, a method for determining which voxels will be sensitive to error in particular cameras and which voxels are extremely insensitive to error as a function of calibration noise and image characteristics is covered. Additionally, Chapter 6 reviews the situations where the voxelization of the space can contribute to errors, and identifies the best case voxelization of a space upon which no improvement can be made.

7.1. Future Work

Given the scope of this thesis, there are a number of interesting areas which were not able to be explored which should considered in the future. Areas such as the construction of more complex linear systems, in depth consideration of measured errors, and the detailed qualities of the perfect voxelizatoin of a space are subjects particularly well suited for future study.

Chapter 4 concludes by presenting a method of solution to the N camera calibration problem where error has been introduced and measured. At this time, it is unclear as to the best method to analyze and handle the resultant error (correcting) vectors and how to work this information back into the model of the projection system. Techniques such as weighting the confidence in a single cameras projection according to the magnitude of its error relative to the other cameras, or collectively considering the error vectors to form a Probability Distribution Function (or functions) to apply across several cameras are obvious areas for consideration. Statistical techniques such as ANOVA could also be used to analyze error vectors. There also exist any number of modifications to optimization algorithms which could make use of the error vectors along with the Rotational and Translational constraints to weight certain estimates more confidently when producing the final calibration model. Certainly, this area provides a number of possible avenues of further research.

As previously mentioned, Chapter 4 dealt with the special case where fundamental matrices were obtained for each consecutive camera pair, and each pair of cameras separated by only one camera. It may well be possible, in any given set of circumstances, to obtain additional fundamental matrices relating cameras. When obtained, these would add additional confidence, and constraints to the overall system. Enumerating the number of possibly linearly independent constraints for a set of cameras and a sufficiently large arbitrary number of fundamental matrices is one area outside the scope of this thesis which could significantly enhance the calibration of a system.

In the previous chapter (5), the perfect voxelization of a space was presented. The properties of this space, and the specific effects of error on the system are not well know,

and have not yet been discussed in depth. It is entirely possible, that use of multiple cameras in this framework could lead to resolution enhancement for a system assuming that the Nyquist rate of the cameras has not been riotously exceeded. Other benefits of dealing with the perfectly voxelized space include the possibility to deal with error for a set of cameras entirely in the Projective space. Lastly, if an efficient way to order voxels in the perfect space were developed, the use of this voxelization in voxel coloring could become feasible, and useful.

Several areas which are related, but slightly outside the scope of this thesis present significant opportunities for future work, which could contribute to both efficiency of computation and correctness of reconstruction.

Bibliography

[1] M. O. Berger and C. Chevrier and G. Simon, “Compositing Computer and Video Image Sequences: Robust Algorithms for the Reconstruction of the Camera Parameters,” *Computer Graphics Forum*, 15(3), p. C23-C32, C457--C458, September 1996.

[2] Blostein, S.D. and Huang, T.S., "Error Analysis in Stereo Determination of 3-D Point Positions," PAMI, vol. 9, no. 6, pp 752-766, Nov 1987.

[3] R. A. Brooks and A. M. Flynn and T. Marill, “Self Calibration of Motion and Stereo Vision for Mobile Robot Navigation,” *Image Understanding Workshop (Cambridge MA, April 6-8, 1988)*, pp. 398-410, Morgan Kaufmann, 1988.

[4] Carver, K.R., C. Elachi, and F.T. Ulaby, 1995, Microwave remote sensing from space, *Proc. IEEE*, 73(6): 970-996.

[5] Qian Chen and Gérard G. Medioni, “Image Synthesis From A Sparse Set of Views,” *IEEE Visualization '97*, IEEE, pp. 269-276, 1997.

[6] R.T. Collins, C. O. Jaynes, Y. Cheng, X. Wang, F. Stolle, A. Hanson, E. Riseman, "The ASCENDER System: Automated Site Modelling from Multiple Aerial Images", *Int. Journal of Computer Vision, and Image Understanding (CVIU)*, Special Issue on Building Detection and Reconstruction from Aerial Images, guest eds. R. Nevatia, A. Gruen 1998.

[7] Bruce Culbertson, Tom Malzbender, Greg Slabaugh, “Generalized Voxel Coloring,” ICCV 1999 Vision Algorithms Workshop

[8] Curless, B., Levoy, M., "A Volumetric Method for Building Complex Models from Range Images," Proc. SIGGRAPH '96 (New Orleans, LA, August 5-9, 1996). In *Computer Graphics* Proceedings, Annual Conference Series, 1996, ACM SIGGRAPH, pp. 303-312.

[9] Dobson, M.C., F.T. Ulaby, and L.E. Pierce, 1995, Land-cover classification and estimation of terrain attributes using synthetic aperture radar, *Remote Sensing Environ.*, 51(1): 199-214.

[10] F. Dornaika and R. Chung, "An Algebraic Approach to Camera Self-Calibration," *Computer Vision and Image Understanding: CVIU*, 83(3), pp. 195-215, September 2001.

[11] Elachi, C. J.B. Cimino, and M.Settle, 1986, Overview of the Shuttle imaging radar-B preliminary results, *Science*, 232: 1511-1516.

[12] Evans, Chris J., Hocken, Robert J., and W. Tyler Estler, "Self-Calibration: reversal, redundancy, error separation, and 'absolute testing,'' *CIRP Annals*, Vol 45/2, 1996, pp617-634.

[13] Evans, D.L., 1992, Current status and future developments in radar remote sensing, *ISPRS J. Photogrammetry and Remote Sensing*, 47: 79-99.

[14] O. D. Faugeras and G. Toscani, "The Calibration Problem for Stereo," *Proceedings, CVPR '86 (IEEE Computer Society Conference on Computer Vision and Pattern Recognition Miami Beach, FL, June 22--26, 1986)"*, IEEE Publ.86CH2290-5, pp. 15-20, IEEE, 1986.

[15] K. D. Gremban and C. E. Thorpe and T. Kanade, "Geometric Camera Calibration Using Systems of Linear Equations," *Image Understanding Workshop (Cambridge MA, April 6-8, 1988)*, pp. 820-825, Morgan Kaufmann, 1988.

[16] R. M. Haralick, "Determining Camera Parameters From the Perspective Projection of a Rectangle," *Pattern Recognition*, Vol. 22, pp. 223-230, 1989.

[17] R. M. Haralick and Y. H. Chu, "Solving Camera Parameters From the Perspective Projection of a Parameterized Curve," *Pattern Recognition*, Vol. 17, pp. 637-645, 1984.

[18] I. Hirschberg (ed.), "Reconnaissance, Astronomy, Remote Sensing, and Photogrammetry (Los Angeles, CA, January 19--20 1989)," Vol. 1070, SPIE, 1989.

[19] Hoepfner, K., Jaynes, C., Riseman, E., Hanson, A., and Schultz, H., "Site Reconstruction from IFSAR and EO," *Proc. ARPA Image Understanding Workshop*, New Orleans, LA, May 1997, Vol. II, pp. 983-988.

[20] C. Jaynes, A. R. Hanson, M. Marengoni, H. Schultz, F. Stolle, E. R. Riseman, Ascender II: a framework for reconstruction of scenes from aerial images, Intl. Workshop on Aerial and Spaceborne Imagery, 2001.

[21] T. Kanade, P. Rander, and P. Narayanan. "Virtualized Reality: Constructing Virtual Worlds from Real Scenes". IEEE MultiMedia, 4(1), 1997.

[22] J. Kittler and J. Illingworth and G. Matas and P. Remnagnino and K. Wong and H. Christensen and J.-O. Eklundh and G. Olofsson and M. Li, "Symbolic Scene Interpretation and Control of Perception," *VAP-Final Report*, p. DD.G.1.3-VAP Deliverable Report, March 1992.

[23] M. Kimura, H. Saito, and T. Kanade, "3D Voxel Construction based on Epipolar Geometry," *Proceedings of the 1999 International Conference on Image Processing*, Vol. 3, October, 1999, pp. 135 - 139.

[24] D. J. Kriegman, "Computing stable poses of piecewise smooth objects," *Computer Vision, Graphics, and Image Processing. Image Understanding*, 55(2), pp. 109-118, March 1992.

[25] K.N. Kutulakos and S. Seitz, "A Theory of Shape by Space Carving," *Technical Report TR692, Computer Science Dept., U. Rochester*, May, 1998.

[26] K.N. Kutulakos and S. Seitz, "What Do N Photographs Tell Us about 3D Shape?," *Technical Report TR680, Computer Science Dept., U. Rochester*, January, 1998.

[27] A. Laurentini, "How far 3D shapes can be understood from 2D silhouettes," IEEE Trans.Pattern Anal. Machine Intell.,Vol.17, No.2, February 1995.

[28] A. Laurentini, "Surface Reconstruction Accuracy for Active Volume Intersection," Pattern Reco. Letters, Vol.17, 1996.

[29] A. Laurentini "The Visual Hull Concept for Silhouette-Based Image Understanding," IEEE Trans.Pattern Anal. Machine Intell.,Vol.16, No.2, February 1994.

[30] David Liebowitz and Antonio Criminisi and Andrew Zisserman, “Creating Architectural Models from Images,” *Computer Graphics Forum (Eurographics '99)*, Vol. 18(3), pp. 39-50, The Eurographics Association and Blackwell Publishers, 1999.

[31] R. Mohan and G. Medioni and R. Nevatia, “Stereo Error Detection, Correction and Evaluation,” *First International Conference on Computer Vision (London, England, June 8--11, 1987)",* pp. 315-324, IEEE Computer Society Press, 1987.

[32] Pierre Poulin and Mathieu Ouimet and Marie Claude Frasson, “Interactively Modeling with Photogrammetry,” *Rendering Techniques '98*, Eurographics, pp. 93-104, Springer-Verlag Wien New York, 1998.

[33] H. Saito and T. Kanade, "Shape Reconstruction in Projective Grid Space from Large Number of Images," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '99)*, June, 1999.

[34] Steven M. Seitz, “Image-Based Transformation of Viewpoint and Scene Appearance,” University of Wisconsin-Madison, 1997.

[35] S. Seitz and C. Dyer, "Photorealistic Scene Reconstruction by Voxel Coloring," *International Journal of Computer Vision*, Vol. 25, No. 3, November, 1999.

[36] Greg Slabaugh, Bruce Culbertson, Tom Malzbender, Ron Schafer, “A Survey of Methods for Volumetric Scene Reconstruction” International Workshop on Volume Graphics 2001

[37] Greg Slabaugh, Tom Malzbender, Bruce Culbertson, “Volumetric Warping for Voxel Coloring on an Infinite Domain,” ECCV 2000 SMILE Workshop

[38] Mark R. Stevens, “Evaluating 2-D Image Comparison Metrics for 3-D Scene Interpretation,” *Computer Vision and Image Understanding: CVIU*, 84(1), pp. 179-197, October 2001.

[39] T. M. Strat, “Recovering the Camera Parameters From a Transformation Matrix,” *Image Understanding Workshop (New Orleans, LA October 3-4, 1984)",*, pp. 264-271, Science Applications International Corp., 1984.

[40] K. Sugimoto and M. Takahashi and F. Tomita, “Scene Interpretation Based on Boundary Representations of Stereo Images,” *Ninth International Conference on Pattern Recognition (Rome, Italy, November 14--17, 1988)*, pp. 155-159, Computer Society Press, 1988.

[41] R. Szeliski. Stereo algorithms and representations for image-based rendering. In *British Machine Vision Conference (BMVC'99)*, volume 2, pages 314-328, Nottingham, England, September 1999.

[42] H. Takahashi and F. Tomita, “Self-Calibration of Stereo Cameras,” *Second International Conference on Computer Vision (Tampa,, FL, December 5--8, 1988)*, pp. 123-128, Computer Society Press, 1988.

[43] Emanuele Trucco, Alessandro Verri, “Introductory Techniques for 3-D Computer Vision,” Prentice Hall, Upper Saddle River, New Jersey, 1998.

[44] G. Q. Wei and Z. Y. He, “Determining Vanishing Point and Camera Parameter: New Approaches,” *Ninth International Conference on Pattern Recognition (Rome, Italy, November 14--17, 1988)*, pp. 450-452, Computer Society Press, 1988.

[45] Kuzu Yasemin, Volker Rodehorst, ”Volumetric Modeling Using Shape from Silhouette,”
In: O. Altan & L. Gründig (eds.) *Fourth Turkish-German Joint Geodetic Days*. Paper presented to the conference organized at Berlin, Germany, April 2-6, Vol. I, pp. 469-476.

[46] Anthony Yezzi, Greg Slabaugh, Adrian Broadhurst, Roberto Cipolla, Ron Schafer, **“**A Surface Evolution Approach to Probabilistic Space Carving,” The 1st International Symposium on 3DProcessing, Visualization, and Transmission (3DPVT) 2002.

Vita

Elwood Talmadge Waddell, Jr.

Date of Birth: 29 December 1977

Place of Birth: Huntington, West Virginia.

Education

2001, Aerospace Basic Course, Maxwell Air Force Base, Al.

2000, BS Mathematics, *summa cum laud*, University of Kentucky, Lexington, Ky.

1999, BS Computer Science, *summa cum laud*, University of Kentucky, Lexington, Ky.

Honors

2000, Distinguished Graduate, Air Force Reserve Officers Training Corps

2000, Mathematics Departmental Honors

Omicron Delta Kappa

Tau Beta Pi

Positions Held

2001, Second Lieutenant, United States Air Force, Scientific Analyst,

Electronic Warfare Analyst, 36th Electronic Warfare Squadron,

Eglin Air Force Base, Fort Walton Beach, Fl.

Elwood Talmadge Waddell, Jr.
23 July 2002