Automation and Calibration for Robot Vision Systems

Steven A. Shafer
May 1988
CMU-CS-88-147(3)

This research was sponsored by the National Science Foundation under Grant DCR-8419990 and by the Jet Propulsion Laboratory, California Institute of Technology, sponsored by the National Aeronautics and Space Administration under Contract 957989. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author and do not necessarily reflect the views of the National Science Foundation, the Jet Propulsion Laboratory, or the United States Government.

Abstract

General-purpose robot vision includes a number of different tasks that impose a great variety of imaging conditions and requirements. To support the full range of these tasks, an imaging system must provide a very wide dynamic range and high precision in both geometric and radiometric characteristics. In general, this can only be accomplished by a highly precise, automated imaging system. This paper defines a twelve-parameter model for a robot imaging system -- six parameters in camera position, three in optical constraints, and three in sensitivity -- that subsumes common TV cameras and "scientific" cameras as special cases. We call this model the "Imaging Space", a configuration space for robot imaging systems. Systematic consideration of this complete model leads to a more comprehensive treatment of camera calibration than has been seen before. While traditional calibration literature refers only to geometric calibration of the imaging system, the new model uses similar concepts to outline the *radiometric* (pixel value) calibration of the system. The concept of "second-order calibration" is also introduced, in which the interaction of geometry and radiometry is explicitly accounted for. Representing this second-order calibration data in a useful form is not yet a solved problem. We also outline some of the issues in specifying imaging constraints in a task-oriented way. This paper is not primarily a report of research results; instead, it is a report on the state of the art in imaging system technology and calibration and an outline of some future directions for work in this area. The emphasis throughout is on the achievement of wide dynamic range, i.e. high precision in geometry and radiometry, because modern theories for robot vision are showing a direct link between the precision of the imaging system and the precision in computed quantities such as object shape.

Table of Contents

1. Imaging Requirements for Robot Vision 1
2. The Imaging Space: A Camera Model for Automated Imaging 1
   2.1. Camera Position Parameters 1
   2.2. Rationale for Automation 1
   2.3. Lens Parameters 1
   2.4. Other Parameters in the Imaging System 2
   2.5. Summary of the Imaging Space Model 2
3. Calibration and the Imaging Model for the Imaging Space
   3.1. First-Order Geometric Calibration
   3.2. First-Order Radiometric Calibration
   3.3. Second-Order Integrated Calibration
4. Higher-Level Issues in Imaging System Control
   4.1. Acknowledgements

List of Figures

Figure 2-1: Imaging Coordinate System
Figure 2-2: Two Imaging Systems
Figure 2-3: Focus Distance Components
Figure 2-4: Adjustment of Focus
Figure 2-5: Overall Imaging System Block Diagram
Figure 3-1: Pinhole Model of Imaging Geometry 1
Figure 3-2: Equipment for Geometric Calibration 1
Figure 3-3: Quantization and Repeatability Confidence 1
Figure 3-4: Pinhole Model v. Finite Aperture Model 1
Figure 3-5: Video and Scientific Cameras 1
Figure 3-6: Polarization of Incoming Light at the Sensor Plane 1
Figure 3-7: Blooming Caused by Bright Highlights 2

1. Imaging Requirements for Robot Vision

Robots are being called upon to perform increasingly ambitious tasks including higher-speed outdoor navigation, more reliable manipulation and inspection, and responding to more "natural" environments with more variation and less structure. To achieve these goals, the demands upon robot vision systems are increasing dramatically. Yet, relatively little attention has been paid to the quality of the imaging equipment and the attendant calibration. Quite the contrary, the proliferation of inexpensive TV-grade equipment has led to an explosion in the use of this technology, regardless of whether it is well-suited to the vision task being undertaken.

Modern theories for computer vision have begun to establish a direct relationship between the precision of the imaging system and the precision of computed quantities such as line position and surface shape [3, 5, 19, 20]. As more quantitative and precise vision systems are demanded, there is an increasing need for a theory of camera modeling that can capture the essential variables in the selection and description of imaging equipment. Also needed is a corresponding calibration methodology to achieve the full potential of the hardware for high-precision image acquisition. In this paper, we present an imaging model and calibration definition that help to set forth the dimensions of interest in high-precision robot imaging systems. This discussion differs from previous literature in this area in its inclusion of both geometric and radiometric factors in a single model, and in its orientation towards *automated imaging*, in which the robot can control all possible imaging parameters.

To begin, let us consider some vision tasks that define the limits of the precision requirements for robot vision systems. We will define the *dynamic range* of a system to be the number of reliably resolvable units in the total sensitivity range of the system. For the geometric domain, dynamic range means the number of pixels in the field of view; for the radiometric domain, dynamic range refers to the signal-to-noise ratio in the brightest pixel value of the image. We use the term *precision* to refer to such a relative property of the imaging system. In contrast, we use the term *accuracy* to refer to the absolute correspondence between the imaging system's scale of measure and the actual physical coordinates in the external world.

To determine some bounds on the geometric dynamic range we might desire for general-purpose imaging, we need an upper bound estimate on a "reasonable" field of view and a lower bound on the pixel resolution requirement. For navigation tasks indoors or outdoors, experiments have shown the need for at least a 60° field of view [6]. For surface characterization, little research has been reported; however, a pixel size of about 0.1 mm at a distance of 1 meter is a reasonable estimate for distinguishing smooth from rough surfaces. This corresponds to a pixel size of 0.0001 radian, or about 1/150°. Combining these values shows a potential need for a geometric dynamic range of 10,000:1, i.e. 10,000 pixels across the image! In the radiometric domain, high-precision quantitative analysis is becoming increasingly important [8, 15, 27]. For this research and its future applications, we will likely require at least a 200:1 pixel value resolution, which is nevertheless less than the full range of an 8-bit pixel. This must be augmented by considering the variation in the potential illumination level of the image, determined by the brightness of the scene as a whole, as well as the need to look into bright highlights or dark shadows within the scene. Assuming conservatively that the overall illumination level may vary by 50:1 [10], that shadows can easily be 10 times darker than the overall scene, and that highlights can be at least 10 times brighter, a total radiometric dynamic range of at least 1,000,000:1 is desirable.

These figures stand in sharp contrast to the capabilities of TV-grade equipment commonly used in

robot vision, which can typically provide perhaps a 500:1 geometric dynamic range and a 100:1 radiometric dynamic range. A higher dynamic range can be achieved generally by three means:

- *Direct Precision Improvement:* Certain technologies are capable of providing wider dynamic ranges than TV equipment. For example, imaging array sensors may soon provide as much as a 2000x2000 image; cooled CCDs can provide on the order of a 10,000:1 radiometric response range. Of course, these figures are still far below the desirable ranges.

- *Nonlinearity:* If a nonlinear scale is used such as a logarithmic scale, the dynamic range can be increased substantially at a cost of resolution. For radiometric response, this corresponds to the common nonlinear intensity response function (i.e. "gamma" response); in geometric terms, nonlinearity would describe, for example, a sensor with larger pixels at the periphery. The tradeoff is that the resolution will be nonuniform, for example proportional to the absolute level, so that high levels (or peripheral pixels) would be measured with poor resolution. A variation on this is the use of multiple cameras with different parameters, e.g. one with its aperture set for normal illumination, and another with a wide aperture used for looking into shadows [12]. The result is a nonlinear resolution scale.

- *Automation:* In view of the modest successes achieved with TV-grade equipment, it should not be surprising to note that many specific tasks do not require such extensive dynamic ranges. In fact, this is how most robot vision is performed: the imaging equipment for a specific task is dedicated to that task, and hand-tuned in all the various parameters such as camera position, lens focal length, aperture, etc. The result is hardware that is not capable of any substantial generality in the variety of tasks it can undertake. Through such parameter settings, the dynamic range can be manually extended by a considerable amount -- easily 10+:1 geometric scaling by a zoom lens, and 1,000:1 radiometric range extension by controlling the aperture, exposure time, and digitization gain. These parameters can all be automated by straightforward means. Thus, while current technology cannot provide all the desired dynamic range *at one time*, the available instantaneous range can be adjusted over a much wider range by a robot equipped with appropriate automation. This is one of the key reasons for utilizing automation in a robot imaging system.

This approach -- beginning with a task description in general terms and deriving the imaging precision needs -- is a kind of "top-down" approach to imaging system configuration. It stands in contrast to the traditional "bottom-up" approach that dominates the field, in which the researcher begins by selecting a technology (usually Closed-Circuit TV technology) that has inherent limitations, and then attempts to determine what tasks can be accomplished with that equipment. While the bottom-up approach is a good paradigm for initial exploration of the vision problem, only the top-down approach can provide the engineering control needed for high-performance robot vision systems.

Measurement quality can be described in both relative and absolute terms. Relative measurement quality is best described by the number of resolvable units in the range of the instrument; this may be referred to by any of the terms *precision*, *dynamic range*, or *SNR (signal-to-noise ratio)*. In a continuous measurement, these terms refer to the ratio of the total measurement range to the standard deviation of the signal (noise); for a discrete measurement, precision and dynamic range refer to the number of steps in the range. The noise level or step size itself may also be called *resolution*. In both the continuous and discrete cases, these terms refer to the repeatability of measurements and therefore characterize the *uncertainty* of the measured value. On the other hand, the term *accuracy* refers to the possible misalignment between the internal measurement scale of the instrument and the external world. This misalignment creates a very systematic and repeatable *error* in the measurements. The error itself may be known to be bounded by a *tolerance*, or it may be continuous and be characterized by its standard deviation.

2. The Imaging Space: A Camera Model for Automated Imaging

In this section, we will briefly develop the "Imaging Space" camera model to describe all of the generally achievable degrees of freedom in automating a camera for robot vision. As we will see, these 12 degrees of freedom can describe a wide variety of systems that includes both closed-circuit TV equipment and so-called "scientific" imaging systems. This set of parameters defines the Imaging Space, a configuration space for robot imaging systems. We also discuss the ramifications of automation of all of these parameters, using a wide variety of vision tasks to illustrate the precision requirements for automation. In the later portions of the paper, we will address the problem of calibrating the imaging system.

2.1. Camera Position Parameters

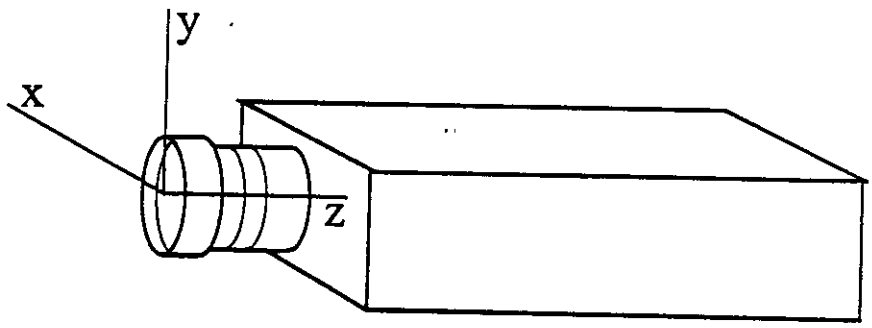

Figure 2-1: Imaging Coordinate System

The most commonly controlled and studied parameters of the imaging system are the six degrees of freedom in the camera position and orientation, which we will call T_x, T_y, T_z, θ_x, θ_y, and θ_z. As shown in Figure 2-1, we will assign *x* to be horizontal in the image plane and *y* vertical; *z* will be the optical axis of the camera. The rotational parameters will be assumed to be rotations about the respective axes (θ_x = "tilt", θ_y = "pan", and θ_z = "roll"), but we won't go into detail on their interpretation since that is beyond the scope of this paper. The origin of the coordinate system is the front principal point of the lens, which acts as the scene perspective center for geometric optics calculations.

Many technologies are commonly available for automation of these parameters, including particularly rotators and translators based on analog, stepping, or microstepping motors, and also mobile robot platforms that provide control over T_x, T_z, and θ_y (or some rotation of these if the optical axis is tilted or rolled out of the standard, horizontal position). It is useful, however, to consider the impact of limited precision in each parameter. Let us hypothesize for a moment an outdoor vehicle using a stereo vision system for object detection and localization, as shown in Figure 2-2(a). It has two cameras on a 2-meter baseline, each having 512 pixel columns spanning a 50° field of view. In this system, the average pixel subtends approximately a 0.1° angle. If the vision software produces accurate matches to the nearest pixel location, an object at a distance of 10m (meters) can be localized to within about 18cm, approximately 7 inches, in *z*. If the rotational precision of the camera alignment is similar, the localization tolerance increases to about 36cm, 14 inches, giving the overall system greater than a 3% relative error. This would be considered rather a crude 3D vision system. If the baseline distance were in error by 0.1 inch, about 0.25cm, which would be a fairly coarse error, the increase in the measurement error would be

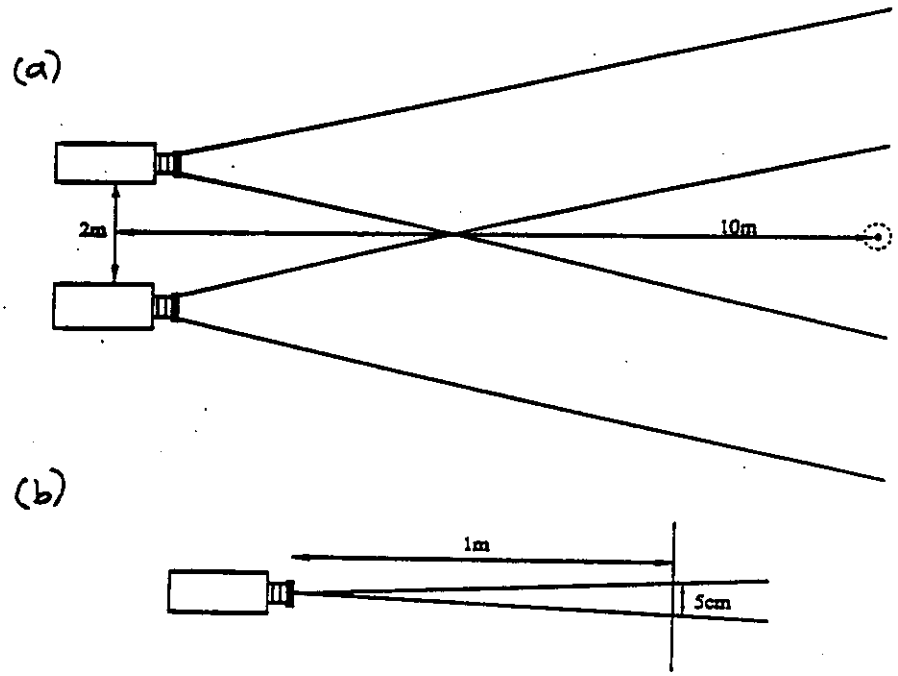

Figure 2-2: Two Imaging Systems

only about 2.5cm. Thus, the translational precision at this magnitude is far less critical in this system than the rotational precision.

In contrast, consider a system for inspection of fine surface details, as shown in Figure 2-2(b), with 10 pixels per millimeter at a distance of 1 meter. In this system, each pixel subtends about 0.006°, so an angular tolerance of 0.1° corresponds to a 15-pixel error. Similarly, a 0.1 inch translational error represents a 25-pixel shift. In this case, increasing the translational precision is more critical. These two example systems illustrate the basic principle of precision analysis for imaging system design: the precision of the control parameters cannot be evaluated in isolation -- it must be evaluated in terms of its effect on the resulting images.

Automation for these degrees of freedom can readily achieve resolution that is a factor of greater than 10 improvement over the above examples. Commonly available technology for microstepping motors and translation stages is advertised as providing repeatability on the order of 0.001 inches (0.025 mm), or with specially graded components, 0.00005 inches (0.00125 mm); standard rotation stages typically claim repeatability of 0.5 arc-minutes (0.008°), and precision grade stages 0.2 arc-minutes (0.003°). Arrangements of such stages usually involve "stacking" them together into a jig of some sort, with appropriate rigid support structures, to provide an automated camera platform; the motors are controlled by specialized boards in the host computer or by a control box with an RS-232 or IEEE-488 connection to the computer.

It is extremely difficult to design such a jig to reliably place the camera's optical center on all the rotational axes; thus, each rotation of the stages will generally induce some translation in the camera's coordinate system as well. To obtain a specific motion then requires that a 6-DOF (degree of freedom)

inverse kinematics problem must be solved in order to move the camera to the desired position and orientation. Even if the jig is designed to align the rotation axes perfectly, some special 6-DOF manually-adjustable assembly is usually required to couple the camera to the jig so that it can be aligned in the desired manner. Some alternative technologies to these jigs include robot arms or robot vehicles used as camera platforms. Usually, those devices provide substantially lower precision than the figures given here, but of course they are useful for many robotics tasks, especially if some on-line calibration procedure allows compensation for the low repeatability in position control. Also, in some cases, the vehicle or arm itself defines the desired frame of reference for calibration, so that the precision of its motion is not an issue for the vision system.

2.2. Rationale for Automation

Even though we have not yet presented the full 12-DOF model for automated imaging, the above discussion is enough to see the gross outlines of the reasons for utilizing automation in the imaging process:

- *Repeatability:* Automation can typically provide finer resolution of control than manual adjustments. In the cases above, very precise manual adjustments can be utilized, typically over very small ranges of travel. However, the knob itself needs to be quite specialized to allow human fingers to manipulate it. Similarly, manual lens focusing and aperture control are typically very crude by comparison with automated lens controls.
- *Separation of Parameters:* Frequently, a real imaging system will not provide for independent adjustment of parameters. For example, as described above, rotating a camera platform may also induce a translation in the camera's coordinate system. By automating and calibrating the imaging mechanism, the system can be designed to automatically compensate for such couplings, in effect providing a system with independently controllable parameters even though their mechanisms may be physically coupled. This is analogous to the distinction in robotic manipulation between joint control and Cartesian control. The same argument applies to very pragmatic issues such as automated v. manual filter-changing: changing filters manually makes the camera vulnerable to being bumped or pushed into a different position or orientation. Another important case is that of focusing the lens, which generally moves the perspective center of the imaging system along the optical axis as described later in this paper. Compensation for this change may involve adjusting the focal length and position of the camera. In all these cases, automation can be used to decouple parameters that are physically connected but conceptually independent.
- *Knowledge of State:* For use of calibration data, the imaging system must generally know the values of all the key imaging parameters that affect the calibration. With a manual system, the system may not know where it is, what the focal length is, etc. Such ignorance leads to calibration data that assumes specific values for all the imaging parameters, and the resulting imaging systems are only useful for a single, dedicated task.
- *Image Quality Control:* To maintain the best image quality, the imaging system may need to make minute adjustments in all parameters. For example, when the camera is moved, it may now be imaging some shadows that require that the intensity sensitivity range be adjusted downward. When a new object is to be imaged, the lens may need to be re-focused. Maintaining high image quality requires that the system adapt to the changing demands imposed by the task and environment, which is only possible if the imaging process is highly automated. Achieving a wide dynamic range, as described in the introduction of this paper, is an example of image quality control.

Hopefully, the above discussion will serve to clarify the immense importance of automated imaging for high precision and flexibility in the imaging process. It does not address the issues of real-time control of the imaging system or of saving labor in repetitive reconfiguration of the imaging system; while those

might sometimes be additional reasons for automation, they do not bear directly on precision and flexibility issues and are hence outside the scope of this paper.

2.3. Lens Parameters

In a fully automated imaging system, the lens and camera also provide several degrees of freedom. While the camera position and orientation are strictly geometric features, the lens and camera parameters will affect both the geometric and radiometric aspects of imaging. First we will examine the lens, which generally provides three degrees of freedom. The *focal length*, F, is an optical magnification parameter (also called "zoom"). Zooming trades off the width of the field of view for the resolution of the individual pixels; zooming is thus the primary means for controlling the geometric sensitivity range. A modern lens generally contains as many as a dozen glass elements arranged in three to eight groups; zooming is implemented by elaborate movement and counter-movement of these groups to achieve the desired optical effect. A convenient measure of lens performance is the ratio of the maximum to the minimum focal length, expressed as "x3" or "x10", etc., sometimes called the "zoom factor" of the lens. Due to the complex optics of focal length adjustment, a high zoom factor is very difficult to achieve. Typically, larger format lenses such as 35mm camera lenses have zoom factors of x3 to x6; smaller format lenses such as C-mount video lenses, which provide more opportunity for optimization of optical parameters, may achieve x10 or even x20.

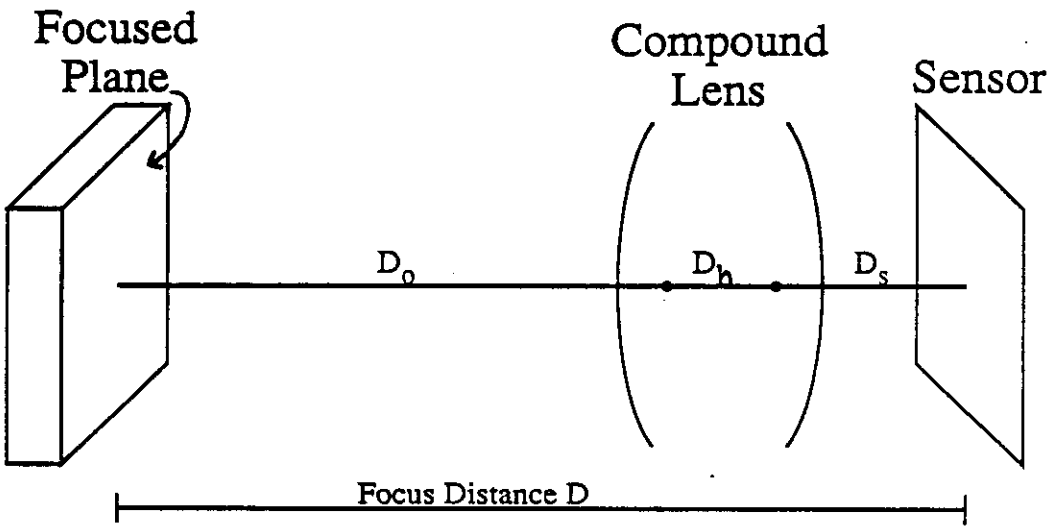

Figure 2-3: Focus Distance Components

The second geometric parameter of a lens system is the *focus distance*, D, by which we refer to the distance from the sensor plane (chip, tube, film, etc.) to the plane in the scene that will be focused on the sensor. This distance is composed of three parts as shown in Figure 2-3: the distance D_s from the sensor plane to the rear principal point of the lens, the length D_h of the *hiatus* between the principal points, and the distance D_o from the front principal point to the focused plane in the scene, according to $D = D_s + D_h + D_o$ [14]. The two components D_s and D_o form a "conjugate pair" of distances according to the lens law $1/F = 1/D_s + 1/D_o$. The focus distance is physically controlled by moving optics within the lens to move the locations of the two principal points, thus changing the values of all the components of D according to the constraints of the above laws. As can be seen in the lens law, if $D_s = F$, then D is infinite and the lens

is "focused at infinity". By increasing D_s, the focus distance D can be reduced down to some minimum value that is an attribute of the lens. Special "close-up" optics can usually be fitted to the front of a lens to modify F, thus allowing the minimum focus distance to be reduced yet further.

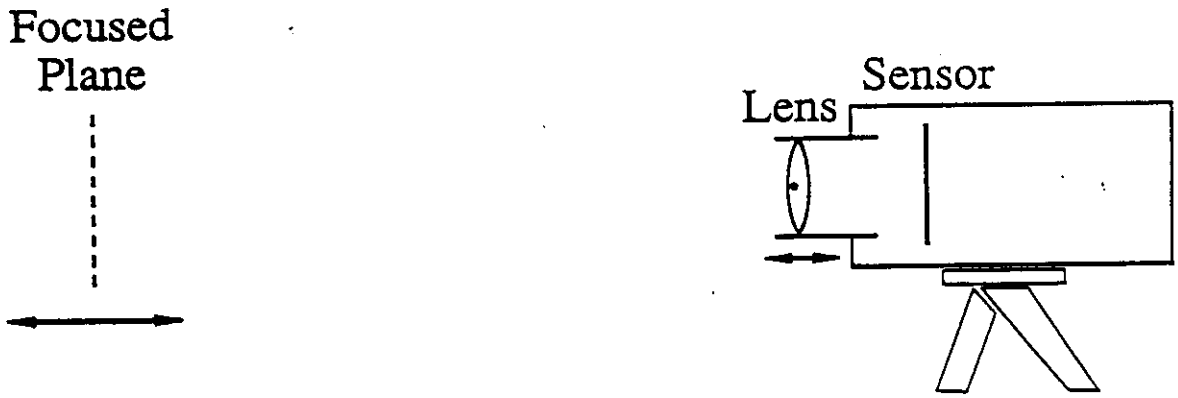

Figure 2-4: Adjustment of Focus

This parameterization describes the typical imaging arrangement of Figure 2-4, in which the camera body containing the sensor is held rigidly while the lens is focused. Adjusting the focus can then be seen to modify the lens component positions, which in turn moves the principal points of the imaging system forward and backward. Since the front principal point is generally taken as the origin of the scene coordinate system, there is an intimate coupling between D and T_z. Similarly, focusing by changing the lens-to-sensor distance alters the angle subtended by the sensor as seen from the optical center, thus modifying the magnification (angular field of view and pixel resolution) of the system. To maintain a fixed focus plane in the scene, the lens-to-sensor distance must also be modified with F. This creates an awkward inter-relationship between F and T_z, with the result that the focus distance must be an explicit parameter in the kinematic computation of the camera position. Such complexity is an inescapable outcome of the desire to decouple control of focal length from control of the sensor-to-object focus distance so that each one can be adjusted without disturbing the other, an arrangement of nearly all modern "equifocal" zoom lenses.

Finally, the lens *aperture*, A, is generally adjustable. Widening the aperture increases the intensity sensivity of the imaging system but reduces the "depth-of-field", the interval about the focus distance within which objects will be satisfactorily focused on the sensor plane. Aperture is measured in inverse ratio to the focal length; this is called the *f-number*. Thus, an aperture of f/4 refers to an aperture diameter A that is one-fourth the focal length of the lens. Aperture adjustment for a lens commonly ranges from about f/3.5 to f/16, with many lenses providing a wider range of adjustment down as far as f/1.2 or up to f/32. A manual lens is generally adjustable only in large discrete steps, i.e. f/2, f/3.5, f/4, f/5.6, f/8, f/11, f/16, f/22. These steps are measured logarithmically to the base 2, with a difference of one log unit (factor of 2) called one *f-stop*. Automation can provide a virtually continuous adjustment of aperture.

For lenses, perhaps more than for the positioning equipment or the camera itself, manual technology is by far the most common for robot vision. Yet, of all the elements, the lens adjustments are typically the

most crudely adjustable and the least repeatable. This presents a serious problem for high-precision computer imaging. Lens price, reflecting the optical and mechanical quality and the degree of automation, tends to vary over about three orders of magnitude (from $50 to $50,000), thus placing a very strong pressure on researchers to use inexpensive lens technology. Unfortunately, such lenses tend to be manual, with poor optical quality (introducing distortions) and poor mechanical quality (reducing repeatability). For example, in common C-mount lenses, the perspective center in the image may "wander" by several pixels as the lens is zoomed; 35mm SLR lenses are better but still provide for poor repeatability of the lens parameter settings.

Automation is also quite expensive, typically multiplying the cost of the lens by an additional factor of 10 over a manual lens with similar performance. Inexpensive automation is sometimes achieved by starting with an inexpensive joystick-controlled lens such as a VCR-grade lens, chopping off the controls, and inserting D/A converters to provide computer control. Typical repeatability for such lenses is poor, and the optical quality is similar to the low-end manual lenses that they resemble. By using a high-quality lens and fitting it with microstepping motors, a precision of about 2,000 steps for each parameter (*F*, *D*, and *A*) can be achieved, with substantially lower distortions than the common lenses described above.

2.4. Other Parameters in the Imaging System

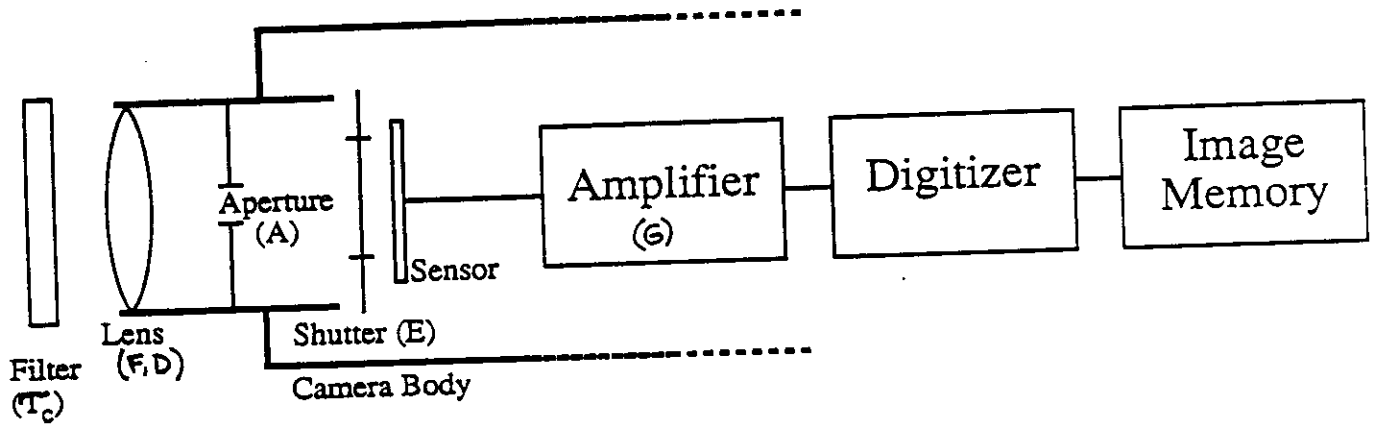

Figure 2-5: Overall Imaging System Block Diagram

In addition to the 6 DOF in positioning and 3 DOF in the lens, the imaging system may provide for several additional adjustable parameters from the components shown in Figure 2-5. First of these is the *exposure time*, *E*, which determines how long the sensor will be exposed to light from the scene. A longer exposure will typically increase the signal-to-noise ratio of the pixel values but will of course slow down the imaging time. Long exposures also allow the use of smaller apertures, thus increasing the depth-of-field in the image. For static imaging scenarios, a long exposure time can thus be very advantageous. Ironically, this is the parameter that is least frequently adjustable in typical robot vision equipment, since nearly all robot vision uses TV-grade video cameras with a fixed 1/30 second rate for exposing and scanning the image. As an alternative to video equipment, there are "scientific cameras" that typically allow controllable exposure time ranging from about 1/5 second up. In addition, the new generation of "fast-frame" CCD cameras may allow adjustable exposure in the range of about 1/10 second to 1/500 sec.

For even greater increase of the SNR, scientific cameras using CCDs sometimes provide for cooling the chip to -40°C or below, using pressurized nitrogen gas to keep moisture from condensing on the sensor chip. Systems with these arrangements advertise SNRs as high as 10,000:1, and are frequently used in astronomical observatories. For comparison, typical TV-grade equipment may provide a SNR of 100:1, and a high-precision vidicon or an uncooled scientific camera may provide 500:1.

The camera head generally amplifies the signal from the sensor so that it will not be swamped by noise while being transmitted and digitized. The gain of this amplifier, G, is a factor that can sometimes be externally controlled. A high gain increases the intensity sensitivity of the system, allowing the use of a narrower aperture and shorter exposure time; however, the noise in the pixel values is generally proportional to the gain. Thus, a high gain does not increase the radiometric precision of the system. In many TV-grade cameras, an "automatic gain control" circuit adjusts G automatically so that the average signal from the camera is maintained at a constant level. Such a design is totally unsuitable for radiometric calibration, since the gain is being adjusted without control by the software, and there is usually no way to measure its setting. Furthermore, such a design offsets the effects of changing the aperture and exposure time of the lens, rendering those control parameters useless for controlling the intensity sensitivity range.

Finally, the spectral responsivity of the imaging system can be adjusted, usually by using a filter wheel to rotate one of several filters in front of the lens. We will use c to denote which filter is in place, and denote its spectral transmittance by $\tau_c(\lambda)$. With a filter wheel, c and thus thus $\tau_c(\lambda)$ can be automatically controlled. When comparing an automated filter wheel to a manual filter system, it may appear at first glance that repeatability is not an issue since the manual system can certainly select each desired filter at the proper time. However, automation is important for two reasons. First, as noted above, changing filters manually may make the system vulnerable to being "bumped" out of position by the human operator. Second, perhaps more important, an automated system will always place the same filter *in the same position* in front of the lens -- thus, any irregularity in the makeup of the filter or any dust or markings on its surface will be placed back into the same position each time. This may seem a trivial matter, but it is an important issue for a "warts and all" end-to-end calibration of the system. Of course, a full and detailed treatment of filter irregularities can be a profound study in itself. Yet, a reasonable model of the most important filter properties ought to include some consideration of the optical uniformity across the area. While any non-uniformity may impair the absolute *accuracy* of the theoretical filter model, at least an automated system will preserve the repeatability of the effect and thus preserve the *precision* of the calculation.

2.5. Summary of the Imaging Space Model

We have now presented a 12-parameter model that captures the controllable parameters of a robot vision imaging system:

- T_x, T_y, T_z, θ_x, θ_y, θ_z: the location and orientation of the imaging system
- F, D, A: the focal length, focus distance, and aperture diameter of the lens
- E, G: the exposure time and gain of the camera and digitizer
- τ_c: the transmittance of the selected filter

This model, which we call the Imaging Space, is a configuration space for robot imaging systems. In a specific system, each of the parameters can be characterized by whether it is manual or automated, by

the range over which it can be adjusted, and the repeatability with which it can be controlled.

In this presentation we have not discussed the issues of controlling the illumination, although one might certainly develop a similar model for each light source under automated control.

3. Calibration and the Imaging Model for the Imaging Space

By itself, the Imaging Space is only a parameterization of the imaging system. To be useful, it must be accompanied by an imaging model and a calibration methodology. The purpose of an *imaging model* is to establish the correspondence between the physical reality of the scene and the data measurements contained in the image, as determined by the Imaging Space characterization of the camera. Each ray in the scene can be parameterized as a point of origin (x,y,z) and a direction given by the vector gradient $(p,q) = (\Delta x/\Delta z, \Delta y/\Delta z)$ [22]. Using λ to represent the wavelength of light, we can then express the radiance of light in any ray as $L(x,y,z,p,q,\lambda)$. The image data can be parameterized as an integer function $P(r,c)$, where P is the pixel value and (r,c) are the integer image coordinates (row and column). With these definitions in mind, the goal of the imaging model can be stated as the establishment of a relationship between the scene, the imaging parameters of the Imaging Space, and the image, which may be symbolized as follows:

$$L(x,y,z,p,q,\lambda) \leftarrow T_x,T_y,T_z,\theta_x,\theta_y,\theta_z,F,D,A,E,G,\tau_c \rightarrow P(r,c)$$

A specific imaging model will consist of one or more equations relating the scene parameters to those of the image. Such equations typically involve several constants in addition to the imaging parameters; *calibration* is the act of determining the values of these constants for a specific implementation.

Errors in calibration typically reduce the *accuracy* of the imaging model, i.e. they cause an incorrect registration between the scene parameters and the image parameters. The *precision* of the image data, however, is determined by three factors:

- The precision of the control of the twelve imaging system parameters. Using an uncertainty model such as a Gaussian model of the control error, the precision of the imaging system can be related to the precision of the image data through the equations of the imaging model.
- The precision of the calibration measurements. Calibration proceeds by imaging some calibration standards (such as grids or gray scales), using some external measurements of physical parameters of the standards, and deriving the desired constants through manipulation of the imaging model equations. The physical measurements themselves are generally subject to limited precision, which imposes a consequent limitation on the precision of the constant values determined through calibration.
- Inadequate image modeling equations. If the modeling equations do not account for all the necessary parameters in the imaging process, the result will be an inaccurate imaging model. This inaccuracy will usually result in slightly mis-registered data (in either the geometric or radiometric domain). Because computer vision theories assume ideal registration, the discrepancy is generally modeled by simply increasing the assumed uncertainty ("noise") in the image data. Unless the researcher is actively seeking to analyze the nature of the uncertainty, the fact that it is systematic rather than "random" may not be detected at all. Thus, an inadequate imaging model leads to a preventable loss of precision and consequent narrow dynamic range in the imaging system, both geometrically and radiometrically.

This last issue -- the precision loss caused by inadequate imaging models -- is of particular concern for two reasons. First, it is entirely preventable inasmuch as it is due to a faulty theory rather than poor equipment. Second, as we will see, the "first-order" theory on which almost all robot vision calibration is currently based has just such an erroneous assumption at its very heart. To remedy this problem requires the development and adoption of a more complete "second-order" model of imaging.

3.1. First-Order Geometric Calibration

Almost all the calibration literature in the field of computer vision addresses only the geometric aspect of the imaging problem, that is, determining a model to relate each ray in space to the corresponding image pixel through the geometric parameters of the imaging system:

$$x,y,z,p,q \leftarrow T_x,T_y,T_z,\theta_x,\theta_y,\theta_z,F \rightarrow r,c$$

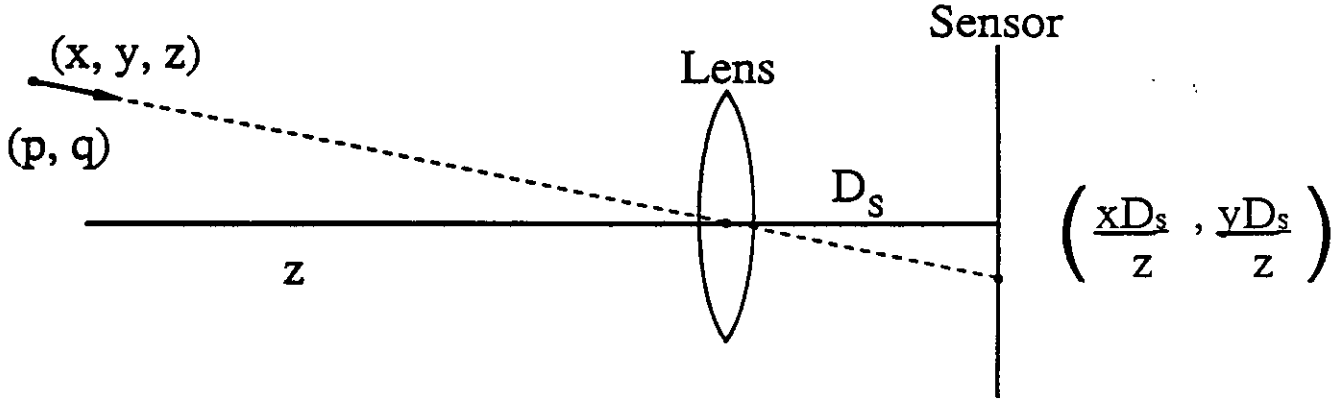

Figure 3-1: Pinhole Model of Imaging Geometry

The most commonly used geometric model is the *pinhole camera* model, in which the lens is modeled as posessing an infinitesimally small aperture (Figure 3-1). This aperture is located at the optical center of the lens, the point used for calculating principal rays from the scene to the sensor plane according to the laws of geometric optics. For an *internal* calibration, the camera coordinate system is used as the frame of reference; thus the *x-y-z* coordinate system aligns with the image coordinates. Only those rays from the scene that pass through the aperture will contribute to the image, thus only rays with $p = x / z$ and $q = y / z$ are considered. If the lens-to-sensor distance is D_s, then such a ray will strike the image plane at coordinates $(pD_S,qD_S) = (xD_S/z,yD_S/z)$ (with suitable assignment of signs). This is the common *perspective projection* imaging model; the value of D_s must be determined by a calibration procedure. In this model, collinear rays are treated identically.

To complete the model, we need to consider the interaction with both the task environment and the image. The assumption that the task and the imaging system use the same coordinate system is rather far-fetched; thus, the transformation between the task and imaging coordinate systems needs to be calculated. This introduces additional variables to be determined by calibration, and allows the imaging system to be related to the world in what we call an *external* calibration. Of course, an external calibration demands that measurements of the scene be made using devices independent of the imaging system, which introduces additional uncertainty into the imaging model.

Similarly, the coordinates of the projection equations must be related to the image coordinates r and c. This is usually modeled by the linear relationships $r = \lfloor a_r + b_r y \rfloor$ and $c = \lfloor a_c + b_c x \rfloor$ where $\lfloor x \rfloor$ is the greatest integer part of x, and the minimum and maximum values of r and c are determined by the camera and digitizer. If desired, this linear model can be back-projected through the perspective projection equations to combine it with the 3-D transformation from scene to image described in the preceding paragraph. Of course, many rays project onto each pixel; for back-projection purposes, usually

the central ray for each pixel is selected. The calculation of the parameters in this type of model has been studied for a number of years [24, 28].

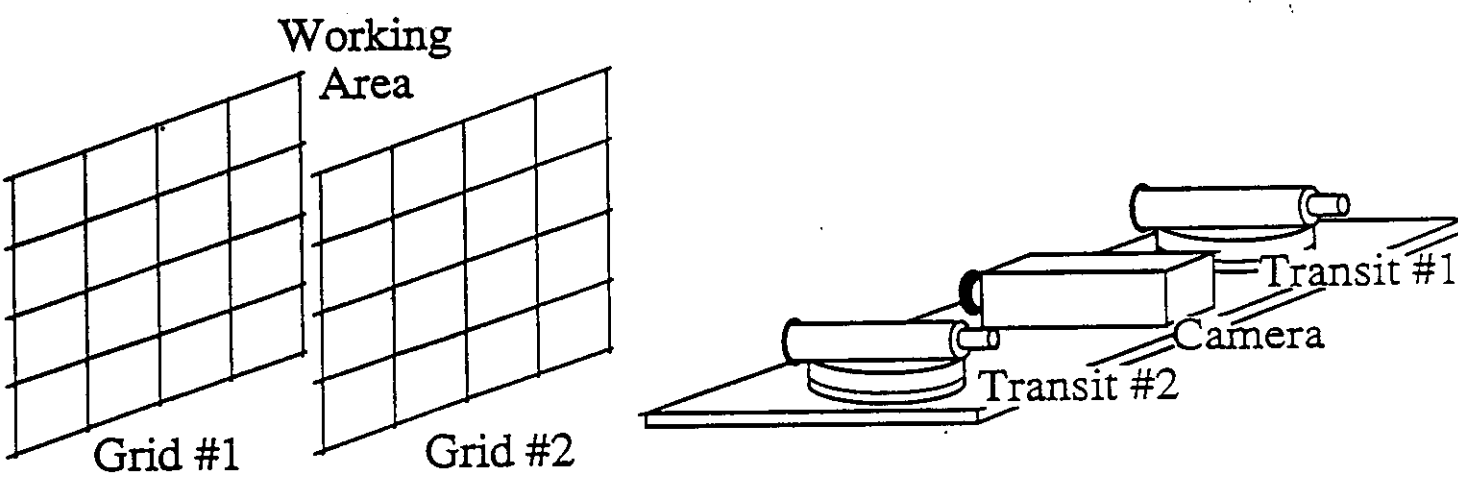

Figure 3-2: Equipment for Geometric Calibration

The assumption that projection coordinates and imaging coordinates are related by a linear model is generally not strictly acccurate, due to optical distortions in the lens and possibly scanning pattern irregularities if a tube-type sensor is used. This type of distortion is frequently modeled as a low-order polynomial or trigonometric function [26]. A more comprehensive model is to relate pixel locations to rays in the scene on a more individual basis, effectively providing separate values for x, y, z, p, and q for each pixel (r,c) in the image (again, with the provision that collinear rays are treated identically) [11, 18]. A lookup table is then used to determine the ray that corresponds to each pixel. In this model, there need be no particular requirement that all the contributing rays in the scene pass through a single point at the optical center of the imaging system. Recent work by Gremban et al. at CMU has developed a refinement of this approach that can be solved by set of linear equations [7]. A sparse lookup table is calculated by the calibration process, and interpolation is used for back-projection from a specific pixel. In this approach, a pair of surveyors' transits are used as shown in Figure 3-2 to give 3D locations of grid points with a resolution of better than 0.2 mm; two images of the grid taken at different distances are then provided to the calibration algorithm. Of course, the precision of moving the grid or camera is also critical to the process. With this approach, back-projection resolution of 0.2 mm can be routinely obtained at a working distance of approx. 0.5 meter for points located in between the two calibration grid positions.

3.2. First-Order Radiometric Calibration

Traditional "camera calibration" methods refer only to the geometric calibration just presented. However, it is equally important to address the radiometric calibration, which relates the spectral radiance in the scene, the lens and camera parameters, and the pixel values in the image as symbolized by:

$$L(\lambda) \leftarrow A,E,G,\tau_c \rightarrow P$$

The radiometric model is developed by following the signal from start to finish through the imaging system. The radiance in the scene, $L(\lambda)$, is transmitted through the filter according to $\tau_c(\lambda)$ yielding a

spectral irradiance proportional to $L(\lambda)\tau_c(\lambda)$ at the sensor. The sensor responds to each wavelength according to the *spectral responsivity* $s(\lambda)$; the output signal from the sensor, which we will designate the *integrated spectral signal* I, is integrated over the spectrum according to $I = \int L(\lambda)\,\tau_c(\lambda)\,s(\lambda)\,d\lambda$, with the integration taken over the portion of the spectrum to which the sensor responds. The pixel value P is then modeled as some function of I such as linear, logarithmic, or exponential (I^{γ}, the "gamma" model, with $\gamma \approx 1/2.2$ for standard NTSC TV [25]). The camera electronics determine which model is appropriate for a given imaging system. The sensitivity parameters A^2, E, and G are simply taken as multiplicative factors. (For a thin-lens approximation, $1/F^2$ should also be a factor. However, modern compound zoom lenses have a much smaller dropoff function; lacking an adequate model, we will simply sidestep this issue in this presentation.) To this model is typically added Gaussian random "noise" denoted $N(\mu,\sigma)$ with mean μ and standard deviation σ. The noise may be factored into the pixel value model; for example, a linear model would be $P = a_p + b_p A^2 EGI + N(0,\sigma)$. Noise can be correlated with the intensity level I and with the control parameters E and G, as described below; these effects have not been included in this simple linear model. Photometric (intensity) calibration for this linear model would be aimed at determination of a_p, b_p, and σ using, for example, an image of a precision gray-scale chart. Of course, a complete radiometric calibration also requires measuring $s(\lambda)$ using a monochromator of some sort.

Several problems arise in the application of such a radiometric model. First, the illumination must be very carefully controlled during calibration or the method must account for the differential illumination across the gray scale. Recalling that a single pixel level of an 8-bit value represents about 1/250 of the maximum response, high accuracy in illumination is seen to be critical. It is virtually impossible to control the illumination to such a degree; thus, some high-precision radiance measurement must be used to provide radiometric calibration data.

The model for the relationship between I and P may not be accurate for any number of reasons. One option to deal with this is to measure a number of ordered pairs (I,P) and interpolate a curve to fit these measurements [17, 15]. Then, table lookup can be used for each pixel value P to determine the corresponding value of I. This is the radiometric analog of the geometric table-lookup method described above, and serves the same function of accounting for all the signal transformations without modeling each one explicitly. Unfortunately, as shown below, such a model cannot account very well for the dependency of the calibration parameters on the imaging system control parameters E and G, and thus would not be suitable for cameras with adjustable values for these parameters.

Radiometric models may require separate calibration at *each pixel*, especially for the constants a_p and b_p. This is primarily due to unevenness in the thickness of the photosensitive material or similar hardware non-uniformities. Thus, these variables should properly be denoted $a_p(r,c)$ and $b_p(r,c)$. To calibrate these values requires imaging flat-fields of uniform intensity rather than a gray chart; since a typical "flat field" is uniform to about 1 part in 30 rather than 1 part in 250, obtaining a suitable large flat intensity field is a severe problem. One approach would be to scan the camera geometrically so that the pixel values can be related to each other by imaging the same areas of the scene. This, of course, requires that a very precise geometric calibration and control system be in place before radiometric calibration can be performed. For the same reasons, the spectral responsivity $s(\lambda)$ can vary across the image, with an exponential form $s(\lambda,r,c) = s(\lambda)^{a_s(r,c)}$ whose exponent a_s is independent at each pixel location (r,c). All of these parameters may also vary with the lens parameters, particularly A, because different parts of the image may be affected differently as the lens optics are changed. Calibrating these effects is beyond the scope of this paper.

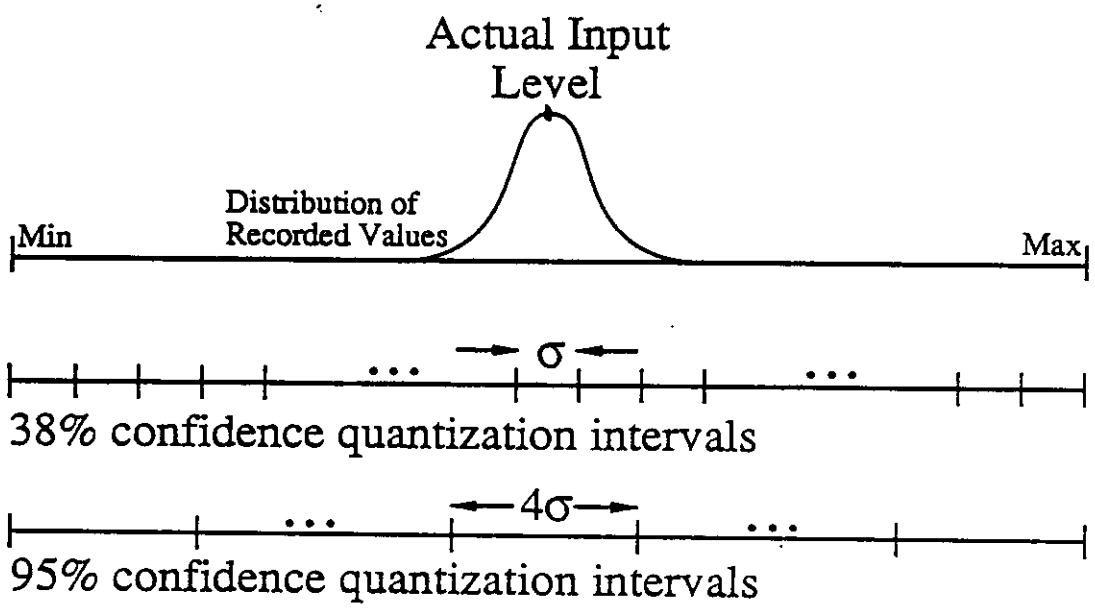

Figure 3-3: Quantization and Repeatability Confidence

The interpretation of the noise parameter σ also merits some comment, since it arises in many disguises in the literature. It must be emphasized that **the number of bits produced by a digitizer is not a measure of the radiometric precision.** The proper measure is the standard deviation of the pixel values, the R.M.S. "noise" value σ; the dynamic range of the system is then defined as the signal-to-noise ratio given by $SNR = \frac{max - min}{\sigma}$, where *max* and *min* are the maximum and minimum pixel values that the digitizer is capable of producing. The SNR may be expressed in *decibels (dB)*, given by $dB = 20 \log_{10} SNR$. Another common way to express the SNR is by the number of "bits per pixel". This is commonly given as $bits\text{-}per\text{-}pixel = \log_2 SNR$. In all these cases, the SNR must be interpreted with great care. Recalling that the underlying model is Gaussian with standard deviation σ, the probability that any given pixel in the image is within a "resolution unit" of size σ is only about 38% (the area under the normal curve from $-\frac{\sigma}{2}$ to $+\frac{\sigma}{2}$), as shown in Figure 3-3. Thus, the confidence level for pixel values at this resolution is only 38%. Using a 95% level instead ($\pm 2\sigma$), the appropriate resolution unit should be 4σ wide; thus, at a 95% confidence level, the number of bits per pixel will be seen to be two bits less than the commonly reported figure!

Finally, the model given above assumes that the noise can be modeled as a simple Gaussian random value added to the pixel value. This model cannot account for the dependence of the final pixel value upon the exposure time or digitizer gain; for example, a long exposure time obviously results in a large integrated noise value from the sensor (though the SNR will increase). To model these effects, we need a somewhat more careful model of the transformation from I to P, with particular attention to the noise introduced at each stage shown in Figure 2-5. First of all, we can model the noise in the sensor element per unit time as the Gaussian $N(\mu_s,\sigma_s)$, which can also be expressed as $\mu_s + N(0,\sigma_s)$. The signal from the chip is then $A^2EI + E\mu_s + N(0,\sigma_s\sqrt{E})$. (For a deeper discussion of noise in CCD imaging devices, including non-Gaussian models for the various processes, see [1].) The amplifier introduces additional noise $N(\mu_a,\sigma_a)$ and scales the result by the gain G; finally, the digitizer may introduce additional noise $N(\mu_d,\sigma_d)$

before quantifying the signal. The resulting model is:

$$P = \lfloor A^2EGI + GE\mu_s + G\mu_a + \mu_d + N(0,\sigma_s\sqrt{EG}) + N(0,\sigma_a\sqrt{G}) + N(0,\sigma_d) \rfloor$$

The value of P is then limited to some maximum possible output from the digitizer, which may or may not utilize the full range representable by the bits available in the pixel value. This digitizing model, of course, presents quite a few parameters to be calibrated (μ_s, σ_s, μ_a, σ_a, μ_d, and σ_d). If all the noise is lumped together as $N(0,\sigma)$, then this model differs from the simple model presented earlier in the inclusion of the terms $GE\mu_s$ and $G\mu_a$ that model the dependence of the signal level on the control parameters G and E; here, the term μ_d corresponds to a_p in the simple model. As described above, the variations over the image sensor plane should be modeled by parameterizing $\mu_s(r,c)$ and $\sigma_s(r,c)$ according to the image coordinates, along with perhaps a linear scaling $a_s(r,c) + b_s(r,c)I$ substituted for the value of I in the term A^2EGI to reflect variations across the sensor plane. Depending on the exact nature of the scanning (e.g. parallel digitization of an entire row at once), some additional correction terms might be appropriate to include in the model. Of course, to this uncertainty model must be added the control uncertainty for the parameters A, E, G, and $\tau_c(\lambda)$.

3.3. Second-Order Integrated Calibration

The above discussion of geometric and radiometric calibration is based on a very simple assumption: that these aspects of the imaging model can be separated from each other and treated independently. Because the resulting models deal only with one domain or the other, we call them "first-order" models of the imaging process. A more comprehensive "second-order" model would be one that deals with the two domains as interdependent upon each other, without attempting to separate them for the sake of tractability. Thus, second-order models can capture the relationship between the entire field of incoming light and the value of each pixel in the image.

Using a second-order model is critical for highly accurate image modeling for several reasons. First, let us consider the effect of focusing the lens. As shown in Figure 3-4, the first-order model assumes a pinhole camera approximation. Only those rays that impinge upon the pinhole are considered in the model; all such rays, of course, will be perfectly focused on the sensor plane. Thus, the first-order models cannot model the effect of varying the focus distance D (to verify this, simply scan over the preceding sections of the paper on geometric and radiometric calibration!). This is an area in which the basic second-order model can be readily characterized, and has been utilized in software-controlled focus systems such as [16]. The key concept is to abandon the pinhole model of perspective and account instead for all the rays incident upon the aperture of the lens. For any point on the focussed plane in the scene, there is a bundle of rays that intercept the effective lens aperture. These rays are bent by the lens to converge at a point on the sensor plane determined by the principal ray passing through the optical center. The direction of the principal ray is not affected by the lens, thus it obeys the laws of perspective. For any point $P = (x,y,z)$ not on the focussed plane, there is a similar bundle of rays intercepting the aperture. This bundle of rays intercepts the focussed plane in a circle, thus the area intercepted on the sensor plane is also a circle called the *blur circle* for P. The center of the blur circle is the perspective projection of P, i.e. $(xD_s/z, yD_s/z)$. The radius will be $AD_s\ (1/D_0 - 1/z)$, which can be seen to vary with all the lens parameters A, D, and F. Thus, the effects of focusing link the geometric and radiometric imaging system parameters closely together. This is a thin-lens model; a more accurate thick-lens model would also account for displacement of the principal ray [13].

It might seem that the literature on software focusing would have developed general second-order

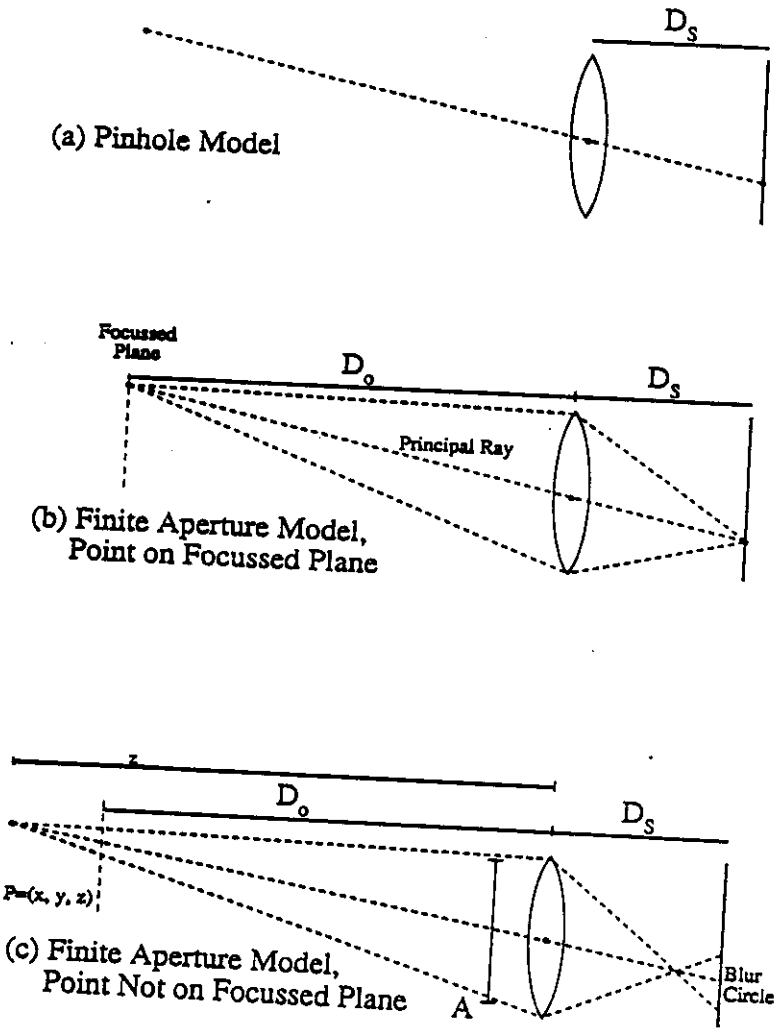

Figure 3-4: Pinhole Model v. Finite Aperture Model

camera modeling and calibration techniques, but such is not the case. Published methods for focus control solve the problem of achieving optimum focus for a single point or area in the image, typically the center point, by adjusting *D* to maximize the high-frequency signal content in the desired neighborhood. Pentland's analysis of focus and aperture comes somewhat closer to the mark, showing the relationship between aperture diameter and local image intensity behavior, but for example does not model the interaction of focus and image magnification [21]. These formulations establish some of the key relationships, but fall somewhat short of modeling the complete interaction between radiometry and geometry caused by focus effects. The modeling of image magnification, in particular, is essential for any subsequent geometric analysis of the image data such as inferring the size of a 3D feature by analysis of its image. Software focussing research does not usually include such geometric inference from the image; on the other hand, geometric robot vision research generally assumes fixed focus which is set manually before calibration is performed. Thus, further research is needed in the geometric modeling of focus adjustment.

The blur circle is an example of *crosstalk* among the incident rays, which is a non-impulse point-spread function for the response of the imaging system to each incoming ray of light. Several other factors can contribute to such crosstalk:

- *Optical flare and diffraction:* Poor optical quality in the lens may cause noticeable amounts of light from a ray to scatter into adjacent pixels. This is generally not a severe problem in robot vision, but it can be noticeable if the pixels are very small, as in a high-resolution detector. Diffraction, which also causes a broadening of the point-spread function, is also an inherent effect of a lens.
- *Mismatched electronics:* Because the signal from the camera to the digitizer is analog, there will be some frequency limitation imposed by the electronics in the signal path. The use of standard analog video regimens such as RS-170 imposes noticeable frequency limitations. In addition, long transmission distances and other aspects of the video signal distribution configuraton can cause additional frequency limitations. Further, since the sensor itself may not have the resolution of the digitizer, the analog signal may be re-sampled by the digitizer in a way that smooths over several distinct sensor elements. For example, common TV-grade CCDs have 384 columns of sensor elements, yet most commonly used digitizers sample each row 512 times.
- *Smoothing over blemishes:* As described above, sensors typically have some non-uniformity across the sensor plane resulting in slightly different radiometric response at each pixel. To compensate for this, it is not unusual in TV-grade equipment to smooth the signal across each row of the image, thus inducing a local averaging process. Such smoothing introduces crosstalk among the pixels.

Such effects as these are typically rather constant over time and space, and can thus be grouped together into the point-spread function for the imaging system. Our lab experiences with TV-grade equipment indicate a point-spread function for our camera that can be modeled as a Gaussian with horizontal σ of about 1.2 (pixel widths) and vertical σ about 0.5.

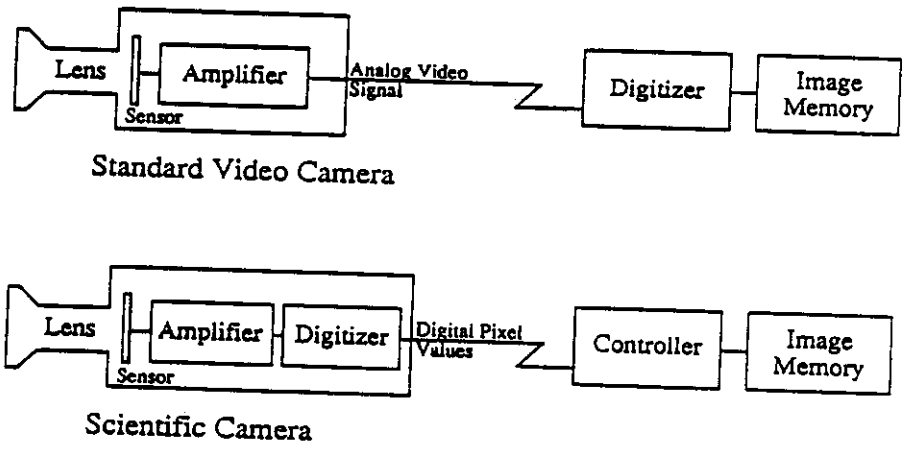

Figure 3-5: Video and Scientific Cameras

Scientific-grade cameras have a telling advantage over TV-grade equipment in posessing low crosstalk. The key advantages of scientific cameras can be summarized as:

- *Controllable exposure time.* The ability to control the exposure time under software means that the imaging system can take advantage of small apertures for a wide depth-of-field, increase the SNR of the pixel values, respond to greater variation in the scene radiance level, etc.
- *High SNR.* As described above, the use of cooled CCDs allows scientific cameras to keep

sensor noise down to a very low level.

- *Low crosstalk.* In a scientific camera, typically each pixel is *independently* digitized inside the camera body itself as shown in Figure 3-5, before any substantial analog signal transmission has occurred. Further, the use of smoothing circuits for covering blemishes is avoided. The result is a narrow point-spread function, which enhances the accuracy of a first-order calibration model and thus contributes noticeably to geometric and radiometric precision.

The price for using a scientific camera includes the cost, the relatively slow digitization time (e.g. 1 frame/second), the need to calibrate each pixel's radiometric response independently, and if the sensor is cooled, the need for external support equipment.

Closely related to crosstalk is the integration of radiometric information over finite pixel areas. This gives rise to *spatial aliasing* in the image. Crosstalk and spatial aliasing can be concisely modeled in the Fourier domain.

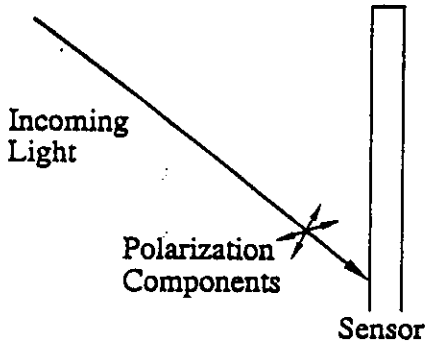

Figure 3-6: Polarization of Incoming Light at the Sensor Plane

All of the above second-order effects are essentially uniform across the sensor plane. However, the value at each pixel may also be affected by a difference in the responsivity of the sensor to incident rays that varies according to the pixel position. One example is the well-known proportionality factor of $cos^4\,\theta$, where θ is the angle of incidence upon the sensor plane for each pixel [14]. In addition, it is generally the case that photosensitive material is differentially sensitive to each incoming ray, depending on the angle between the ray and the sensor plane. This is primarily due to a concept not yet addressed: *polarization.* In addition to the well-known interaction between a polarizing filter and the incoming light, there is a less-well-known interaction with the sensor plane. Light that strikes the sensor plane at an angle can be characterized as having two polarization components perpendicular to the incoming ray: one is coplanar with the surface normal of the sensor, and the other is perpendicular to the surface normal (Figure 3-6). It is frequently assumed in robot vision that the sensor will be equally responsive to both components. This is not generally so: the coplanar component will cause less response from the sensor [2] (as much as 10% less in our simple lab studies). If the incoming light is unpolarized, the effect is simply a reduced sensitivity at the periphery of the image. However, if the incoming light is highly polarized such as specular reflection from a smooth surface, the effect can be quite substantial [27]. To properly parameterize polarization requires additional parameters for the incoming scene radiance, additional control parameters (if polarizing filters are used), and new imaging models to account for the effect of polarization. Clearly, the robot vision community needs a satisfactory model for the effect of polarization on imaging.

All of the above discussion can be characterized by the use of the principal ray for each image pixel, along with aggregate concepts such as the blur circle. If a more detailed model is needed, each individual ray in the scene can be parameterized as *x-y-z-p-q* and the focus model can be used to calculate a corresponding ray incident on the sensor plane itself. The radiometric calculations of the sensor's responsivity should ideally be based by considering the field of incoming light at the sensor plane, given by $L(x,y,p,q,\lambda)$, rather than assuming a single spectral distribution $L(\lambda)$ incident at each pixel location. (Here, *(x,y)* is the location on the sensor plane, *(p,q)* is the direction of each incident ray at that point, and λ is the wavelength.) This field of light is then integrated over *p* and *q* for each point on the sensor plane, and over *x* and *y* within the area of each pixel, to determine the spectral irradiance for each pixel. This model can account more accurately for orientation-specific effects at the sensor plane.

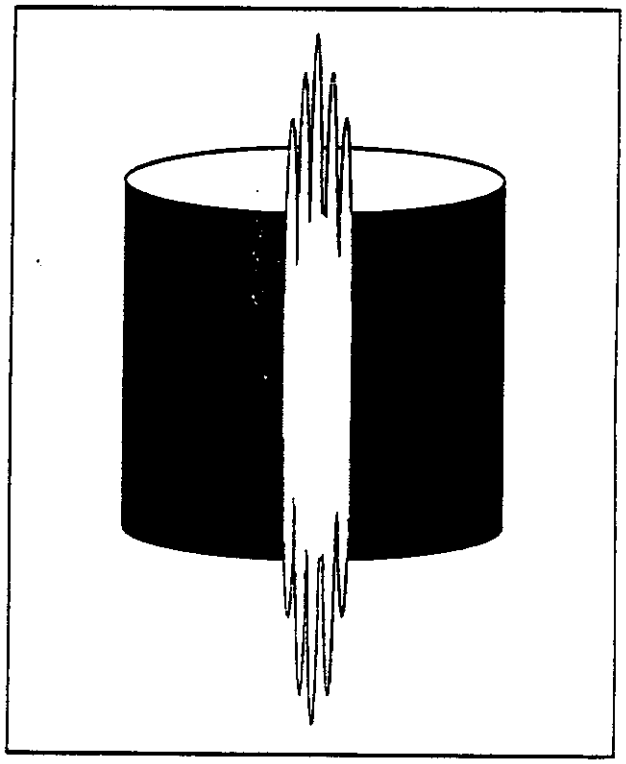

Figure 3-7: Blooming Caused by Bright Highlights

A final issue to motivate the use of second-order models is the effect of *blooming*, which is quite noticeable in CCD sensors. Blooming occurs when a pixel receives so much irradiance that the photosensitive layer cannot hold all the resulting charge; the excess then spills over into neighboring pixels, causing their reported values to be higher than they ought to be. Since most CCD imaging sensors have pixels that are arranged in columns of pixel cells, blooming tends to be mostly or entirely in the vertical direction as shown in Figure 3-7. Blooming can cause very dramatic effects in the image, including hue changes in color imaging [15]. Prospects appear dim for modeling blooming in a useful way, because it is frequently a very non-local effect in the image.

In all these situations, the first-order models commonly in use fail to adequately model the interaction

between geometric and radiometric aspects of the imaging process. When such models are used in practice, the resulting inaccuracy is subsumed in "noise" measurements that are actually much higher than the inherent limitations of the hardware. The development of useful representations and calibration procedures for second-order imaging models is thus an area that requires further research.

4. Higher-Level Issues in Imaging System Control

This paper has presented a discussion of imaging precision that shows the relationship between dynamic range, precision, imaging models, and calibration. Unlike past literature, this presentation has shown how the same concepts can be applied to understand both the geometric and radiometric domains, either independently through "first-order" models or interdependently in a "second-order" framework. This unified view is made possible by the introduction of the Imaging Space model of the complete set of imaging system parameters. We have also described how the uncertainty inherent in measurement and control processes gives rise to uncertainty in the image data, which can be modeled explicitly. We have included a discussion of some existing technologies for robot imaging, along with indications of their inherent precision limitations.

One of the thrusts of this paper has been the proposition that automation in the imaging process is a key ingredient for obtaining high precision, ease of control, and a wide dynamic range. With automation arises an important new question: What strategy or goals should be used for controlling the parameter settings? This can be modeled formally by allowing each control variable to be a function of time t, e.g. $T_x(t)$ for lateral motion, $T_z(t)$ for forward motion, $\theta_y(t)$ for pan, $D(t)$ for focusing, etc. The imaging system state thus follows a trajectory in the Imaging Space, which can be traversed through automation in the imaging system. The physical system parameters of course change correspondingly over time; thus, the sensor plane receives an image that varies continuously over time and can be captured when desired. Using the imaging model, the changes in the control parameters can be related to image properties. Some questions that arise for each parameter are:

- What is the effect (or motivation) for controlling this parameter? For example, altering $F(t)$ changes the angular width of the field of view, and also the magnification of the image.
- How can data be usefully aggregated as this parameter is varied? For example, aggregating as $T_z(t)$ is varied is the paradigm of "optical flow" [9].
- What control strategy should be used for this parameter? For example, $D(t)$ can be controlled in an autofocus mode to find a single "optimal" value for imaging the scene (for a recent survey of this topic, see [16]). Another strategy would be to sweep $D(t)$ continuously from its minimum to its maximum setting, recording at each pixel the setting producing the best focus; the result would be a complete depth image produced by focus manipulation.
- How might this parameter be usefully sampled? For example, sampling $T_x(t)$ is known as "stereo"; sampling $\theta_y(t)$ gives a "panoramic" composite image with a very wide field of view useful for navigation [6]; sampling $A(t)$ can be used for looking into shadows [12].

By considering each of these issues for each control parameter, much of the design of a robot imaging system can be cast into a single comprehensive framework. This can serve to clarify some issues such as the use of multiple cameras as opposed to a single controllable camera.

Based on such considerations, it should be possible to construct fully automated robot imaging systems that act as "smart cameras", using active control techniques to optimize the imaging quality based on the task being performed, the needs of imaging (such as wide apertures for precise focusing and narrow apertures for large depth-of-field), and producing super-high quality images by compensating for distortions (such as the coupling of focus distance and chromatic aberration with image magnification).

There appear to be four inter-related aspects of the imaging specifications to be studied:

- *Aiming* the camera -- adjusting the position and orientation. Constraints include obtaining a clear line of sight, focus and framing limits, obtaining a desirable viewing angle, and modeling the kinematics and dynamics of the camera mount which may include motion of a vehicle.

- *Framing* the desired features -- adjusting magnification. Constraints include pixel resolution, field of view, obtaining uncluttered images, and avoiding motion blur.
- *Optimizing* the image quality -- focus and exposure. Past efforts have generally optimized only a single point in the image such as the perspective center point. Also, the interaction between these parameters and image magnification is still an open research area.
- *Acquiring* the resulting images. In some situations, different imaging parameters may be used for optimizing the system and for actually acquiring the desired image data for vision analysis.

Some of these imaging configuration issues, particularly Aiming, are being studied by Cowan and Kovesi, although they have not cast the corresponding modeling and calibration problems in the same framework [4].

We call this area of research "Imaging Understanding" -- developing a deeper understanding of the imaging process and all the attendant issues for robot task automation. At the Calibrated Imaging Laboratory of Carnegie Mellon University, we are developing a fully automated twelve-degree-of-freedom imaging system, including high-precision positioning equipment, an uncooled scientific camera, an automated lens and filter wheel, and external equipment for geometric and radiometric calibration (surveyors' theodolites and spectroradiometer) [23]. With this facility, we are pursuing research in Imaging Understanding as well as obtaining high-quality data for research in robot vision theory and practice.

4.1. Acknowledgements

Thanks to John Krumm for comments on this paper, to Robert Nathan at JPL for his thoughts on radiometric calibration, and to Thad Druffel, Keith Gremban, and Richard Volpe for their work on geometric calibration in the Calibrated Imaging Lab.

References

[1] Beynon, J.D.E. and Lamb, D.R.
Charge-Coupled Devices and Their Applications.
McGraw-Hill, London, 1980.

[2] Budde, W.
Optical Radiation Measurements. Volume 4: *Physical Detectors of Optical Radiation.*
Academic Press, New York, 1983.

[3] Canny, J. F.
A Computational Approach to Edge Detection.
IEEE PAMI 8(6):679-698, November, 1986.

[4] Cowan, C. K. and Kovesi, P. D.
Automatic Sensor Placement from Vision Task Requirements.
IEEE Trans. on Pattern Analysis and Machine Intelligence PAMI-10(3):407-416, May, 1988.

[5] Gennery, D. B.
Modelling the Environment of an Exploring Vehicle by Means of Stereo Vision.
PhD thesis, Stanford U., June, 1980.

[6] Goto, Y., Shafer, S.A., and Stentz, A.
The Driving Pipeline: A Control Scheme for Mobile Robots.
Technical Report, CMU Robotics Institute, 1988.
In press.

[7] Gremban, K. D., Thorpe, C. E., and Kanade, T.
Geometric Camera Calibration Using Systems of Linear Equations.
In *Proc. Intl. Conf. on Robotics and Automation.* IEEE, 1988.
To appear.

[8] Healey, G. and Binford, T.O.
Local Shape from Specularity.
In *Proc. First Intl. Conf. on Computer Vision*, pages 151-161. IEEE, June, 1987.

[9] Horn, B.K.P. and Schunck, B.G.
Determining Optical Flow.
Artificial Intelligence 17:185-203, 1981.

[10] IES.
IES Lighting Handbook, Reference Volume.
Illuminating Engineering Society of North America, New York, 1981.

[11] Isaguirre, A., Pu, P., and Summers, J.
A New Development in Camera Calibration: Calibrating a Pair of Mobile Cameras.
In *Proc. Intl. Conf. on Robotics and Automation*, pages 74-79. IEEE, 1985.

[12] Kanade, T. and Thorpe, C.E.
NAVLAB Road Following: 1988 Annual Report.
Technical Report, CMU Robotics Institute, 1988.
In preparation.

[13] Kingslake, R.
Lens Design Fundamentals.
Academic Press, New York, 1978.

[14] Kingslake, R.
Optical System Design.
Academic Press, 1983.

[15] Klinker, G.J., Shafer, S.A., and Kanade, T.
The Measurement of Highlights in Color Images.
Intl. J. of Computer Vision , 1988.
In press.

[16] Krotkov, E.
Focusing.
Intl. Journal of Computer Vision 1:223-237, 1987.

[17] Leclerc, Y.
Private communication.
1986.

[18] Martins, H. A., Birk, J. R., and Kelley, R. B.
Camera Models Based on Data from Two Calibration Planes.
Computer Graphics and Image Processing 17:173-180, 1981.

[19] Matthies, L., Szeliski, R., and Kanade, T.
Kalman Filter-Based Algorithms for Estimating Depth from Image Sequences.
Technical Report CMU-CS-87-185, CMU Computer Science Dept., December, 1987.
Submitted to Conference on Computer Vision and Pattern Recognition.

[20] Mitchell, O.R., Lyvers, E.P., Dunkelberger, K.A., and Akey, M.L.
Recent Results in Precision Measurements of Edges, Angles, Areas and Perimeters.
In *Proc. Conference 730 on Automated Inspection and Measurement.* SPIE, October, 1986.

[21] Pentland, A.
A New Sense for Depth of Field.
In *Proc. Intl. Joint Conf. on Artificial Intelligence*, pages 988-994. IJCAI, 1985.

[22] Shafer, S. A., Kanade, T., and Kender, J. R.
Gradient Space Under Orthography and Perspective.
Computer Graphics and Image Processing 24:182-199, 1983.

[23] Shafer, S. A.
The Calibrated Imaging Lab Under Construction at CMU.
In *Proc. Image Understanding Workshop.* DARPA, December, 1985.

[24] Sobel, I.
On Calibrating Computer Controlled Cameras for Perceiving 3D Scenes.
Artificial Intelligence 5:185-198, 1974.

[25] Sproson, W.N.
Colour Science in Television and Display Systems.
Adam Hilger Ltd., Bristol, 1983.

[26] Tsai, R. Y.
An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision.
In *Proc. Intl. Conf. on Computer Vision and Pattern Recognition*, pages 364-374. IEEE, 1986.

[27] Wolff, L.B.
Spectral and Polarization Stereo Methods Using a Single Light Source.
In *Proc. Image Understanding Workshop*, pages 810-820. DARPA, February, 1987.

[28] Yakimovsky, Y. and Cunningham, R.
A System for Extracting Three-Dimensional Measurements from a Stereo Pair of TV Cameras.
Computer Graphics and Image Processing 7:195-210, 1978.