A Model for Image Sensing and Digitization in Machine Vision

Muralidhara Subbarao   Arman Nikzad

Department of Electrical Engineering
State University of New York at Stony Brook
Stony Brook, NY, 11794-2350

Abstract

A mathematical model for a typical CCD camera system used in machine vision applications is presented. This model is useful in research and development of machine vision systems, and in the computer simulation of camera systems. The model has been developed with the intention of using it to investigate algorithms for recovering depth from image blur. However, the model is general and can be used to address other problems in machine vision. The model is based on a precise definition of *input* to the camera system. This definition *decouples* the *photometric properties* of a scene from the *geometric properties* of the scene in the input to the camera system. An ordered sequence of about 20 operations are defined which transform the camera system's input to its output, i.e. digital image data. Each operation in the sequence usually defines the effect of one component of the camera system on the input. This model underscores the complexity of the actual imaging process which is routinely underestimated and oversimplified in machine vision research.

1 Introduction

Experimental verification of computer vision theories often requires sophisticated and expensive camera systems. When such camera systems are not available, then one can simulate such a camera system on a computer. This simulation however requires a mathematical model of the camera system. In this paper we present a mathematical model for a typical camera system used in machine vision applications.

The motivation for developing a model of a camera system came from our need for a precision camera system whose parameters could be controlled and measured accurately. We needed such a camera system for conducting experiments on a new method of finding the distance of objects [12]. However, no such camera system was available in the market.

The model presented here is also useful in machine vision research. It makes explicit the sequence of transformations that the light incident on a camera system undergoes before being sensed by the image detector, and the sequence of operations by which the sensed image is converted to digital data.

We found that previous literature in computer vision (e.g. Ch. 1 in [8], [6], [9]) on a camera system is descriptive in nature. Shafer's [9] work is perhaps the first to deal with a *detailed* description of the image sensing process. Many important aspects of the image sensing process which are routinely ignored by computer vision researchers are discussed in detail by him. A twelve parameter model for a robot imaging system - six parametrs in camera position and orientation, three in the optical system, and three in sensitivity - is presented. In addition, valuable information on the state of the art in imaging system technology and calibration is presented by Schafer [9].

A camera system (or any physical system) can be modeled at many levels of detail and abstraction. Here we have attempted to focus on developing what *we* believe is a useful mathematical model rather than a *detailed* mathematical model. However, it is certainly possible that there are situations where a more detailed model (or a less detailed model) is needed than the one presented here.

Fig. 1 shows a block diagram of a solid-state video camera system which models a typical system used in machine vision applications. Fig. 2 provides additional information about the roles of different parts of the system. In these diagrams, the order in which certain parts appear from left to right can be changed without affecting the image recorded by the system. For example, the position of Light Filter (LF) and Aperture Stop (AS) can be interchanged, or the position of Light Filter and Shutter (implemented electronically) can be interchanged. However, the amount of computation required for simulating the system on a computer may be affected by the order of the parts.

2 Generalized Image Forming Optical System

Fig. 1 shows a single lens and a single aperture. However, an actual image forming optical system may contain many lenses and apertures. In this case, all these elements may be lumped into a single "black box", and the significant properties of the optical system can be completely described by specifying only the *terminal properties* of the aggregate [3]. The terminals of the black box are an *enterance pupil*, representing a finite aperture (effective or real) through which light must pass to reach the imaging elements, and an *exit pupil* (again effective or real), representing a finite aperture through which light must pass as it leaves the imaging elements on its way to the image detector plane. It is usually assumed that passage of light between the entrance and exit pupil planes can be adequately described by geometric optics [5]. The smallest aperture in the system (the aperture which determines the amount of light reaching the image detector), i.e. the Aperture Stop (AS), determines the entrance and exit pupils.

Entrance pupil is the geometrical image of the aperture stop as seen from an axial point on the object through those imaging elements(lenses) preceding the aperture stop [5]. If there are no elements preceding the aperture stop, then the aperture stop itself is the enterance pupil.

Exit pupil is the image of the aperture stop as seen from an axial point on the image detector plane through the interposed (imaging elements) lenses, if there are any. If there are no interposed lenses, then the exit pupil is the same as the aperture stop. We will restrict our camera system to a circularly symmetric optical system, therefore the entrance and exit pupils are circular.

3 External Parametrs

We define a spherical coordinate system with its Z-axis along the optical axis and the origin fixed in the plane of the entrance pupil (see Fig. 3). We call this the Entrance Pupil coordinate system, or the ENP coordinate system for short.

The *external parameters* of the camera system are the position and orientation of the entrance pupil coordinate system with respect to an external coordinate system (X_o, Y_o, Z_o, O_o) in the object space (see Fig. 3). Let the position of the entrance pupil origin O with respect to the external coordinate system be (O_x, O_y, O_z) and orientation of the entrance pupil coordinate system with respect to the external coordinate system be specified by three suitable rotation parametrs $(\theta_x, \theta_y, \theta_z)$.

4 Scene Parameters

We shall consider a scene which contains only objects with opaque surfaces. In this case, for every direction defined by (θ, ϕ) in the entrance pupil coordinate system, there is a unique point P (Fig. 3) on a visible object in the scene. If r is the distance of P from the orogin O, then the geometry of visible surfaces in the scene can be completely defined by a function denoted by $r(\theta, \phi)$. $r(\theta, \phi)$ gives *geometric information* about visible surfaces in the scene. We shall assume that the geometry of the scene remains the same during the time interval of interest, i.e. there are no moving objects in the scene. If there are moving objects, then one may take it into consideration by specifying the geometry as function of time t as $r(\theta, \phi, t)$.

What the camera "observes" due to a point light source at P is the electromagnetic field distribution produced by the source at the entrance pupil (Fig. 3). Although this field distribution can be specified by a very general function, we will only consider those cases and quantities which are useful for our purposes. The case we consider corresponds to an incoherent light source and unpolarized light, and the quantities we consider are radius r of the wavefront incident on the entrance pupil, and the energy strength of the wavefront as a function of its wavelength λ. The other quantities one can consider are the degree of coherence of the light source, and the direction of polarization of light. We shall assume that these quantities remain the same during one exposure period (T) of the camera system. We consider the radius r of the wave front (i.e. the distance of the point) because, one of our interest is to model defocusing in the camera system [12, 13].

We take the inputs to the camera system to be the geometry of visible surfaces specified by $r(\theta, \phi)$, and the *photometry* of visible surfaces specified by $f(\theta, \phi, \lambda, t)$ which is the power of light of wave length λ incident on the entrance pupil from the direction (θ, ϕ) at time t.

For our purposes, the function $r(\theta, \phi)$ and $f(\theta, \phi, \lambda, t)$ are the inputs to the camera system. We shal call $f(\theta, \phi, \lambda, t)$ as the *apperance of the scene* or the *focused image of the scene*, and $r(\theta, \phi)$ as the *geometry of the scene*. The parameters of the function $r(\theta, \phi)$ and $f(\theta, \phi, \lambda, t)$ are the *scene parameters*. Our view that the input to the camera system is *scene appearance* $f(\theta, \phi, \lambda, t)$ and *scene geometry* $r(\theta, \phi)$ as defined earlier is very important in modeling the camera system. In particular we note that $f(\theta, \phi, \lambda, t)$ and $r(\theta, \phi)$ as defined by us are quantities which are directly observable by the camera system. There is no interdependence between these quantities. This should be compared to the traditional convention of characterizing the photometric aspects of the scene in terms of scene radiance. The radiance of a small surface patch in the scene is the light power emitted by the surface patch into a unit solid angle per unit area, and has units of $Watts/m^2/Steradian$. Since the area of a surface cannot be estimated without a knowledge of the orientation of the surface patch with respect to the camera system, the radiance of the surface patch cannot be observed by the camera system independently. Therefore, characterization of scene appearance in terms of the radiance of surface patches in the scene invariably *couples* the photometric and geometric properties of the scene. In contrast, our characterization of scene properties in terms of $f(\theta, \phi, \lambda, t)$ and $r(\theta, \phi)$ *decouples* the photometric and geometric properties.

Another problem with using radiance is that it cannot be used to define the "brightness" of a point light source as a point has no surface area. We will find it necessary to define the brightness of a point light source in order to be able to define the point spread function of the camera system.

In our methods for finding the distance of objects from their blurred images [11, 12, 13], we will need to compare two or more images of the same object recorded with different camera parametrs. The camera parameters that are set to different values for the two images include one or more of: the diameter of the aperture, the focal length of the lens, and the distance between the lens and the image detector. In order for us to decouple the effect of the geometry of the scene from the photometric properties of the scene, the aperture stop experinced by all objects in the scene should be the same. In particular, it should not depend on the distance of objects in the scene.

In Fig. 4, there are two stops A_1 and A_2. For a point P_1 at distance U_1, A_1 detrmines the aperture stop. For a point P_2 at distance U_2, A_2 determines the aperture stop. This example illustrates how the aperture stop depends on the distance of objects in the scene. Note that, by making A_2 sufficiently small, we can make sure that A_2 determines the aperture stop for both P_1 and P_2. Suppose that we assume that the minimum distance at which an object appears is U_{min}, then the camera can be designed such that one stop determines the aperture stop for all objects at distance U_{min} or higher (i.e. $U \geq U_{min}$). In particular, U_{min} may be taken to be f (focal length) because, for objects nearer than the focal length, the images are virtual.

We further assume that the entrance pupil is not moved when any of the camera parameters (e.g. the focal length of the optical system, or the diameter of the aperture stop) are changed. This is a desirable characteristic in our method of determining distance [11, 12], because, in order to avoid the correspondence problem explained in the Appendix, we require the entrance pupil to be unmoved. Since entrance pupil is not moved, all object distances are measured with respect to it.

5 Input Transformation

In order to compute the image of a scene, information about the scene is required. The scene information may be specified in terms of scene radiance at all visible points on the scene, and scene geometry. Alternatively, one may have to compute the scene radiance given scene illumination and scene reflectance at every point in the scene. Given the necessary scene information, and the positon, orientation, and the diameter of the entrance pupil, one can compute the inputs to the camera system - $f(\theta, \phi, \lambda, t)$ and $r(\theta, \phi)$.

Having defined the input to the camera as the photometric characteristics of the scene specified by $f(\theta, \phi, \lambda, t)$ and the geometry of the scene specified by $r(\theta, \phi)$, we will now define a sequence of transformations of the input signals which transforms the input to the output. The output of the camera is the digital image data. Each step in the sequence of transformations typically corresponds to the effect of one component of the camera system on the input signals. The output of the ith transformation step is denoted by a function of the form f_i for $i = 1, 2, 3, \ldots$. The sequence of transformations defined here can be easily translated into a flow-chart form.

5.1 Light Filter

Typically, camera systems have a light filter to control the spectral content of light entering the camera system. Filters that block infrared rays are widely employed. The characteristic of a light filter can be specified in terms of a transmittance function $T_{LF}(\lambda)$ where $0 \leq T_{LF}(\lambda) \leq 1.0$. The effect of the light filter on the camera input is specified by

$$f_1(\theta, \phi, \lambda, t) = f(\theta, \phi, \lambda, t) \cdot T_{LF}(\lambda), \tag{1}$$

where $f(\theta, \phi, \lambda, t)$ is the input light energy defined earlier, and $f_1(\theta, \phi, \lambda, t)$ is the light energy transmitted by the filter. Usually, light filters have uniform characteristics in space and time. However, if the characteristics change with time and space, then the transmittance function takes the form $T_{LF}(\theta, \phi, \lambda, t)$.

5.2 Vignetting

When there are multiple apertures in the optical system along the optical axis displaced with respect to each other (Fig. 6), the effective light energy transmitted by the system decreases with increasing inclination of light rays with respect to the optical axis [4, 5]. The effect of vignetting can be specified by a function $T_V(\theta, \phi)$ where $0 \leq T_V(\theta, \phi) \leq 1.0$ and the relation between the input $f_1(\theta, \phi, \lambda, t)$ and the output $f_2(\theta, \phi, \lambda, t)$ as

$$f_2(\theta, \phi, \lambda, t) = f_1(\theta, \phi, \lambda, t) \cdot T_V(\theta, \phi) . \tag{2}$$

5.3 Optical System

As discussed earlier, an image forming optical system can be characterized by specifing its properties at the entrance pupil and the exit pupil of the optical system. An image sensing system contains an image sensor or an image detector such as a CCD array or photographic film. An ideal image forming optical system (see Fig. 7) converts a *divergent* spherical wavefront WI incident on the entrance pupil ENP from a point object source P into a *convergent* spherical wavefront WO at the exit pupil EXP which converges onto a point P' on the image detector ID. However, practical image sensing systems convert a divergent spherical wavefront WI incident on the entrance pupil into an emergent wavefront WNI which is different from the convergent spherical wavefront WO. The deviation of the actual emergent wavefront WNI from a hypothetical convergent spherical wavefront WO is associated with optical aberrations. Given the parameters of the optical system, the shape of the emergent wavefront WNI can be determined [1], and given the position of the image detector, the intensity distribution produced on the image detector by the emergent wavefront WNI can be determined. This intensity distribution in fact corresponds to the Point Spread Function (PSF) of the imaging system. If point object sources in the scene are *incoherent*, then the intensity distribution produced on the image detector by different point sources can be simply summed (or added) to obtain the overall image [1]. Therefore, for incoherent point sources, the imaging system acts as a linear system. Hence the imaging can be characterized in terms of a point spread function.

The point spread function of an imaging system is determined by the parameters of the optical system, the distance between the exit pupil and the image detector, the radius of the incident wavefront on the entrance pupil, and the direction of location of the point source. Therefore, the PSF can be specified by a function of the form $h(\theta, \phi, \theta', \phi', r(\theta, \phi), \mathbf{e})$ where (θ, ϕ) is the direction of the point source, (θ', ϕ') is the direction of the point on the image detector in the image space where the intensity value is specified by the PSF, $r(\theta, \phi)$ is the function specifying the geometry of visible surfaces in the scene, and $\mathbf{e}$ is a vector specifying the parameters of the imaging system, such as the diameter of the aperture stop, effective focal length of the optical system, distance between the exit pupil and the image detector, etc. In order to specify the effect of the imaging system on the input, we define normalized Cartesian coordinates in the scene space and the object space as shown in Fig. 8. The relations between the corresponding Cartesian coordinates and the spherical coordinates are

$$x = \sin\theta \, \cos\phi, \qquad x' = \sin\theta' \, \cos\phi', \tag{3}$$

$$y = \sin\theta \, \sin\phi, \qquad y' = \sin\theta' \, \sin\phi'. \tag{4}$$

Now the input $f_2(\theta, \phi, \lambda, t)$ can be equivalently represented by $f_2(x, y, \lambda, t)$, the output $f_3(\theta', \phi', \lambda, t)$ by $f_3(x', y', \lambda, t)$, and the PSF $h(\theta, \phi, \theta', \phi', r(\theta, \phi), \mathbf{e})$ by $h(x, y, x', y', r(x, y), \mathbf{e})$ (to be more precise, we should use different notations for the functions which take the above coordinate transformation into account). The effect of the imaging system can now be specified by the relation

$$f_3(x', y', \lambda, t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y, x', y', r(x, y), \mathbf{e}) \, f_2(x, y, \lambda, t) \, dx \, dy \ . \tag{5}$$

[1] If the point sources are coherent, then the electromagnetic field distribution will have to be summed instead of their intensities [3].

If the scene and the optical system are such that the PSF is spatially invariant in the region of interest (*isoplanetic* region [1]), then the above integral becomes a convolution operation:

$$f_3(x', y', \lambda, t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x' - x, y' - y, r(x, y), \mathbf{e})\, f_2(x, y, \lambda, t)\, dx\, dy \quad . \tag{6}$$

Examples of PSF models can be found in Section 8.8.2 of [1] and in [12].

5.4 Field Stop

The region on the image detector plane which contains photosensors is limited in extent. For example, a CCD array has a finite rectangular shape, and in a photographic camera, only a limited rectangular part of the film is exposed to light. The extent of the image detector containing photosensors determines the field of view of the imaging system and is called Field Stop. The effect of the field stop can be expressed in terms of a transmittance function $T_{FS}(x', y')$ which has a value of 1 inside the image detector region and a value of 0 outside. The relation between the input and the output is

$$f_4(x', y', \lambda, t) = f_3(x', y', \lambda, t)\, T_{FS}(x', y') \quad . \tag{7}$$

We shall assume that the field stop is rectangular in shape with length A and width B. Therefore we have

$$T_{FS}(x', y') = \mathrm{rect}\left(\frac{x'}{A}, \frac{y'}{B}\right) \tag{8}$$

where the function rect is as defined in [2]. The definitions in [2] will be used throughout this report for standard functions such as rect, comb, δ, etc.

5.5 CCD Sensor

5.5.1 Spectral sensitivity of the CCD sensor

The transducer on the image detector which converts light energy to electrical (or other type of) energy is not uniformly sensitive with respect to the wave length of light. If the sensitivity as a function of wave length λ is denoted by $T_s(\lambda)$ where $0 \leq T_s(\lambda) \leq 1.0$, then the relation between the input $f_4(x', y', \lambda, t)$ and the output $f_5(x', y', \lambda, t)$ can be specified by

$$f_5(x', y', \lambda, t) \;=\; T_s(\lambda)\, f_4(x', y', \lambda, t) \quad . \tag{9}$$

If the spectral sensitivity changes in space also, then this dependence can be made explicit by denoting the sensitivity by $T_s(\lambda, x', y')$. Time t can also be included as another variable of the sensitivity function if necessary.

5.5.2 Transduction

Output of the photosensitive detectors on the image detector depends on the total light energy incident on the detectors. Therefore, the light energy has to be integrated with respect to wave length. This step removes the dependence of the image signal on λ. The relation between the input $f_5(x', y', \lambda, t)$ and output $f_6(x', y', t)$ is

$$f_6(x', y', t) = \int_{-\infty}^{\infty} f_5(x', y', \lambda, t)\, d\lambda \quad . \tag{10}$$

5.5.3 Exposure Time

The image sensor is exposed to incident light for a finite duration of time. During the period when the sensor is exposed, the strength of the incident light may vary because of the changing area of the aperture stop with time (as in the case of shutter in photographic cameras). The area of the aperture stop as a function of time can be denoted by $T_{AS}(t)$. This exposure function remains almost unchanged each time the shutter is opened to record a new picture. Therefore, the effect of the exposure function is to integrate in time the input $f_6(x', y', t)$ weighted by $T_{AS}(t)$ to give the output $f_7(x', y', t)$, i.e.

$$f_7(x', y', t) = \int_{-\infty}^{\infty} f_6(x', y', \tau)\, T_{AS}(\tau - t)\, d\tau \quad . \tag{11}$$

If $T_{AS}(t) = T_{AS}(-t)$ (i.e., it is symmetric), then the above operation becomes a convolution operation. In a CCD camera, typically there is no physical shutter, but the equivalent efffect is achieved by periodically measuring the charge collected by the CCD elements and then clearing the charge in the elements. In most cases, the exposure function can be approximated by a rectangular form and is given by $T_{AS}(t) = \text{rect}(\frac{t}{T})$ where T is the duration of exposure. Typical duration of exposure is 1/30th of a second.

5.5.4 Integration (summation) Over Space

The photosensitive elements on the image detector have a very small but finite surface area. These elements produce an output which depends on the total light energy incident on their surface. The shape and size of the surface area of each photosensitive element is usually the same. Let $R(x,y)$ be a function whose value is 1 inside the surface area of a photosensitive element centered at the origin and zero outside. The photosensitivity of an element is usually uniform at each point on its surface. Therefore, the effect of collecting light energy over a finite region can be specified by the following relation between the input $f_7(x',y',t)$ and the output $f_8(x',y',t)$:

$$f_8(x',y',t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_7(\alpha,\beta,t)\, R(\alpha - x', \beta - y')\, d\alpha\, d\beta \quad . \tag{12}$$

If $R(x,y)$ is symmetric, i.e. $R(x,y) = R(-x,-y)$, then $R(\alpha - x', \beta - y') = R(x' - \alpha, y' - \beta)$. Then the above equation specifies a convolution operation. As a particular example, consider a CCD array having the geometry shown in Fig. 9. In this case,

$$R(x,y) = \text{rect}\left(\frac{x}{b}, \frac{y}{d}\right) \quad . \tag{13}$$

Therefore, for any pixel with indices (m,n)

$$\begin{aligned} f_8(mx_s, ny_s) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_7(\alpha,\beta,t)\, \text{rect}\left(\frac{\alpha - mx_s}{b}, \frac{\beta - ny_s}{d}\right) d\alpha\, d\beta \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_7(\alpha,\beta,t)\, \text{rect}\left(\frac{mx_s - \alpha}{b}, \frac{ny_s - \beta}{d}\right) d\alpha\, d\beta \\ &= f_7(x,y,t) \star \text{rect}\left(\frac{x}{b}, \frac{y}{d}\right) \end{aligned} \tag{14}$$

where $\star$ denotes the convolution operation.

5.5.5 CCD Sensor Noise

In practice the output of the photosensitive elements includes a noise component $n(x',y',t)$. For CCD sensors, this can be modeled as electronic additive noise. Therefore, the input $f_8(x',y',t)$ at this stage is changed to $f_9(x',y',t)$ where

$$f_9(x',y',t) = f_8(x',y',t) + n(x',y',t) \quad . \tag{15}$$

5.5.6 Sampling in Time and Space

We will assume that the continuous signal $f_9(x',y',t)$ is sampled in time periodically at fixed intervals of τ_{s1} and sampled in space on a discrete rectangular grid of points seperated by a distance of x_s along the horizontal direction and by a distance of y_s along the vertical direction (Fig. 9). The resulting output function is the summation of Dirac delta functions with different strengths. The output function $f_{10}(x',y',t)$ is

$$\begin{aligned} f_{10}(x',y',t) &= f_9(x',y',t)\, \frac{1}{|\tau_{s1}|}\text{comb}\left(\frac{t}{\tau_{s1}}\right) \\ &\qquad \frac{1}{|x_s|}\text{comb}\left(\frac{x}{x_s}\right) \cdot \frac{1}{|y_s|}\text{comb}\left(\frac{y}{y_s}\right) \qquad (16) \\ &= f_9(x',y',t)\, \frac{1}{|\tau_{s1}||x_s||y_s|}\text{comb}\left(\frac{x}{x_s}, \frac{y}{y_s}, \frac{t}{\tau_{s1}}\right) . \qquad (17) \end{aligned}$$

The function $f_{10}(x', y', t)$ is actually a continuous function. This function is converted to a discrete function represented as a three-dimensional matrix $f_{11}[i, j, k]$ where

$$\begin{aligned} f_{11}[i,j,k] &= f_9(i \cdot x_s, j \cdot y_s, k \cdot \tau_{s1}) \\ &= \int_{i \cdot x_s{}^-}^{i \cdot x_s{}^+} \int_{j \cdot y_s{}^-}^{j \cdot y_s{}^+} \int_{k \cdot \tau_{s1}{}^-}^{k \cdot \tau_{s1}{}^+} f_{10}(x', y', t)\, dx'\, dy'\, dt \end{aligned} \tag{18}$$

for $i = 0, 1, 2, ..., M-1$, $j = 0, 1, 2, ..., N-1$, $k = 0, 1, 2, ..., K-1$, where M is the number of columns, N is the number of rows, and K is the number of image frames.

5.5.7 Sensor response

Ideally, we require the output of the photosensitive elements to be proportional to the light energy incident on them. However, in practice, their output is not proportional but some other function of light energy. If $f_{11}[i, j, k]$ is the light energy, then the corresponding output $f_{12}[i, j, k]$ is given by

$$f_{12}[i,j,k] = s\ (\ f_{11}[i,j,k]\ ) \tag{19}$$

where $s(I)$ is called the sensor response function. For a CCD sensor, several models have been proposed for $s(I)$ based on experimental observations. Some examples are listed below:

(i) Exponential Model (the *gamma* model)

$$s(I) = I^{\ \gamma}\ . \tag{20}$$

For standard NTSC TV, $\gamma \approx 1/2.2$ [10].

(ii) Linear Model

$$s(I) = aI + b \qquad a, b \text{ are constants.} \tag{21}$$

(iii) $s(I)$ can also be specified by a dense table of values. It can also be specified by a sparse table of values with an interpolation scheme to determine the intermidiate values.

If the sensor response function is different for different photosensitive elements, then s becomes a function of these variables instead of one variable as $s(i, j, I)$. Then the relation between f_{12} and f_{11} is given by

$$f_{12}[i,j,k] = s\ (\ i, j, f_{11}[i,j,k]\ )\ . \tag{22}$$

5.5.8 Conversion of an array of numbers to a one-dimensional sequence of numbers

The discrete function $f_{12}[i, j, k]$ specifies the value of an image pixel in the k^{th} image frame, j^{th} row, and i^{th} column. Usually (to reduce image flicker observed on a monitor) alternate rows are scanned in sequence (*interlaced mode*, e.g. all odd rows first, and then all even rows), the numbers in the rows are converted to analog signal and transmitted over a cabel. Between each row, a horizontal synchronization pulse is added when the scanning shifts from odd rows to even rows and vice versa. However, in our model, for simplicity, we shall assume that all rows are scanned in sequence one by one. We shall also ignore the horizontal and vertical synchronization pulses that are required in practice. With these assumptions, we can use the following transformation to map the 3D array $f_{12}[i, j, k]$ to a one dimensional sequence $f_{13}[l]$.

$$f_{13}[i + j \cdot M + k \cdot M \cdot N] = f_{12}[i,j,k] \tag{23}$$

where M is the number of columns per image frame, and N is the number of rows per image frame. In a CCD array, a set of horizontal and vertical charge transport registers are used to convert an array of 2D image data into a sequence of 1D data .

5.5.9 Reconstruction / Interpolation Filter

The discrete sequence of numbers represented by $f_{13}[i]$ is converted to an analog signal using a sample-and-hold circuit. The time interval τ_{s2} between two numbers is an input to the circuit. Note that τ_{s2} should be consistent with the frame period τ_{s1}, i.e.

$$\tau_{s2} \leq \frac{\tau_{s1}}{M \cdot N} \quad . \tag{24}$$

Mathematically, the operation of the reconstruction circuit can be thought of as interpolating the sequence of values $f_{13}[i]$ defined at $i \cdot \tau_{s2}$ to obtain a continuous signal $f_{15}(t)$ as output. This can be described in two steps as

$$f_{14}(t) = \sum_{l=0}^{K \cdot M \cdot N-1} f_{13}[l] \, \delta(t - l \cdot \tau_{s2}) \, , \tag{25}$$

$$f_{15}(t) = h_{sh}(t) \, \star \, f_{15}(t) \tag{26}$$

where $h_{sh}(t)$ is the effective impulse response function of the sample-and-hold interpolation circuit and $\star$ is the convolution operator.

5.5.10 Amplifier

$f_{15}(t)$ is the video signal. Before this is transmitted over a video cable, it is amplified so that any noise introduced in the cable will not dominate the signal. If $h_a(t)$ is the impulse response of the amplifier, then the output of the amplifier can be modeled as

$$f_{16}(t) = f_{15}(t) \star h_a(t) + n_a(t) \tag{27}$$

where $n_a(t)$ represents additive amplifier noise, and $\star$ is the convolution operator. $f_{16}(t)$ is the output of the camera system which is transmitted over a video cable.

5.5.11 Cable

The relation between the input signal $f_{16}(t)$ at one end of the video cable and the output $f_{17}(t)$ at the other end of the cable can be specified as

$$f_{17}(t) = f_{16}(t) \star h_c(t) + n_c(t) \tag{28}$$

where $\star$ is the convolution operator, $h_c(t)$ is the impulse response function of the cable, and $n_c(t)$ is the additive cable noise.

5.5.12 Sample and Hold Circuit

The analog signal $f_{17}(t)$ is sampled at intervals of τ_{s3} by multiplying it by (see page 115,Oppenheim and Schafer [7])

$$\frac{1}{\tau_{s3}} \mathrm{comb}\left(\frac{t}{\tau_{s3}}\right) = \sum_{n=-\infty}^{\infty} \delta(t - n\tau_{s3}) \quad . \tag{29}$$

If $\tau_{s3} \neq \tau_{s2}$, this causes geometric distortion of the picture. This phenomenon has been called *mismatched electronics* in Schafer[9]. The sampled values are interpolated by an n-th order (usually 0^{th} order) sample-and-hold filter. The resulting signal is again sampled by a slightly shifted sampling function. If we assume that the sampling occurs at the middle of the interval, the effect of sample-and-hold circuit on the input can be summarized as below:

$$f_{17a}(t) = \frac{1}{\tau_{s3}} \mathrm{comb}\left(\frac{t}{\tau_{s3}}\right) \cdot f_{17}(t) \tag{30}$$

$$f_{17b}(t) = f_{17a}(t) \, \star \, h_f(t) \, + \, n_f(t) \tag{31}$$

$$f_{18}(t) = \frac{1}{\tau_{s3}} \mathrm{comb}\left(\frac{t - \tau_{s3}/2}{\tau_{s3}}\right) \cdot f_{17b}(t). \tag{32}$$

The output $f_{18}(t)$ is an impulse train. This impulse train is converted to a sequence of numbers $f_{19}[i]$ where

$$f_{19}[i] = \int_{(i+1/2)\tau_{s3}^{-}}^{(i+1/2)\tau_{s3}^{+}} f_{18}(t) \, dt \quad . \tag{33}$$

5.5.13 Quantization

The effect of quantization can be summerized by the transformation rule

$$f_{20}[i] = Q(f_{19}[i]) \tag{34}$$

where, after appropriate scaling, $Q(x)$ can be defined as

$$Q(x) = \begin{cases} Q_{max} & \text{for } X > Q_{max} \\ \lfloor x + 0.5 \rfloor & \text{for } 0 \le x \le Q_{max} \\ 0 & \text{otherwise} \end{cases} \tag{35}$$

5.5.14 Digital Image Sequence

The one-dimensional sequence $f_{20}[l]$ is converted to a three-dimensional array representing a sequence of two-dimensional images or image frames using the following rules:

$$i = l\%M, \quad j = ((l-i)/M)\%N, \quad k = (l-i-jM)/(MN), \quad \text{or} \quad k = \lfloor l/(MN) \rfloor, \tag{36}$$

$$f_{21}[i,j,k] \;=\; f_{20}[l], \tag{37}$$

where % denotes the *mod* (i.e. remainder) operator.

6 Conclusion

The camera model described here has been developed to help simulate on a computer the formation and sensing of defocused images. Some additional refinements to the basic model presented here are needed to implement it on a computer. This model can also be extended to address other specific problems.

7 Acknowledgements

We are thankful to Mr. Sanjay Sethi for his help in preparing this report. This research was supported in part by the National Science Foundation (IRI-8821923), and by the Olympus Optical Co.

A Correspondence Problem

In Fig. 5 we show how the motion of the entrance pupil introduces the correspondince problem, and also show how correspondence problem is avoided by keeping the entrance pupil stationary even when the exit pupil is moved.

Let a first image be acquired with a first set of camera parameters. Fig. 5a shows the position of the entrance pupil ENP1, and the exit pupil EXP1 with respect to an external coordinate system X_o, Y_o, Z_o, O_o for the first camera setting. For any point $P_1^{'}$ on the first image, the direction (θ_1, ϕ_1) of the object point P in the scene can be determined. Now, suppose that the camera parameters are changed so that the positions of the entrance pupil and exit pupils are as in Fig. 5b. Here, the position of the entrance pupil ENP_2 remains unmoved with respect to the external coordinate system. In this case, the direction(θ_1, ϕ_1) of the point P remains the same, and the position of P's image, P_1'', in the second image I_2 can be determined from (θ_1, ϕ_1). Consequently, for any point such as P_1' in the first image the position of the corresponding point P_1'' in the second image can be determined, that is, the correspondence between images I_1 and I_2 can be established.

Suppose that the camera parameters are changed so that the entrance pupil moves with respect to the external coordinate system as in Fig. 5c. Then the direction of the point P changes from (θ_1, ϕ_1) in Fig. 5a to (θ_2, ϕ_1) in Fig. 5c. The angle θ_2 cannot be determined even if the displacement $\triangle Z$ of the entrance pupil with respect to the external coordinate system is given because the distance r is not known. Therefore the position of the geometric image P_2' of the point P cannot be determined given the position of P_1' in image I_1, that is, the correspondence between the points P_1' and P_2' cannot be established.

In summary, given the position of a point in one image, the position of the image of the same point in another image cannot be determined if the entrance pupil is at different positions for the two images. This is the correspondence problem.

References

[1] M. Born and E. Wolf, *Principles of Optics*, Pergamon Press, Oxford, Sixth Edition, 1980.

[2] J. D. Gaskill, *Linear Systems, Fourier Transforms, and Optics*, John Wiley and Sons, 1978.

[3] J. W. Goodman, *Introduction to Fourier Optics*, McGraw-Hill Book Co., 1968.

[4] B. K. P. Horn, *Robot Vision,* McGraw-Hill Book Company, 1986.

[5] E. Hecht, *Optics,* Addison-Wesley Publishing Co., 1987.

[6] M. D. Levine, *Vision in Man and Machine* , McGraw-Hill Book Co., 1985.

[7] A. V. Oppenheim, and R. W. Schafer *Discrete-Time Signal Processing* Printice Hall, 1989.

[8] A. Rosenfeld, and A. C. Kak, *Digital Picture Processing,* Vol. I . Academic Press, 1982.

[9] S. A. Shafer, Automation and Calibration for Robot Vision Systems, Technical Report No. CMU-CS-88-147, Computer Science Department, Carnegie Mellon University, 1988.

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

[11] M. Subbarao, Parallel depth recovery by changing camera parameters, *Proceedings of Second International Conference on Computer Vision,* Dec. 1989, Florida, USA, pp. 149-155.

[12] M. Subbarao, Efficient depth recovery through inverse optics. Editor: H. Freeman, *Machine Vision for Inspection and Measurement,* Academic Press, Boston, 1988, pp. 101-126.

[13] M. Subbarao, Determining distance from defocused images of simple objects, Technical Report No. 89.07.20, Computer Vision Laboratory, Department of Electrical Engineering, State University of New York, Stony Brook, NY 11794-2350.