January 1997

The Effect of Radiometric Correction on Multicamera Algorithms

Gerda Kamberova
University of Pennsylvania

Ruzena Bajcsy
University of Pennsylvania

University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-97-21.

The Effect of Radiometric Correction on Multicamera Algorithms

Abstract

We present results confirming the importance of radiometric correction in multicamera applications. Although, we compensate for systematic noise only, we review all noise sources in the video sensor (systematic and random). We use a simple model for radiometric correction of digital images. The correction procedure is tested on the disparity map computation in stereo matching, particularly in a case where stereo usually fails -- almost textureless white surface. Without correcting radiometricly, the matching algorithm matches systematic noise components in the two images. With the correction, after removing the systematic noise, an improvement of 26% to 59% in relative rms of the disparity map is demonstrated (the higher the intensity of the flat field, the better the improvement).

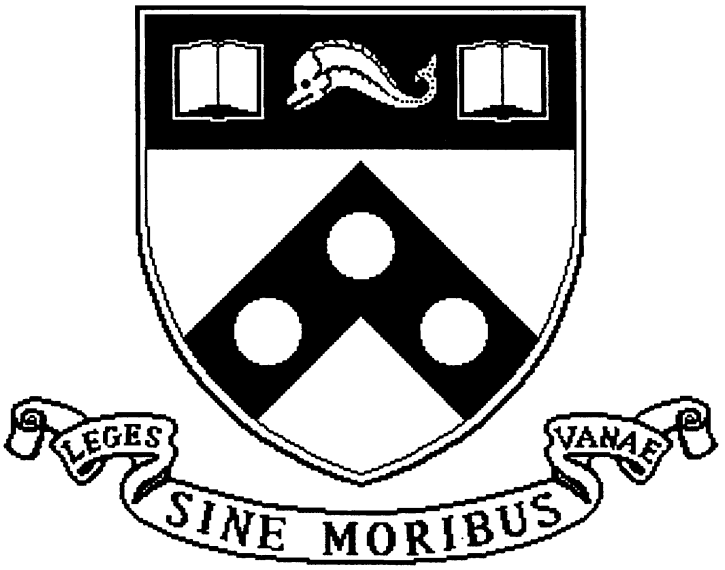
Comments

University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-97-21.

The Effect of Radiometric Correction on Multicamera Algorithms

MS-CIS-97-21 (GRASP LAB 418)

Gerda Kamberova, Ruzena Bajcsy



University of Pennsylvania
School of Engineering and Applied Science
Computer and Information Science Department
Philadelphia, PA 19104-6389

1997

The Effect of Radiometric Correction on Multicamera Algorithms

Gerda Kamberova and Ruzena Bajcsy
E-mail: kamberov@cis.upenn.edu, bajcsy@cis.upenn.edu
Department of Computer and Information Science
University of Pennsylvania
Philadelphia, PA 19104

Abstract

We present preliminary results confirming the importance of radiometric correction in multicamera applications. Although, in this paper we compensate for systematic noise only, we review all noise sources in the video sensor (systematic and random). We use a simple model for radiometric correction of digital images. The correction procedure is tested on the disparity map computation in stereo matching, particularly in a case where stereo usually fails — almost textureless white surface. Without correcting radiometricly, the matching algorithm matches systematic noise components in the two images. With the correction, after removing the systematic noise, an improvement of 26% to 59% in relative rms of the disparity map is demonstrated (the higher the intensity of the flat field, the better the improvement).

1 Introduction

When multiple sensors are used in applications requiring hard performance guarantees, correcting for errors and obtaining objective confidence measures for the uncertainty of the results cannot be neglected. Empirical tests, validation, and analysis of robustness of existing systems are necessary [7, 6, 21]. The use of vision algorithms with physically different sensors, necessitates the evaluation of the performance of the algorithms given the parameters of the sensors. In this paper we study the video sensor. Our short term goal is to understand the important parameters of the video sensor, and to show how this knowledge can be used for improving the performance of vision algorithms in certain multicamera applications. We want to identify the minimal set of experiments and calibration steps one should use in order to "equalize" jointly multiple video sensors. Our objective is the correction of systematic errors which originate in the sensor. Our **final goal** (and object of current research) is to present a full sensor model, digital image model together with a camera error measurement model. We will use this model, together with an environment model, in performance evaluation and characterization of subsequent vision algorithms. In particular, we are interested in stereo reconstruction. We will use the models in a minimax confidence set approach to disparity estimation [13].

First, we review briefly the principle of operation of the video sensor: later we discuss different sources of errors originating in it. The video sensor consists of a lens, a CCD camera, and a framegrabber connected to a computer. The sensor "looks" at the world, and records an image of the observed scene. The input to the sensor is visible light, and the output is a digital image which is stored in a computer memory. The incoming light from the scene is collected by the lens, and directed onto the photosensitive array of identical Sensing ELements (*sels* [14]). The imaging process with a CCD camera is based on the physical principle of converting photons into a measurable quantity (voltage). The photons generate charges. For each sel, the charges generated during the exposure time are collected into a charge packet. When suitable clock voltages are applied potential wells are created which store and move the charge packets. During transport, first, the charge packets are transferred, in parallel, along vertical CCDs

toward a *horizontal CCD shift register*. From this register, the charge packets are shifted to the output stage where they are measured (converted to voltages) and amplified. The image array content is transferred line-by-line to the horizontal shift register, and from there, a line is pumped out, sel-by-sel, at time periods set by the *camera pixel clock*. The camera output is an analog signal. The framegrabber reads in the analog signal, usually in *composite video* format[1]. This format consists of image content periods (corresponding to the content of a single line) separated by *horizontal synchronization pulses* (HSYNC). A single image is called a *frame*. It consists of two fields – one of all odd number lines and one of all even number lines. The transmission of the image from camera to frame grabber is in fields. The fields are separated by *vertical synchronization pulses* (VSYNC). Two consecutive lines in a digital image are actually temporally spaced – the image is *interlaced*[2]. The framegrabber samples the analog signal at a *sampling frequency* set by its A/D converter. The sampled voltage levels are converted to integer gray values which range depends on the bit depth of the digitizer. The framegrabber reassembles the 1D signal into a 2D digital image[3]. Following again [14], we call an individual Picture ELement of the digital image a *pel*. In the text, we use *pixel* to denote sel or pel, which one it will be clear from the context.

Geometric and radiometric uncertainties and discrepancies in the digital image are due to the optics, the CCD camera, the joint operation of the camera, digitizer and other electronic components, and to the discretization process. Distortions related to optics have been analyzed extensively [23]: discretization effects are in the center of the signal processing research [10]. Both of these are important, but out of the scope of our paper. We focus only on noise sources originating in the CCD camera and the framegrabber, except spatial quantization error. An indepth treatment of the video signal formats can be found in [5, 20]. A short and clear introduction is given in [14, 2]. Charge-coupled devices and solid state imaging arrays are analyzed by designers. Test procedures for their evaluation with the use of precise measuring equipment is discussed in [4, 26, 8, 11, 12]. Various sources of geometric distortions in the video sensor are analyzed in [14] where the term videometrics is coined. A detailed study of different framegrabber architectures and synchronization mechanisms, and related geometric and radiometric distortions is given in [3]. A comprehensive discussion of linejitter problem, its sources and detection, is given in [14, 2, 3]. In [15] a method for estimating linejitter based on Fourier analysis is presented, it is appropriate for cases in which it is known that the jitter is no more than a pixel. The computer vision community has long ago recognized the importance of the geometric calibration in accounting for geometric distortions, but work on radiometric calibration and noise models is limited. Radiometric noise models are discussed in [9, 24]. The model in [9] is very good, detailed — capturing various random noise sources. The actual procedure for radiometric correction is based on a simplified model (many assumptions are made along the way of deriving it). The simplified model falls in the class of linear models plus additive noise. The model of [24] is very attractive with its simplicity, the noise is divided into signal-dependent, signal-independent/position-dependent, and singal- and position-independent components (it is used in the context of Bayesian framework for image restoration). Most of the work in sensor noise modeling on practice is done with Gaussian approximations, linearization of nonliner functions, and rms estimation. In our future work we address a sensor model faithful to the physics of the sensor. In this paper, we report ongoing work, and directions for future research. We bring the attention of the computer vision practitioners to the importance of understanding the sensor, and dealing with all sources of errors (systematic and random), we illustrate these errors with examples.

[1] NTSC standard used in North America and Japan, 525/59.94/2:1; PAL standard used in Europe and Asia, 625/50/2:1.

[2] Other modes of transmission are possible, but we mention only interlacing since it is most common for the applications we looked at – it delivers high resolution and signal level, but it imposes restrictions on the speed of moving objects that can be faithfully imaged.

[3] The reasons for the conversion from 2D representation in the CCD array to 1D camera output (and then again to the desired 2D digital image format) is for historical reasons, compatibility with existing standards requirements, and, up to very recently, limitation of the technology.

We emphasize that an assumption about "uniformity" of the sensors in multicamera applications is not valid. In particular, if an application requires high accuracy and performance guarantees, the sensor has to be calibrated (geometrically and radiometricly).

Our paper is organized as follows. In Section 2 we review main parameters and sources of error of the video sensor. Systematic and random errors are illustrated in dark images. Spatial nonuniformities of response to a constant stimuli are illustrated in flat fields. In Section 3 we present a digital image model and a procedure for radiometric correction based on it. We use this procedure to "equalize" different sensors. In Section 4 we use the problem of a disparity map computation in stereo as an example of multicamera application. Test results are presented, in particular, in cases when the stereo usually fails — almost textureless card board (without the radiometric correction, noise processes are matched instead of signal). The results demonstrate a nontrivial improvement in the disparity map computation after the correction. In the last section conclusions are drawn and future research directions outlined.

2 Parameterization of the video sensor

More often than suspected video sensors are used without understanding of their internal limitations and operational parameters, and the impact these have on subsequent algorithms. We review parameters characterizing the video sensor (camera and framegrabber/digitizer) and limit its performance. Any undesired features which cause discrepancies in the output signal we consider *noise*. The noise may be deterministic, systematic, or random. We focus on systematic and random noise which originate in the CCD camera and the framegrabber. We want to study **all noise components**. This paper shows that not only random but also **systematic errors** have to be accounted for.

2.1 Camera related parameters and noise

The individual sels comprising the CCD array have their own physical characteristics. The *quantum efficiency* (QE) characterizes the physics of the charge accumulation process. It is the ratio of collected electrons to incident photons. The QE is dependent on the technology used and the wavelength of the light. A related parameter is the spectral responsitivity of the sels (the response of the sel to light stimuli varies with the wavelength). The difference in the spectral responsitivity of the sels is manifested in the *fixed pattern noise* in *photoresponse nonuniformity* (PRNU) of the whole array to scenes of uniform brightness.

After the charge packets are formed in the sels, they have to be transferred. A basic limitation of the performance of the CCD is the efficiency with which a charge packet can be transferred from one sel to the next. A charge packet traveling through the CCD array picks up additional charges (i.e., thermally generated or diffused); it also looses charges (due to trapping by surface states)[4]. A *charge transfer efficiency* (CTE) of 0.9999995 is reported [12]. For computer vision applications CTE is usually assumed to be 1.

The *spatial* resolution is the ability of the sensor to discriminate between closely spaced points in the image, and the *temporal* resolution the ability to resolve temporal variations in the incoming signal. The spatial resolution depends on the geometry of the imaging array (number of sels, their size and shape, and organization in the imaging array). The minimal time necessary for the CCD to collect enough charges so that the signal raises above the *read noise*[5] puts a bound on the temporal resolution.

The *dynamic range* is a derived parameter defined as the ratio of the full-well capacity (saturation level) to read noise. It specifies the range between the brightest and the darkest levels within a single

[4] A way of minimizing the loss due to trapping is to keep the potential wells always semi-empty so charges from passing packets will not be trapped (*fat zero* is introduced).

[5] Read noise is reviewed later in this section.

image. The saturation level dependent of the CCD technology (material properties and pixel size). The number of the intensity levels in the dynamic range gives *intensity resolution* of the camera.

The *total random noise*[6] of the CCD has three major components: photon (shot) noise, read noise. and fixed pattern noise. The total noise is dominated by the read noise at low illumination levels, by the fixed pattern noise at high levels, and by the photon shot noise in between [11].

The, *photon (shot) noise* is related to the natural variation of the incident photon flux. The total number of photons emitted by a steady light source over time interval varies according to Poisson distribution. Any shot noise limited process is well modeled by Poisson distribution, that distribution is used to describe other noise components related to the CCD[7]. For Poisson distribution the mean equals the variance, so the root mean square error (rms) of the photon shot noise in electrons is equal to the square root of the mean signal. The shot noise is always present in the data. accounting for systematic errors does not remove it.

The second component of the total noise is the *read noise.* It is one of the most important characteristics of the sensitivity of the CCD. It defines the level of the lowest detectable signal. Factors contributing to the read noise are: *background noise*, *trapping noise*. *reset noise*, *charge transfer noise*, and *output amplifier noise*[8]. The output amplifier noise is the definite lower bound on the read noise. Background noise and trapping noise are primary technology related. while reset noise and amplifier noise are mostly related to signal processing of the CCD output [26]. First we discuss the background noise.The *background noise* has three main components: dark current. optical or electronic fat zero, and luminescence in the device [11].

Dark current consists of thermally generated charges caused by motions of the semiconductor atoms, even at normal room temperature. The generation of dark current is a random process modeled by Poisson distribution. Dark current is added to the charge packets and cannot be distinguished from the photo-generated charges. In general. dark current doubles with every 8° increase in temperature [4], to reduce dark current cameras are cooled.

Fat zero is introduced to aid the charge transfer efficiency and the consistency of the quantum efficiency. Optically generated fat zero follows Poisson distribution. If the fat zero is electrically introduced, on input, it is less than shot noise.

The *internal luminance* in the CCD device may have variety of sources [11]. One source is the clocking of the voltages to the *gates* which control the potential well levels. It is manifested in an exponential decline in average line intensity (see our examples of dark images.) The phenomena is explained by generation of long wave photons by the clocking of the register, which photons get absorbed in the lines close to the register, thus increasing the background charges [11]. Statistically this noise is modeled by Poisson distribution. Second source of luminance is diffusion. It is related to input-output mechanism. It explains the "radiation" of light in the CCD dark image from the position of the output amplifier. Another damaging source of luminance are blemishes. These are single defective sels, "hot pixels", that get saturated fast.

Trapping noise is caused by random variations in the "trapping" states of CCD: charges get trapped, and are kept trapped for some random period. For buried-channel CCD this noise is very low, on the order of 5 electrons [17][9].

The *amplifier noise*, due to the on chip amplifier[10], is associated entirely with the output stage. It may have two components: a white noise (due to thermal generation) and a component introduced by interaction of charges and "traps" present in the transistor channel. By good manufacturing, this noise

[6] In the literature often the term noise is used to mean the amount of noise itself *and* the variance/rms of the noise.

[7] An example of the derivation of the distribution in nuclear physics context could be found in [1].

[8] In the literature, there is no agreement on these names. We follow definitions from [11], [4], [17], and [19].

[9] One should keep in mind that in astronomy applications, where very low level signals have to be detected, high performance digital cooled cameras are usually used. We assume that the numbers cited are for digital cameras.

[10] The amplification after the CCD also introduces amplifier noise.

can be reduced substantially [17].

When a charge packet arrives at the output node. it produces a voltage change. To measure the voltage of a charge packet a reference voltage level is needed. The readout capacitors are reset to a nominal voltage level at each readout cycle. The *reset noise* relates to the uncertainty this voltage level. This noise is effectively removed by *correlated double sampling* [17].

The third component of the total noise is the *fixed pattern noise.* Dark current can originate at different locations in the CCD but has, in all cases, to do with the irregularities in the crystal structure of the silicon. This gives rise to a systematic noise called *fixed pattern noise* (FPN) in dark current ("stuck" sels which have constant, fixed output contribute to FPN too). In this paper, we deal with systematic, FPN, in dark current or photoresponse nonumiformity.

In the following subsections, we illustrate some noise components, in particular the background noise, the fixed patter noise in dark current, and the photoresponse nonuniformity in flat fields.

Remark 2.1 All intensity images (except Figure 7 (page 13)) are scaled to enhance the display.

In the following table, we give some specification for the cameras and framegrabbers we use; we use only single channel (black and white) 8 bit images.

Camera	CCD array size	Pixel clock frequency
`HITACH KP 230(231)`	384(H)x485(V)	8MHz
`SONY XC-77`	768(H)X493(V)	14.318MHz
Framegrabber	Effective digital image size	Sampling frequency
DT1451	512(H)x480(V)	about 10MHz
S1V	640(H)x480(V)	not available
TIM40	756(H)x480(V)	set to 14.318MHz

2.2 Dark images and background noise

Manifestation of the background noise can be observed in the *dark images.* A dark image is an image taken with no access of light to the video sensor. Dark images are taken with a tight, opaque cap on the lens.

Before a camera is used for quantitative measurements, it must be warmed up. A simple procedure for determining the minimal warm up time is to observe, **once**, starting with cold camera, a sequence, over time, of average spatial intensities of dark images[11]. Initially the sequence of averaged intensities exhibits nonmonotone behavior (increasing then decreasing), but eventually it stabilizes to a constant level, within some nominal noise bounds. The index of the first term in the sequence for which the stabilization occurs is directly related to the minimal amount of time needed to warmup the camera. For the sensors we were working with, a warm up had to be from 20min to almost 2 hours, depending on the physical camera (warm up time vary between same make).

Example 2.1 Figure 1 (page 6) shows a typical dark image for the camera/framegrabber configuration `SONY XC-77/DT1451`. In image A the background random noise and internal luminance are noticeable. In the graphs, notice the decline in the average intensity due to the internal luminance and the DC-restoration problem which originates in the framegrabber [3], see Section 2.4. Systematic component in the background noise is prominent in image B. The vertical stripes are systematic noise due to the mismatch between the camera pixel clock and the ADC sampling frequency, and "cross-talk" from clocking of other electronic components. For comparison, observe the dark image for another camera

[11] In the experiments we used sequences of length 500 to 960. Each element of the sequence was a pair (μ_i, σ_i^2). For fixed i we took 100 dark images, and computed the overall (spatial and temporal) average intensity, μ_i, and variance, σ_i^2, over the 100 images.

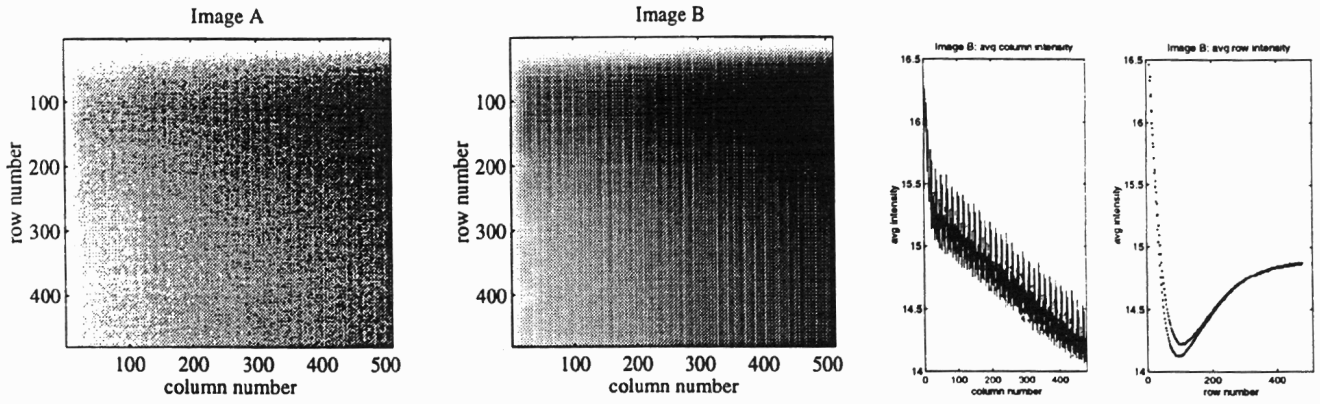


Figure 1: SONY XC-77/DT1451: Image A is a single dark intensity image, image B is a temporal average of 100 dark images. On the right are graphs representing image B – average intensity vs row, and vs column, respectively.

configuration, SONY XC-77/TIM 40, Figure 3 (page 8). There are no high frequency components in that image, the camera pixel clock and sampling frequencies are almost equal (there is one-to-one correspondence between sels and pels).

Example 2.2 Figure 2 (page 7) shows data for 4 different sensor configurations, based on single dark image for each configuration: average column intensity (for a fixed column, on the vertical axes the average intensity over all rows is plotted), and the amplitude of the Fourier power of the that average intensity (to enhance the display, we have plotted the *log* of 1 plus the amplitude of the shifted fft). Refer to the table on page 10 for specifications of the sensor configurations. For configuration SONY XC-77/DT1451 the high frequencies in power spectrum are prominent, the camera line is under-sampled. Contrast this to the configuration, SONY XC-77/TIM40, where the sampling is almost one-to-one. For each configuration, we tested several combinations of physically different cameras and framegrabbers. Each video sensor configuration, camera/framegrabber, had its own dark image signature. The difference between dark image signatures for video sensors within the same configuration were not as dramatic as the differences between dark image signatures for sensors from different configurations. We can use averaged (over time) dark images to characterize systematic components in the background noise for different sensor configurations. Within a particular configuration, the background noise can be parametrized.

Example 2.3 Anther phenomena which can be observed in dark images are residual, "ghost" images and phase patterns. Both are illustrated in Figure 3 (page 8). On the left, the result of averaging 10 **dark** images for SONY XC-77/TIM-40 is presented, on the right, 100 SONY XC-77/DT1451 dark images. A brighter rectangle ("ghost image"), and a wavy phase pattern are noticeable in the SONY XC-77/TIM-40 image. The rectangle results from the scene of a white piece of paper on black background to which the camera had been exposed prior to taking the dark images. For SONY XC-77/DT1451, an image of the target which we use for geometric calibration is visible. In both experiments, the "ghost image" is a result of an artificial increase in the quantum efficiency of some the sels. The ghost image results from charges trapped during use of the sensor prior to taking the dark images. While the charges remain trapped, "the effective QE for subsequent images will be increased because less trapping of newly generated signal electrons will occur" [11]. To clear up such residual images, some cameras have a "clear" command, in our case simply switching off the camera removes the residual image. The origin of the ghost images is the CCD chip (camera). In [3] phase phenomena similar to that in SONY XC-77/TIM-40 is reported, again the source of it is not exactly located. Possibilities are the DC-

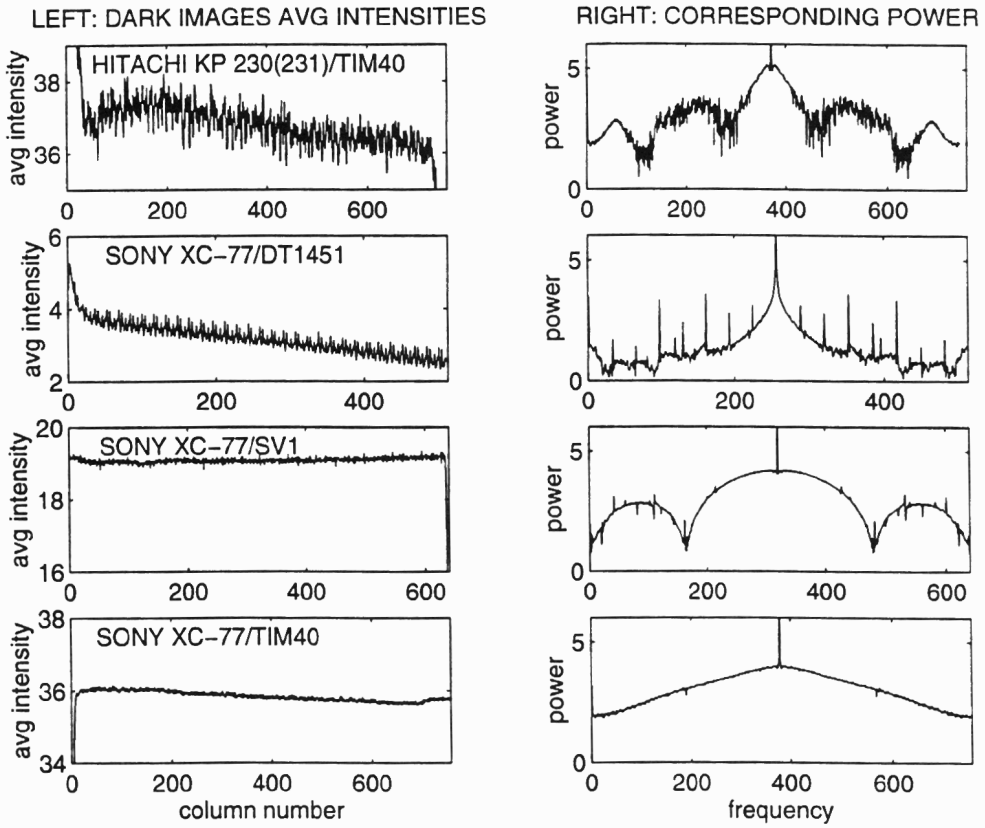


Figure 2: Data for single dark images obtained with different `camera/framegrabber` configurations. The graphs represent average intensities vs column (left), and the corresponding amplitudes of the power spectra (right).

restoration process, linejitter (Section 2.4). Thus this phenomena is related to the camera/framegrabber interface and the digitization.

2.3 Flat fields and photoresponse nonuniformity

We will illustrate the intrinsic nonuniformities of the sels. An image. which is a response of the sensor to a scene of uniform brightness (preferably close to saturation level) is called a *flat field.* It is used in radiometric calibration. Obtaining a flat field in a standard engineering laboratory is a nontrivial task. We investigated different methods of *approximating* a flat field. The flat field approximations reported here were obtained with a diffusing white glass filter instead of the lens.

We present two independent experiments regarding two instances of the configuration `SONY XC-77/DT1451`: the same framegrabber but two physicly different cameras. Figure 4 (page 8) shows the temporal average of 100 flat fields and their histograms for each of the two video sensors tested. The flat field approximations for each sensor were obtained independently, under fixed illumination conditions provided by 3 distributed incandescent light bulbs illuminating a white planar surface. The acquisition of the flat fields without the use of a lens removes any nonuniformity which could have risen from lens vignetting (fall off in intensity from the center of the image to the boundary) or other lens distortions. The bright square frames noticeable in the two intensity images, are due not to reflections from the square aperture in front of the sensor array (a part of the camera architecture). Apart from that artifact, the variation in intensity in the images is mostly due to fixed pattern noise of the photoresponse nonuniformities of

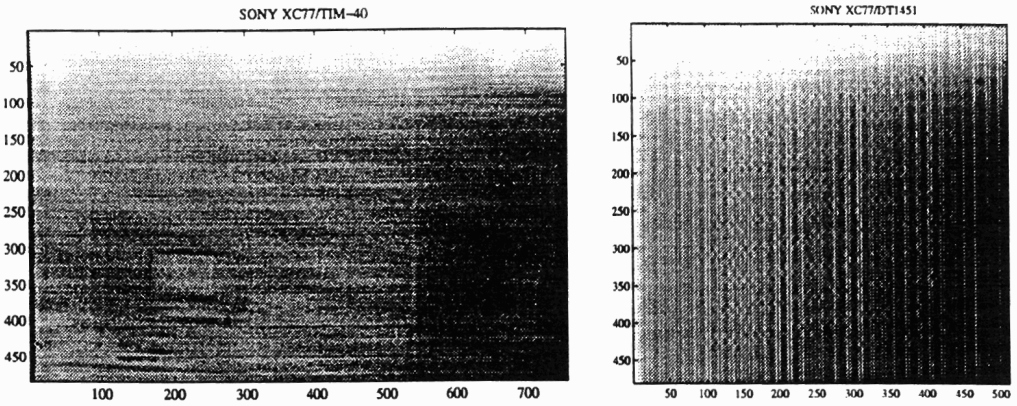


Figure 3: Ghost images: For SONY XC-77/TIM-40 (left), the average of 10 darkimages; visible are a residual image of a rectangle and a phase pattern. For SONY XC-77/DT1451 (right), the average of 100 dark images; visible are a checker-board pattern and vertical stripes.

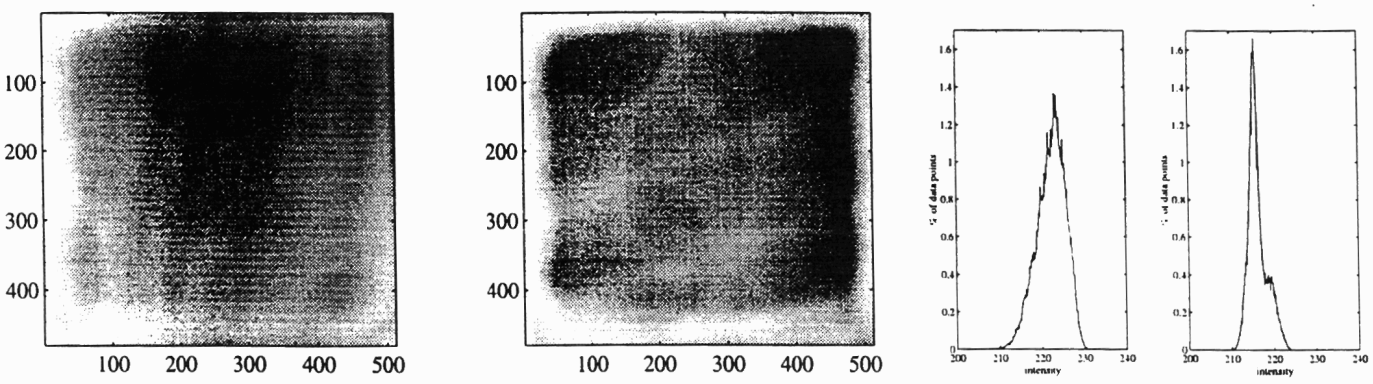


Figure 4: SONY XC-77/DT1451: averaged flat field approximations and their histograms.

the individual sel (the read noise is negligible at that high level, a gray value of about 220)[12].

2.4 Framegrabber related noise

There are different types of geometric discrepancies due to the ADC in the framegrabber.

When the individual lines in the digitized image are not aligned properly, the discrepancy is called *linejitter* (it is a result of the failure of the framegrabber to detect the HSYNC). When the digitizer under-samples or over-samples the "locked" line (the sampling frequency is different from the camera pixel clock frequency), a scaling factor (different than one) between pixel size and sel size is introduced. This parameter, *horizontal scale*, is one of the objects of geometric calibration. We have also observed a constant shift within a line for a fixed frame (up to 15 pixels for some sensor configurations). This shift varies ± a pixel for different frames. Thus (as with linejitter) extra caution has to be paid when using temporal pixelwise averages of images, so that data related to different pixels are not mixed. A geometric distortion in the digital image which relates to VSYNC detection is that the top, up to 100, lines in the image have relatively higher jitter.

An effect in radiometric distortion from interlacing is the shift in gray level between odd and even fields. This is clearly observed in Figure 1 (page 6), in the most right graph. The appearance of two curves is due to the intensity difference between odd and even fields. In the process of digitization, the digitizer has to be able to restore the zero reference level from which to measure (this process is called

[12]The printing of the images has introduced minor defects as well.

DC-restoration); failure to detect the reference level correctly results in shift in the gray values, and thus radiometric distortions. "The fall-off of the sample-and-hold mechanism used in many DC-restoration circuits" leads to uniform (for all images with this configuration) component in the background and could be easily removed [3].

One of the most noticeable manifestations of distortions due to the camera/framegrabber interface is a systematic error component in the background noise of the dark images, as discussed earlier, due primary to the mismatch between the frequency of the camera pixel clock and the digitizer sampling frequency.

The effect of discretization is a main objective in the area of signal processing. Severe distortions (*Moiré-effect*) occur when the sampling theorem[13] is violated. The sel spacing puts limitation on the highest frequency in the input beyond which severe distortions occur, and the area of the CCD chip limits the lowest frequencies which can be detected.

Another type of radiometric uncertainty due discretization is captured by the *quantization error*. It results from the analog to digital conversion of the signal. There are different conversion schemes: with uniform or nonuniform quantization step. The quantization error is modeled as an additive absolutely continuous random noise. Its distribution depends on the distribution of the input signal (output voltage from the camera). If the input signal satisfies the *quantization theorem* (or its extension)[14], the quantization error for round off quantizer with uniform step q, is uniformly distributed in the interval $[-\frac{1}{2}q, \frac{1}{2}q]$, [25]. Under this model the variance of the quantization error is $q^2/12$. The value of q depends on the dynamic range (full well capacity of the potential wells) of the CCD camera and the bit depth of the digitizer. When the distribution of the input signal is Gaussian, the necessary and sufficient condition of the quantization theorem is not satisfied, so the very popular assumption about the uniformity of the quantization error given Gaussian input, is not valid; depending on the variance of the input, the uniform model for the quantization error could be close or very far from the true distribution of the quantization error [25].

3 The model and the radiometric correction procedure

The variation of intensity in dark images and in flat fields for physically different sensors clearly shows that sensors are not "equal", contrary to the assumptions made in many multisensor computer vision applications. Given, even, absolutely the same scene and illumination, due to all the factors we discussed in Section 2, physically different sensors "see" differently.

There are different radiometric correction procedures in use to adjust pixel values in digital images. They all assume a linear model for the pixel response[15]. The value, p, at a given pixel is changed to the value pc, $pc = (p - offset) \times gain$, where $gain$ and $offset$ are pixel dependent parameters. For an ideal camera the gain should be 1 and the offset 0 over the whole array. For real cameras, these parameters vary spatially. The radiometric correction methods differ in the way the parameters $gain$ and $offset$ are estimated, but in all cases, they are based on two sets of images: dark images and flat fields.

The radiometric correction procedure of [9], despite of the very complete and detailed model, falls actually in the class of the simple procedures we discuss. From the experiments conducted in [3], it is clear that there are no theoretical reasons for choosing one of the presented radiometric correction methods over the other, but in both cases, the two radiometric correction methods improve the uniformity of the pixel response over the CCD array and have effect on detection algorithms (detecting dots, lines

[13]The input signal can be fully reconstructed from the samples if the input signal frequency, f, is at most half of the the sampling frequency.

[14]The characteristic function of the absolutely continuous input random variable is 0 at multiples of $2\pi/q$ iff the quantization error is uniformly distributed in $[-\frac{1}{2}q, \frac{1}{2}q]$.

[15]CCD cameras have good linearity, but even if no gamma correction or extra nonlinear processing is done in the sensor, the amplifier introduces non-linearity [3].

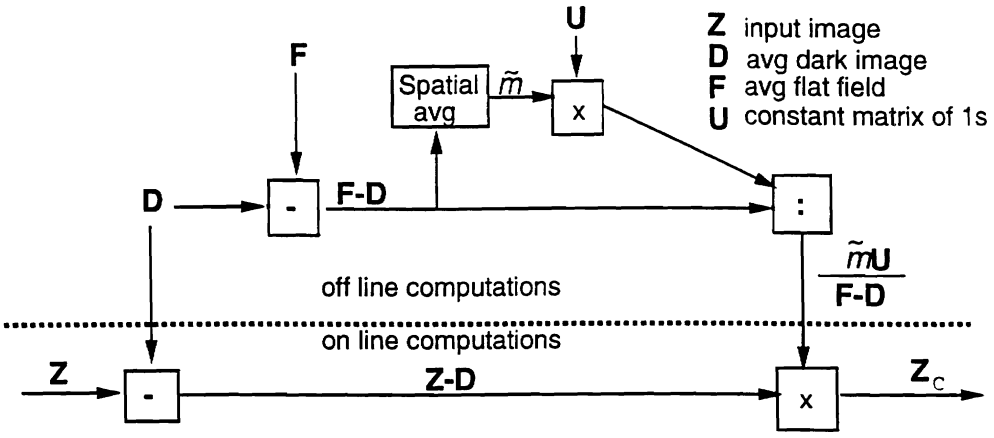


Figure 5: Input-output diagram for the radiometric correction procedure. The operations subtraction, $-$, multiplication, $\times$, and division, :, are pixelwise; $\tilde{m}$ is a scalar.

or edges, [3]) and matching algorithms (as we illustrate). Such procedures are crucial in applications which rely on intensity, or in stereo when the texture signal is not strong enough (as illustrated later).

In the following section we give the radiometric correction procedure, the significance of which in disparity map computation (an example of multicamera algorithm) we studied. We not only use the correction procedures, but also provide the **theoretical justification** for it.

3.1 The radiometric correction procedure

Individual pixel values are corrected for nonuniformities due to systematic errors from fixed pattern noise in dark current and photoresponse nonuniformities. The process is called *flat fielding* or *shading correction* [19]. Intuitively, in order to achieve a corrected image with zero offset and constant gain. we removed from the original image the offset observed in dark images. and scale the difference it inversely proportional to the photoresponse. The procedure is schematicly represented on figure 5 (page 10). Do the following steps **off line**: (1) Take a sequence of dark images and compute the average dark image D of the sequence: a pixel (u, v) in D is assigned the average intensity of the pixel (u, v) over all dark images. Store D. (2) Take a sequence of flat fields with gray level close to saturation. and compute the average flat field, F, of the sequence. (3) Subtract the average dark image from the average flat field. (4) Compute the average intensity, $\tilde{m}$, for the difference, $F - D$. (5) Let U be a constant image of intensity 1, then $\tilde{m}U$ is an image of uniform intensity $\tilde{m}$. (6) Divide each pixel in $\tilde{m}U$ by the value of the corresponding pixel in $F - D$. Store the resulting image, $C = (\tilde{m}U)/(F - D)$ (note that the division here is is pixel-by-pixel).

Steps 1-6 are executed once for the given sensor. Images acquired with the sensor under physical (temperature) conditions similar to the one in steps 1-6 are corrected using D and the image C calculated in step 6. For a specific image Z, execute **on line**: (7) Subtract the average dark image from Z. the result is $Z - D$. (8) Now scale every pixel in $Z - D$ according to the corresponding pixel value in C.

3.2 The digital image model

We justify the validity of the procedure under a simple image model. The noise is spatially varying over the array, but signal-independent. In this paper, we do not model the random component of the noise. In our current studies we will investigate signal dependent noise models.

We assume that the state of a pixel is represented by a random variable[16] with fixed distribution over

[16]We denote random variables with capital letters, observations of these random variables with the corresponding lower

time; random variables associated with different pixels are independent. We use the time averages as means of estimating the ensemble averages of the processes involved[17].

We fix a pixel $(u,v), 1 \leq u \leq M, 1 \leq v \leq N$, where MN is the size of the digital image. The digital image model (sensor model) for the intensity, $Z(u,v;\theta)$, is given by

$$Z(u,v;\theta(u,v)) = \beta(u,v)\theta(u,v) + m(u,v) + W(u,v), \tag{1}$$

where $\theta(u,v)$ denotes the true, unobservable (ideal) intensity of the pixel, the gain factor $\beta(u,v)$ models the PRNU effect [18]), $m(u,v)$ is the offset (fixed pattern in dark current). $W(u,v)$ is a zero-mean random variable which is assumed independent of θ (thus, independent of the scene radiance). We use the model to correct any particular image so that in the corrected image the offset will be zero, and the gain will be constant over the whole image; this gain is equal to the average spatial photoresponse nonuniformity $\tilde{\beta}$; additive noise with zero mean is present in the corrected image. We manipulate dark images and flat fields to achieve the correction.

3.2.1 Estimation of $m(u,v)$

In case of dark images (no access of light), $\theta(u,v) = 0$. From (1), we obtain that the expected value $E(Z(u,v;0)) = m(u,v)$. We take a sequence of m dark images $\{z_i(0) = \{z_i(u,v;0) : 1 \leq u \leq M, 1 \leq v \leq N\}\}_{i=1}^{m}$, and compute for the pixel (u,v), the sample mean, $\bar{z}(u,v;0)$. Under the ergodicity and stationarity assumption, $\bar{z}(u,v;0)$ is an estimate of $E(Z(u,v;0))$, and thus of the offset $m(u,v)$: $\hat{m}(u,v) \stackrel{def}{=} \bar{z}(u,v;0) = m^{-1}\sum_{i=1}^{i=m} z_i(u,v;0)$. The variance of that estimator is $Var(\hat{m}(u,v)) = Var(W(u,v))/m$.

3.2.2 Expression for $\beta(u,v)$ in terms of the average $\tilde{\beta}$

Let $\tilde{\beta} = (MN)^{-1}\sum_{u=1}^{M}\sum_{v=1}^{N}\beta(u,v)$. We take a sequence of m flat fields at uniform scene radiance, $\{z_i(\theta_0) = \{z_i(u,v;\theta_0) : 1 \leq u \leq M, 1 \leq v \leq N\}\}_{i=1}^{m}$, θ_0 is the ideal intensity corresponding to the constant scene radiance. For a fixed pixel (u,v), $z_i(u,v,\theta_0), i = 1,\ldots,m$ are independent identically distributed observations according to the model (1) with $\theta(u,v) = \theta_0$, and $E(Z(u,v;\theta_0)) = \beta(u,v)\theta_0 + m(u,v)$. As before, the time average $\bar{z}(u,v;\theta_0)$ approximates the mean $E(Z(u,v,\theta_0))$ with variance $Var(W(u,v))/m$. Using the estimate $\hat{m}(u,v)$, we obtain,

$$\bar{z}(u,v;\theta_0) - \hat{m}(u,v) = \beta(u,v)\theta_0. \tag{2}$$

If we average now spatially, (2) becomes,

$$\tilde{z}(\theta_0) - \tilde{m} = \tilde{\beta}\theta_0, \text{ where } \tilde{z}(\theta_0) = (MN)^{-1}\sum_{u=1}^{M}\sum_{j=1}^{N}\bar{z}(u,v;\theta_0),\ \tilde{m} = (MN)^{-1}\sum_{u=1}^{M}\sum_{v=1}^{N}\hat{m}(u,v). \tag{3}$$

Since $\tilde{\beta}\theta_0 \neq 0$, taking the ratio of (2) and (3), after algebraic manipulations, we obtain,

$$\beta(u,v) = \tilde{\beta}\frac{\bar{z}(u,v;\theta_0) - \hat{m}(u,v)}{\tilde{z}(\theta_0) - \tilde{m}}. \tag{4}$$

---

case letters. We will put a bar, $\bar{}$, above a symbol to denote temporal average, and a tilde, $\tilde{}$, to denote a spatial average.

[17] The stochastic process representing the whole digital image is stationary, uncorrelated and ergodic

[18] Note that in this context the PRNU factor refers to the nonuniformity of the digital pixel values, which will be proportional to the PRNU for the camera sels only when the frequency of the pixel clock of the camera and the sampling frequency of the framegrabber are the same.

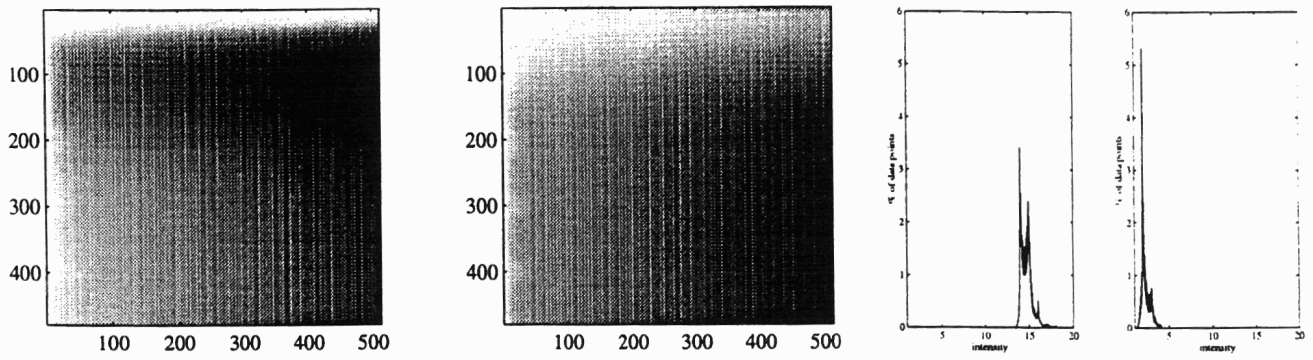


Figure 6: SONY XC-77/DT1461:Left and right averaged dark images and their histograms.

3.2.3 The model for the corrected image, Z_c

For a pixel (u, v), the model for the radiometricly corrected image is obtained according to

$$Z_c(u, v; \theta(u, v)) = (Z(u, v; \theta(u, v)) - \hat{m}(u, v)) \frac{\tilde{z}(\theta_0) - \tilde{m}}{\tilde{z}(u, v, \theta_0) - \hat{m}(u, v)}. \tag{5}$$

Making substitutions in the right-hand side of (5) according to (1) and (4):

$$Z_c(u, v; \theta(u, v)) = \tilde{\beta}\theta(u, v) + W_c(u, v), \tag{6}$$

where the uncertainty in the corrected image is captured by the random noise $W_c(u, v)$, and

$$\begin{array}{lll} E(W_c(u, v)) = 0, & \text{and} & Var(W_c(u, v)) = \frac{(\tilde{z}(\theta_0) - \tilde{m})^2}{(\tilde{z}(u, v; \theta_0) - \hat{m}(u, v))^2} Var(W(u, v)); \\ E(Z_c(u, v; \theta(u, v))) = \tilde{\beta}\theta(u, v) & \text{and} & Var(Z_c(u, v; \theta(u, v)) = Var(W_c(u, v)), \end{array}$$

i.e. the corrected image has expected uniform photoresponse over the whole array. Note that since the simple model (1) does not account for the photoresponse and noise dependence on the scene radiance, the correcting procedure depends on the flat filed level θ_0. As we illustrate in the example in the next section, for best results, θ_0 should be as close as possible to saturation (but not saturated).

4 Disparity map: an example of multicamera algorithm

We give an example of multiple-camera algorithm for which the radiometric correction matters:: disparity map in stereo matching for images which lack strong texture. We demonstrate the positive effect which the radiometric correction has. When the texture is not strong, the matching algorithm matches systematic noise components in the images. We show that by correcting radiometricly the images, the number of false matches is reduced. For a given stereo pair of images we compute, first, the disparity map; second, the radiometricly (flat-field) corrected images for each sensor; third, the disparity maps based on the uncorrected and on the flat-field corrected pairs; last, we evaluated the disparity maps.

4.1 Experimental setup

Configurations SONY XC-77/DT1461 was used. The two cameras were connected to the same physical frame grabber, each camera was equipped with 25mm lens. The pair of cameras were approximately $80cm$ apart, and at a verging angle of approximately 40°.

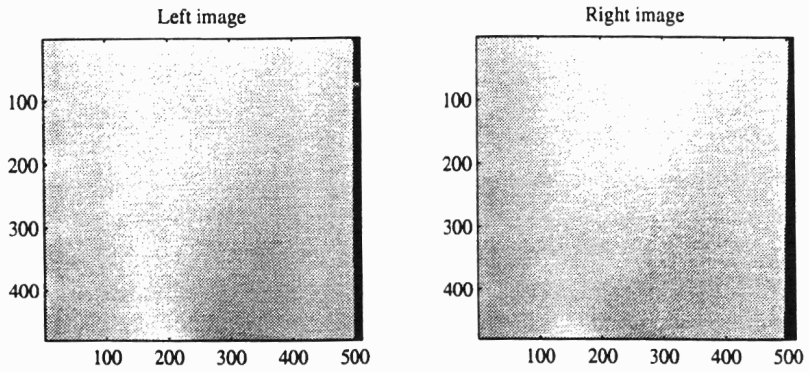


Figure 7: The original stereo pair.

Steps 1-6 from the radiometricly correcting procedure were executed[19]. Figure 6 (page 12) shows the average dark images and their corresponding histograms, for each of the two sensors. The average flat fields are given in Figure 4 (page 8).

On the day of the experiment, the cameras were warmed up, and geometrically calibrated (intrinsic and extrinsic camera parameters, using least squares, were recovered). The target was a planar white card, Lambertian reflectance surface. The approximate orientation of the card was parallel to the y axes of each camera. The lens parameters and illumination were controlled to prevent saturation. The matching uses normalized cross-correlation [18], and match selection method forbidden zone constraint [27]. To account for the suppression of high frequencies due to low pass filtering process of the image (in camera and digitizer) a constant gain k is used in the calculation of image derivatives during the subpixel disparity map computation (see below). The gain is constant for a fixed pair of cameras [22].

4.2 The test

First, a stereo pair of raw images was acquired (Figure 7 (page 13)), then the radiometricly corrected pair of images was computed. Second, a subpixel disparity map based on the *rectified original stereo pair* images was computed (we call this "uncorrected" disparity map). The gain constant k used by the matching algorithm was tuned to an optimal value 0.71 [22]. Third, using the same rectification and matching algorithms, a subpixel disparity map based on the *rectified flat-field corrected stereo pair* of images was computed (we refer to this map as "corrected"). No tuning was attempted, $k = 1$. Figure 8 (page 14), shows the disparity maps as 3D plots. The full size of the disparity maps is 256x256 since the rectification was done in half resolution. For the purpose of displaying the result, the 3D plots show the subsampled maps at a quarter of that resolution. "Holes" in the disparity maps can be observed, where the algorithm failed (matches were rejected). Spikes are outliers which erroneously were accepted as valid. The "corrected" map has less holes and spikes.

4.3 Evaluation

The two subpixel disparity maps were evaluated by fitting a plane to each one of them, and comparing statistics of the residuals from the planar fit. Figure 9 (page 14), shows the histogram for the residuals in the two cases. The standard deviation of the residuals, for the "uncorrected" map was 0.1668, and for the "corrected" 0.0679, relative improvement of about 59%.

[19] Average dark images and flat fields were based sequences of 100 images. When the distributions of the noise sources for the individual sensors are empirically confirmed, the sample size of the images to be averaged could be adjusted accordingly. In any case, the averaged dark image and flat filed acquisition is one-time, off-line process. For flat field acquisition caution should be paid to eliminate a possibility of image residues.

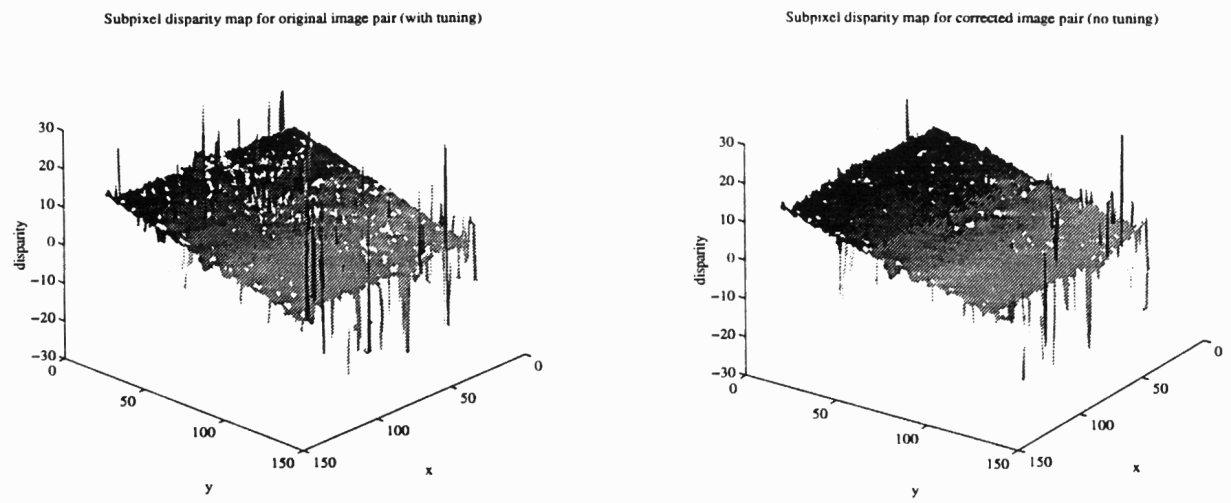


Figure 8: The subpixel disparity maps shown as 3D plots.

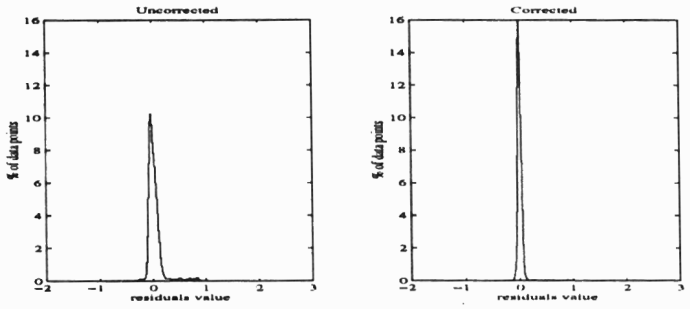


Figure 9: Histograms the of residuals

The experiment was repeated with flat field pairs at four different average levels, approximately 60, 110, 160, 220. Subpixel disparity maps without and with flat-field correction were computed based on these, planes were fitted to each one, and statistics of residuals were compared. The dominance of the performance of the matching algorithm (in relative rms) with flat-field correction over the matching without flat-field correction is observed: from 26%, for the low intensity level flat field (level 60), to 59%, for the matching with flat field correction at level 220.

5 Conclusions

We reviewed systematic and random sources of errors in the digital images. Before a video sensor is used for *quantitative* measurements, it is necessary at least that the sensor warmup time is determined; the offset in background noise, and PRNU is investigated; care should be taken to minimize image residue. The CCD array should be inspected for "hot" and "stuck" pixels. "Stuck" pixels are defective and should be excluded from consideration by measurement algorithms. The camera/digitizer interface should be evaluated, if the digitizer allows it, the sampling frequency should match the pixel output frequency. In digital cameras, the ADC is on the chip, so at least the aliasing due to camera/framegrabber mismatch is eliminated, still the rest of the sources of error we discussed are present. We showed that accounting for the systematic radiometric variation in the measurements of multiple sensors and "normalizing" the digital images prior to matching improves significantly the results in the computation of the disparity maps for weak texture scene. Such correction is very important in algorithms relying primary on intensity values, like shape from shading for example. In multisensor application, the effort should be

made to minimize the sources of errors in the input, the radiometric correction procedure is simple and fast (after the off line steps are performed, the online consists of one pixelwise subtraction and a multiplication). The procedure we discussed is oversimplified: it does not model well the video sensor, still it improves significantly the results in cases where the uncorrected algorithm experienced problems. This confirms the importance of the radiometric calibration, so we intend to address the problem in its full complexity. We plan to use the procedure to help stereo matching locally (at palaces where it fails due to uniformity of the patches). We continue our work on deriving an image model for the radiometric distortions and error measurement models which we will use in multisensor vision applications for: (i) "equalizing" the output from different sensors; (ii) estimating the random noise error statistics in the digital images; (iii) propagating the errors distributions through the vision algorithms and deriving a quantitative measure for their performance. We will obtain for each pixel in the left image, a fixed size confidence interval, for the true disparity at the pixel. This confidence interval has highest confidence coefficient (i.e., it has maximal minimal probability of capturing the true disparity in the interval). The theory of these intervals has been developed, [13], and tested successfully in the application of mobile robot pose estimation, [16].

Acknowledgements

This work is supported by, or in part by, the U.S. Army Research Office under grant number DAAH04-96-1-0007, the Defense Advanced Research Projects Agency under grant number N00014-92-J-1647, and the National Science Foundation under grant SBR89-20230.

We would like to thank Radim Šára for the useful discussions and comments. Thanks to all colleagues who provided dark images.

References

[1] Y. Beers, *Introduction to the theory of error*, Addison-Wesley, 1957.

[2] H. Beyer, "Linejitter and geometric calibration of CCD cameras", *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 45, pp. 17-32, 1990.

[3] H. Beyer, "Geometric and radiometric analysis of a CCD-camera based photogrammetric close-range system", *Dissertation, Institut für Geodäsie und Photogrammetrie, Zürich*, 1992.

[4] E. Beynon and D. Lamb,edt., *Charge-coupled devices and their applications*, McGraw-Hill, 1980.

[5] H. Ennes, *Television broadcasting: equipment, systems, and operating fundamentals*, Howard W. Sams & Co, Inc., 1971.

[6] W. Förstner, "10 Pros and cons against performance characterization of vision algorithms", *Workshop on performance characterization of vision algorithms*, Robin College, Cambridge, 1996.

[7] R. Haralick, "Performance characterization in computer vision", *Performance versus methodology in computer vision*, Haralick and Meer, editors, University of Washington, Seattle, 1994.

[8] G. Holst, *CCD arrays, cameras, and displays* , Winter Park, FL : JCD Pub., 1996.

[9] E. Healey and R. Kondepudy, "Radiometric CCD Camera Calibration and Noise Estimation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 16, No. 3, pp. 267-276, 1994.

[10] B. Jähne, *Digital image processing: concepts, algorithms and scientific applications*, Springer-Verlag, 1993.

[11] J. Janesick et al, "The future scientific CCD", *Proc. of SPIE State-of-the-art imaging arrays and their applications, San Diego, California*, Vol. 501, pp. 2-33, 1984.

[12] J. Janesick et al, "Sandbox CCDs", *Proc. of SPIE Charge-coupled devices and solid state optical sensors V, San Jose,California*, Vol. 2415, pp. 2-42, 1995.

[13] G. Kamberova and M. Mintz, "Minimax rules under zero-one loss for a restricted location parameter", *Journal of statistical Planning and Inference*, Submitted, 1996.

[14] R. Lenz and D. Fritsch, "Accuracy of videomery with CCD sensors", *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 45, pp. 90-110, 1990.

[15] R. Lenz and R. Tsai, "Techniques for calibration of the scale factor and image center for high accuracy 3D machine vision", *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 45, pp. 90-110, 1990.

[16] R. Mandelbaum and M. Mintz, "A Confidence Set Approach to Mobile Robot Localization", *IEEE-ICIROS*, Osaka, Japan, 1996

[17] I. Mclean, *Electronic and computer-aided astronomy: from the eyes to electronic sensors*, Ellis Horwood Limited, England, 1989.

[18] H. Moravec, "Robot rover visual navigation", *Computer Science:Artificial Intelligence*, pp. 13-15, 105-108, UMI Research Press 1980/1981.

[19] Photometrics Homepage,
Photometrics High performance CCD Imaging, `http://www.photomet.com`, 1996.

[20] C. Poynton, *A Technical introduction to digital video*, John Wiley & Sons, Inc., 1996.

[21] K. Price, "Anything you can do, I can do better (No you can't)...", *Computer Vision, Graphics, and Image processing*, Vol. 36, pp. 387-391, 1986.

[22] R. Šára and R. Bajcsy, "Reconstruction of 3-D geometry and topology from polynocular stereo", GRASP Laboratory, University of Pennsylvania. work in progress.

[23] C. Slama, edt., *Manual of photogrammetry, IVth edition*, American Society of Photogrammetry, 1980.

[24] D. Snyder *et al.*, "Compensation for readout noise in CCD images", *Journal of the Optical Society of America, A*, Vol. 12, No. 2, pp. 272-283, 1995.

[25] A. Sripad and D. Snyder, "A necessary and sufficient condition for the quantization error to be uniform and white", *IEEE Trans. on acoustic, speech, and signal processing*, Vol. ASSP-25,No 5, pp. 442-448, 1977.

[26] A. Theuwissen, *Solid-State imaging with charge-coupled devices*, Kluwer Academic Publishers, 1995.

[27] A. Yuille and T. Poggio, "A Generalized Ordering Constraint for Stereo Correspondence", *MIT, Artificial Intelligence Laboratory Memo*, No. 777, 1984.