ISPRS Journal of Photogrammetry and Remote Sensing, 45 (1990) 90–110
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

Accuracy of videometry with CCD sensors*

Reimar Lenz[a] and Dieter Fritsch[b]
[a]*Institute for Telecommunications, and* [b]*Institute for Photogrammetry, Technical University München, FR Germany*

(Received 30 December 1988, revised and accepted 04 July 1989)

ABSTRACT

Lenz, R. and Fritsch, D., 1990. Accuracy of videometry with CCD sensors. *ISPRS J. Photogramm. Remote Sensing*, 45: 90–110.

For accurate computer vision based on standard video signals, the term Videometry is introduced. Some geometrical, optical and electrical properties of CCD-cameras in conjunction with analog/digital-converters and frame buffers are investigated: lens distortion, sensor distortion, anisotropic modulation transfer function, space-variant impulse response due to discrete sensor elements and insufficient optical low-pass filtering, horizontal line jitter and scaling factor due to mismatch of sensor-shift- and A/D-conversion-clock, noise etc. Based on these results, a very simple camera model with a special radial lens distortion equation is proposed. This allows for a fast, fully linear calibration algorithm (15 msec calibration time for 36 coplanar calibration points) with good accuracy (1/30 of a frame buffer pixel residual error). It requires independent pre-calibration of the principal point and the horizontal scale factor. The latter is performed by Fourier analysis of the aliasing patterns produced by interference of camera- and A/D-converter clock. Only small improvements (3% average error, 30% maximum error) were obtained by subsequent nonlinear optimization (self-calibration with bundle adjustment) using the results from the linear approach as initial guess.

1 INTRODUCTION

This paper is concerned with the accuracy of imaging with solid-state, discrete-array sensors (short: CCD-sensors). The interface between camera and digitizer/computer is assumed to be the standard, B/W video signal (RS170 or CCIR). Theoretical predictions are compared with actual measurements on a modern, well designed camera (Panasonic WV-CD50) with a Sony interline transfer 2/3″ CCD-Sensor (500 Sensor Elements (Sels) horizontally, pitch 17 μm and 583 Sels vertically, spaced at 11 μm), digitized with several frame grabbers [Imaging Technologies AP512, Kontron IBAS II and Matrox

*Updated version of a paper presented at the 16th ISPRS Congress in Kyoto, Japan.

0924-2716/90/$03.50

PIP-1024A, all with 512×512 Picture Elements (Pels)]. In detail, the following was investigated:

(a) Geometrical camera model
- calibration of exterior parameters: rotation, translation
- calibration of interior parameters: principal distance, lens distortion, principal point, scaling factors
- analyzed model errors such as line jitter, spatial quantization, center offset caused by perspective imaging and lens distortion, sensor distortion
- neglected model errors are: 5th order radial lens distortion, tangential lens distortion, thermic camera instability.

(b) Signal transfer model
- optical transfer function, influenced by diffraction, defocussing, phase errors of lens surface
- sensor transfer function, influenced by local integration, sampling, linearity
- electrical transfer functions (x-direction only), in particular sample-and-hold circuitry, low-pass filtering, sampling for analog to digital conversion
- random noise such as photon noise, amplifier noise and quantization noise
- fixed pattern noise such as sensor noise, A/D converter noise and computer noise
- analyzed errors are: periodically space-variant and asymmetrical impulse response.

2 THE GEOMETRICAL CAMERA MODEL

Figure 1 shows the annotations used for the camera model which originates in work from Tsai (1985) and was modified by Lenz (1987b) to allow for a fully linear calibration algorithm capable of real-time performance.

The geometrical imaging process may be subdivided into four steps:

(a) Rigid-body transformation (object coordinate system (CS) to camera-CS):

$$\begin{pmatrix} x_c \\ y_c \\ z_c \end{pmatrix} = \begin{pmatrix} r_{xx} & r_{xy} & r_{xz} \\ r_{yx} & r_{yy} & r_{yz} \\ r_{zx} & r_{zy} & r_{zz} \end{pmatrix} \begin{pmatrix} x_w \\ y_w \\ z_w \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix} \tag{1}$$

(b) Perspective transformation with principal distance b (camera-CS to undistorted sensor-CS):

$$x_{su} = b\, x_c / z_c; \qquad y_{su} = b\, y_c / z_c \tag{2}$$

(c) 3rd order radial lens distortion (undistorted sensor CS to distorted sensor CS and vice versa):

$$x_{su} = x_{sv} / (1 + k_3 r_{sv}^2); \qquad y_{su} = y_{sv} / (1 + k_3 r_{sv}^2); \qquad r_{sv}^2 = x_{sv}^2 + y_{sv}^2 \tag{3a}$$

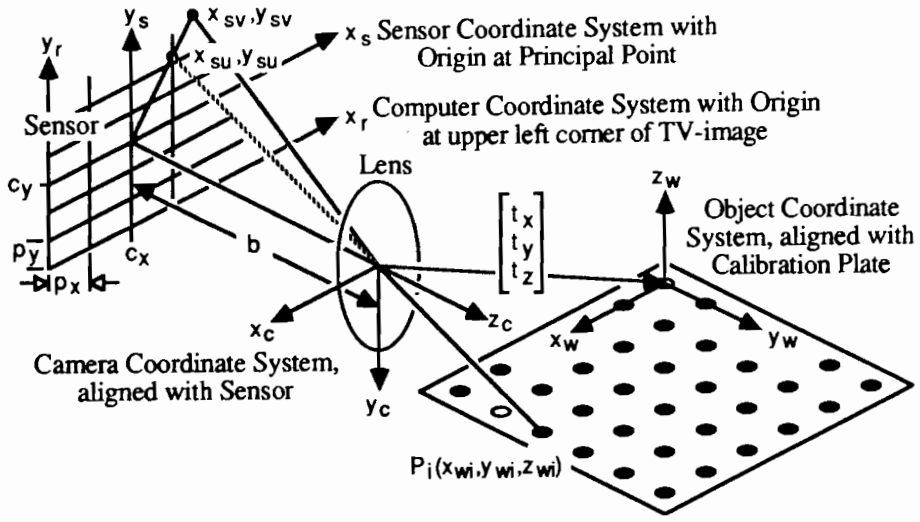


Fig. 1. Geometrical camera model with six exterior and six interior parameters.

$$x_{sv}=2x_{su}/(1+[1-4k_3r_{su}^2]^{1/2}); \qquad y_{sv}=2y_{su}/(1+[1-4k_3r_{su}^2]^{1/2});$$

$$r_{su}^2=x_{su}^2+y_{su}^2 \tag{3b}$$

whereby (3a) has an analytical inverse (3b) and allows for a fully linear calibration algorithm.

(d) Scaling and principal point location (distorted sensor CS to computer CS and vice versa):

$$x_r = x_{sv}/p_x + c_x; \qquad y_r = y_{sv}/p_y + c_y \tag{4a}$$

$$x_{sv} = (x_r - c_x)p_x; \qquad y_{sv} = (y_r - c_y)p_y \tag{4b}$$

(c_x, c_y) are the frame buffer ot computer coordinates of the principal point, p_x (p_y) is the distance on the image sensor between two horizontally (vertically) adjacent frame buffer Pels (*not* sensor Sels). These last four intrinsic parameters are assumed to be already known for the camera calibration procedure described in the following.

2.1 Calibration

2.1.1 Calibration of exterior parameters

Tsai (1985) invented a trick to eliminate k_3, sacrificing one equation per observation:

Dividing the first two equations of (3b) and substituting (1) and (2) eliminates b, k_3, t_z and r_{zx}, r_{zy}, r_{zz}:

$$x_{sv}/y_{sv} = (x_w r_{xx} + y_w r_{xy} + z_w r_{xz} + t_x)/(x_w r_{yx} + y_w r_{yy} + z_w r_{yz} + t_y) \tag{5}$$

When using a coplanar set of calibration points P_i with known coordinates

(x_{wi}, y_{wi}, z_{wi}), the objects CS can be chosen such that $z_{wi} \equiv 0$ and $t_z > 0$ without losing generality. This eliminates r_{xz} and r_{yz} and leads to one equation per observation (x_{svi}, y_{svi}, E expectation):

$$E(y_{svi})x_{wi}r_{xx} + E(y_{svi})y_{wi}r_{xy} - E(x_{svi})x_{wi} - E(x_{svi})y_{wi}r_{yy} + E(y_{svi})t_x - E(x_{svi})t_y = 0 \quad (6)$$

of a homogeneous system of equations $\mathbf{Av}=\mathbf{0}$ which is solved in a least-squares sense for the nontrivial solution $\mathbf{v}=(r_{xx}, r_{xy}, r_{yx}, r_{yy}, t_x, t_y)^T$ with $\mathbf{A}^T\mathbf{Av}=\lambda\mathbf{v}$ for the smallest eigenvalue λ. Because r_{xx}, r_{xy}, ... are elements of a 3×3 orthonormal matrix, all elements of $\mathbf{v}$ must be scaled with a common constant such that:

$$[(r_{xx}+r_{yy})^2+(r_{xy}-r_{yx})^2]^{1/2}+[(r_{xx}-r_{yy})^2+(r_{xy}+r_{yx})^2]^{1/2}=2 \quad (7)$$

introducing a sign ambiguity. Using the orthonormal property again, we can solve for:

$$r_{zx}=[1-r_{xx}^2-r_{yx}^2]^{1/2}; \qquad r_{zy}=-[1-r_{xy}^2-r_{yy}^2]^{1/2}\,\mathrm{sign}(r_{xx}r_{xy}+r_{yx}r_{yy}) \quad (8)$$

introducing yet another sign ambiguity, both of which are resolved later. sign(...) is +1 or −1, depending on the sign of the argument.

2.1.2 Calibration of interior parameters

With (1,2,3) and the results by (7,8), a linear system of equations (two for each observation) is set up and solved with least squares for b, bk_3 and t_z:

$$\begin{aligned} x_{ci}b + x_{ci}r_{svi}^2bk_3 - x_{svi}t_z &= x_{svi}(x_{wi}r_{zx}+y_{wi}r_{zy}) \\ y_{ci}b + y_{ci}r_{svi}^2bk_3 - y_{svi}t_z &= y_{svi}(x_{wi}r_{zx}+y_{wi}r_{zy}) \end{aligned} \quad (9)$$

with $x_{ci}=x_{wi}r_{xx}+y_{wi}r_{xy}+t_x$; $y_{ci}=x_{wi}r_{yx}+y_{wi}r_{yy}+t_y$; $r_{svi}^2=x_{svi}^2+y_{svi}^2$.
Because b and t_z must be positive, we can now resolve all sign-ambiguities by multiplying:

$$\{r_{xx}, r_{xy}, r_{yx}, r_{yy}, t_x, t_y\} \text{ with } \mathrm{sign}(b/t_z),$$
$$\{r_{zx}, r_{zy}, t_z\} \text{ with } \mathrm{sign}(t_z) \text{ and } \{b, bk_3\} \text{ with } \mathrm{sign}(b). \quad (10)$$

Obtaining r_{xz}, r_{yz}, r_{zz} with the outer product completes the calibration, using only physically meaningful, independent parameters and linear equation systems without the need of an initial guess. The plane of calibration points *must not* be nearly parallel to the image sensor. Residuals with $\sigma=0.18$ μm were reached.

2.1.3 Calibration of scale factors

Due to TV line scanning convention, p_y in (4a,b) is identical to s_y, the distance between two vertically adjacent sensor elements (Sel-pitch) and known precisely from the manufacturer's specification. Fairchild 1984 specifies ±5

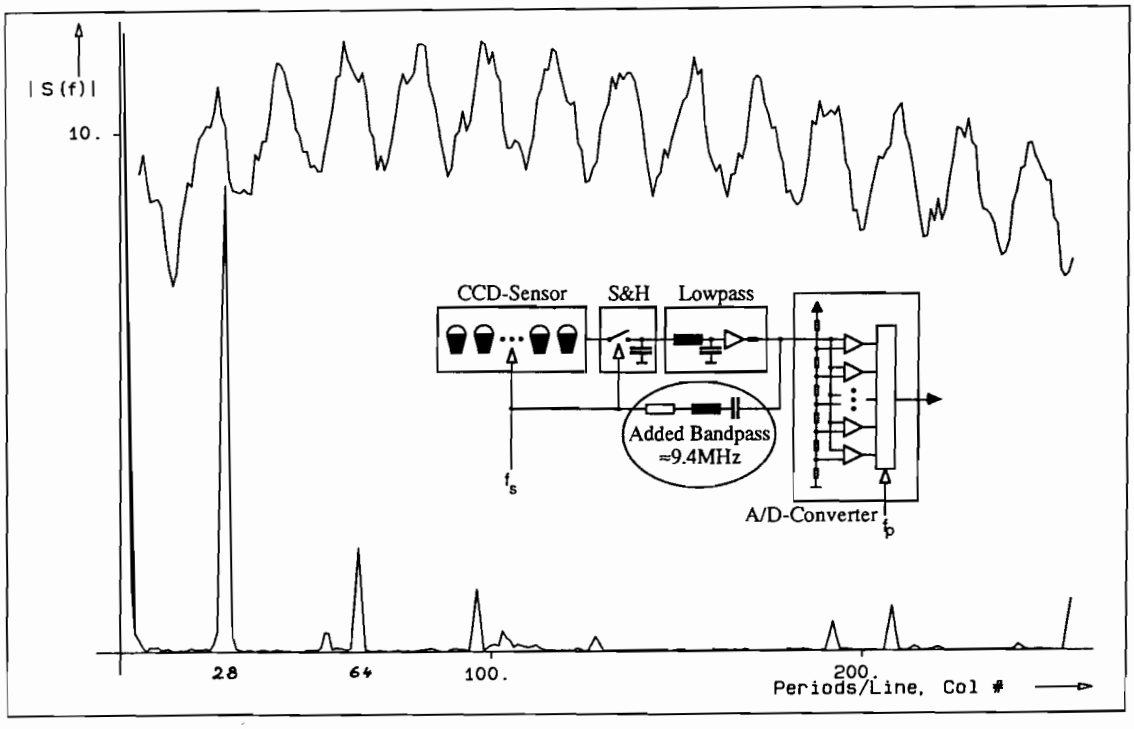


Fig. 2. The Sensor element (Sel) clock signal, added as reference to the video signal with the circuitry shown, can be observed aliased in a TV-line digitized with the Picture element (Pel) clock rate. The horizontal scale factor and line jitter are determined by phase and amplitude analysis of the peak at 28 periods/512 Pels in the Fourier spectrum.

ppm cumulative pitch error for the electron beam written mask, $\pm 0.016\ \mu m$ between any two adjacent Sels and -0.23% to -0.46% isotropic contraction due to subsequent high temperature processing steps. Because the latter affects x- and y-direction in the same fashion, the — more important — ratio between horizontal and vertical Sel-pitch is known to ± 10 ppm. A small common contraction is absorbed by the principal distance b in (2). For the Sony sensor used by the authors, this contraction was found to be $-0.03\% \pm 0.04\%$ in x-direction by measuring a distance of 6798 μm between two sensor elements 400 Sels apart, the camera in Figure 8 mounted on a micrometer stage with $\pm 3\ \mu m$ measurement error. (Measuring the distance between 500 Sels failed, because out of the 500 specified by Panasonic only 484 or 485 show up in the analog output signal, thus reducing the active image width from 500 Sels $\times$ 17 μm/Sel = 8.5 mm to 8.25 mm (8.8 mm were specified). Falsely specifying important features (the number of Sels and the imaging area are often 'overestimated') seems to be symptomatic of CCD-camera manufacturers — Fairchild is a commendable exception). With the limited accuracy of the micrometer no sensor distortion could be found. More accurate interferometric measurements are under way. In nonpixel-synchronized, TV-standard-based systems, the Sels with the pitch s_x are read out and sampled and held with the CCD-shift register clock frequency f_s, converted into an analog

signal with added TV line and field synchronization pulses and subsequently sampled, A/D-converted and digitally stored as Pels with the clock frequency f_p of the frame grabber (Fig. 2), leading to a Pel-pitch of:

$$p_x = s_x f_s / f_p \approx s_x (N_{Sels} / N_{Pels}) \quad (11)$$

in which N_{Sels} is the number of active Sels per line and N_{Pels} is the number of Pels per line.

2.2 Errors of the geometrical camera model

2.2.1 Line jitter

Frame grabbers with true Phase Locked Loop (PLL) line synchronization enforce a fixed number of clock cycles per line (e.g. 640 for the ITI-AP512 with 512 Pels, leaving 128 cycles for the horizontal blanking period). Since the horizontal sync pulse coming from the camera is usually derived from the same master clock which is used for the CCD-clock (e.g., 455 cycles/line for the Fairchild CCD3000 with 380 Sels/line), the average ratio f_s/f_p is fixed for such systems (e.g. $455/640 \approx 0.711$, which is only approx. $N_{Sels}/N_{Sels} = 380/512 \approx 0.742$ due to differing *active* line length) and not subject to drift of the involved oscillators. However, PLL circuits are not perfect and there will inevitably remain some line jitter (1/4 Pel is e.g. specified by Matrox). Most of the line jitter will occur after the vertical blanking period, where line sync is usually lost due to either missing or falsely interpreted serrated horizontal sync pulses.

Non PLL-controlled frame grabbers have either an interruptable oscillator with an integer multiple of f_p, which is started at the beginning of each line, or (as the system analyzed in detail by Beyer, 1987) have a continuously running crystal oscillator with $4 f_p$ or higher and only the clock-dividing circuitry is reset at the beginning of each line, leading to a clock quantization error observable as a sawtooth-shaped line jitter as a function of the line number with a peak to peak amplitude of 1/4 Pel or less at line start. Additional, sometimes much larger, errors result from relative drift between camera and frame grabber oscillator. During warm-up time, a camera drift $\Delta f_s / d_s \approx 2\%$ (Aqua-TV HR600), fully affecting the horizontal scale factor p_x, has been observed in a thorough investigation carried out by Dähler (1987). In such cases, or for rapidly multiplexed cameras, the camera(s) should be synchronized by the frame grabber.

Non TV-Standard, pixel, synchronized systems, strongly advocated by Grün (1987) and other authors, where the camera Sel-Clock is used to trigger the A/D-converter Pel-clock or vice versa, should not have these problems — a Sel becomes a Pel ($p_x = s_x$), without jitter and drift. The scale factors would only be subject to 2 ppm/°K, the rather small linear thermal expansion coefficient of silicon.

Since systems with standard TV-Signal input are still very common, a

scheme to determine line jitter, drift and p_x/s_x with high precision was proposed by Lenz and Tsai (1986). If the Sel clock frequency f_s is superimposed onto the video signal it will show up in the digitized image as Sel reference signal (aliased of $2f_s > f_p$). Many cameras add f_s all by themselves, e.g. Fairchild CCD3000, Javelin JE2063C (MOS), General Electric TN2506 (CID); if this 'noise' is very well suppressed by electrical filters (WV-CD50) a small fraction (≈ 20 mV) was added with a bandpass filter, see Figure 2. A $n=512$-point one-dimensional Fourier analysis of a vertically averaged, low light-level blank image [$I_{av}(x_r) = \sum$ over y_r of image $I(x_r, y_r)$], taken with the WV-CD50, digitized with the ITI-AP512 and weighted with a raised cosine window [$1-\cos(2\pi x_r/n)$] to suppress image border and FFT artifacts, clearly shows the aliased Sel clock at:

$$f_a = n(f_p - f_s)/f_p \approx 28 \text{ periods/line } (\approx n(N_{Pels} - N_{Sels})/N_{Pels} = 512(512-485)/512 = 27) \quad (12)$$

An accurate estimate for the peak location f_a is obtained by solving the narrow-band approximation:

$$\underline{U}'(m) = u_a \exp -j[\pi(m-f_a)+\phi_a] \sin[\pi(m-f_a)]/ [2\pi(m-f_a)(m-f_a-1)(m-f_a+1)] \qquad j=\sqrt{-1} \quad (13)$$

of the discrete complex FFT-spectrum $\underline{U}(m,\ 0 \le m < n)$ of a cosine-weighted sine function with frequency f_a [periods/nPels], amplitude u_a and phase ϕ_a. Three consecutive spectral values $\underline{U}(m)$, $\underline{U}(m-1)$ and $\underline{U}(m+1)$ with the integer m next to f_a are used to eliminate u_a and ϕ_a in (13) and to solve for f_a:

$$f_a = m + 2 \text{ Real Part of } \{[\underline{U}(m-1) - \underline{U}(m+1)]/ [2\underline{U}(m) - \underline{U}(m-1) - \underline{U}(m+1)]\} \quad (14)$$

With $f_a n(f_p - f_s)/f_p$ from (12) we have the ratio f_s/f_p and therefore p_x:

$$\begin{aligned} p_x &= s_x f_s/f_p = s_x(1 - f_a/n) && \text{for } N_{Pels}/2 < N_{Sels} < N_{Pels} \text{ (most common case)} \\ p_x &= s_x f_s/f_p = s_x(f_a/n) && \text{for } 0 < N_{Sels} < N_{Pels}/2 \text{ (very low resolution sensor)} \\ p_x &= s_x f_s/f_p = s_x(1 + f_a/n) && \text{for } N_{Pels} < N_{Sels} < 1.5 N_{Pels} \text{ (very high resolution sensor)} \end{aligned} \quad (15)$$

Now the jitter of each individual line y_r can be determined by measuring the phase $\phi(y_r)$ of f_a with selective Fourier analysis:

$$\phi(y_r) = \text{atan}[\text{imag}\{\underline{U}(f_a, y_r)\}/\text{real}\{\underline{U}(f_a, y_r)\}], \text{ where } \underline{U}(f_a, y_r) = \sum_{x_r=0}^{n-1} [1 - \cos(2\pi x_r/n)] I(x_r, y_r) \exp(-j\pi x_r f_a/n) \quad (16)$$

In (16), f_a is taken real-valued from (14), not integer. Again, raised cosine weighting is used. The phase $\phi(y_r)$ should be accumulated from line to line (both TV-fields individually) to cope with the ambiguity of 2π, which corresponds to a shift of 1 Pel. Figure 3 shows the line jitter/shift as a function of the TV-line number. Only the first field with the phase of the first line being arbitrarily set to zero is shown, the other field behaves very similar. The total shift from first to last line is about $\frac{1}{4}$ Pel ($\pi/2$). PLL control oscillations similar to those observed by Luhmann (1987) using optical line synchronization can be seen. For the camera and frame grabber analyzed, the behavior in Figure 3 seems to be long- and short-term stable within $\frac{1}{20}$ Pel and can therefore be used to pre-correct the computer image coordinate x_r as function of y_r up to this accuracy without the need of going through the Fourier analysis for every image. The frequency modulation of f_p due to PLL deficiencies is related to the phase gradient $\Delta\phi(y_r)/\Delta y_r$. The relative global variation from top to bottom is $\approx(\pi=2)/(256\text{ lines/field}\times 2\pi\times 640\text{ cycles/line})<2$ ppm and thus negligible, whereas the error between consecutive lines of the same TV-field is significantly larger. especially at the image top with larger than average phase gradients (a maximum $\Delta p_x/p_x\approx\frac{1}{40}$ Pel/line/(640/line) ≈ 40 ppm was found). Systems with free-running oscillators are not subject to frequency modulation.

Some sources of noise can be detected in the spectrum in Figure 2: the peaks

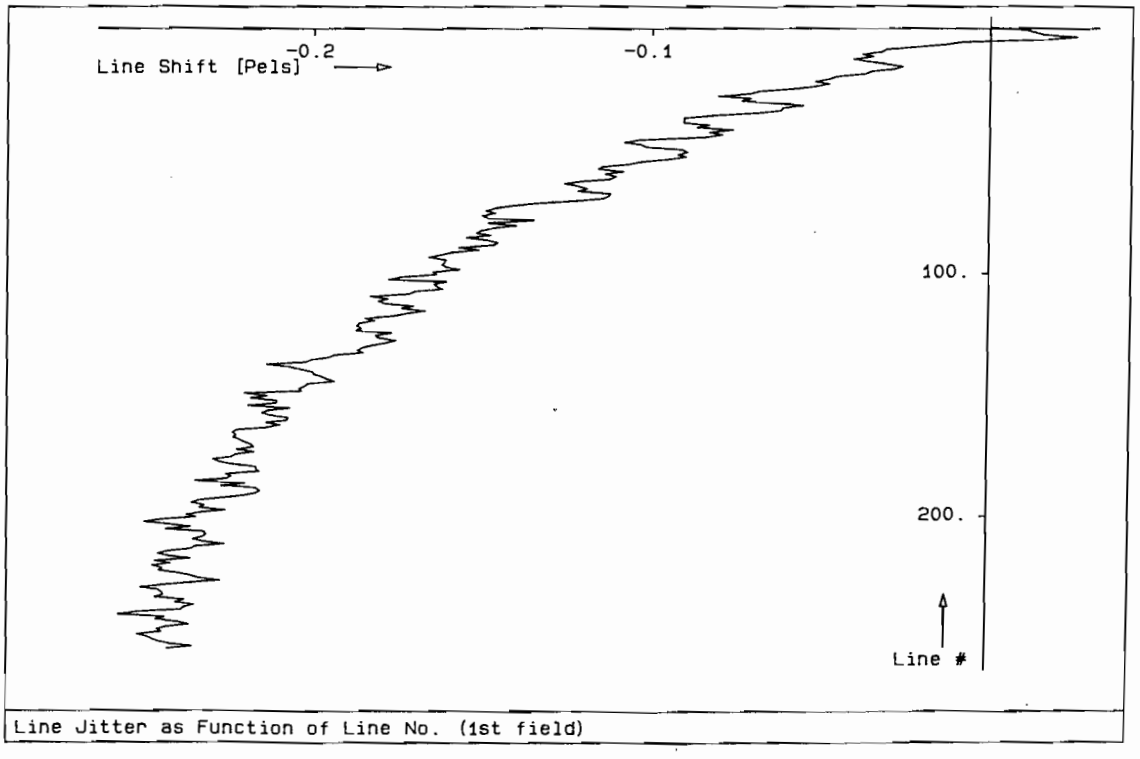


Fig. 3. Line jitter due to imperfect synchronization between camera and frame grabber, first field. Most errors are caused by the loss of perfect sync during the vertical blanking period, much less result from PLL control oscillations.

at 64, 128, 256 periods/line ($f_p/8,4,2$) are due to A/D-converter clock noise and have a fixed phase with respect to Pel sampling, some more peaks come from our host computer clock by straying of electrical fields onto the analog video signal. The (artificially added) 'noise' from the sensor clock and the ADC-noise can be subtracted from the digitized image, because its phase is more or less constant with respect to the frame buffer CS, whereas randomly phased computer noise is not pre-compensatible.

2.2.2 Further errors

Some errors of the geometrical camera model have been discussed, remedies were given. The calibration of the principal point and higher-order radial or tangential lens distortion are not subject of this paper.

Other sources of systematic error in locating the center of calibration points are treated in the following: In order to reduce the effect of spatial sensor quantization, one is tempted to use rather large calibration points for uncorrelated error averaging. This, however, has the disadvantage of introducing systematic errors. Let us assume, that circular calibration points (disks) are used. Due to central perspective imaging and/or lens distortion, the center of the circle image does *not* coincide with the image of the circle center. This deviation is quite noticeable in close-range Videometry using large calibration points. For the setup in Figure 4, where we have a calibration point observed at angle β with radius r sitting on a plane which is tilted around the camera x-axis by an angle α, the difference $y_S - y_{su}$ in y between the image y_{su} of the circle-center and center of gravity y_S of the imaged circle in Figure 5 becomes approximately:

$$y_S - y_{su} \approx b\frac{r^2}{z_c}\sin\alpha\,(\cos\alpha + \sin\alpha\tan\beta) = \frac{a^2}{b}\sin\alpha\,(\cos\alpha + \sin\alpha\tan\beta) \qquad (17)$$

where a is the radius of the circle circumscribing the imaged calibration point, z_c is the the z-coordinate of the calibration point in the camera-CS and b is the principal distance. Due to lack of space, the purely geometrical derivation of (17) is left to the reader. For the dimensions given in Figure 4 ($r = 1$ mm,

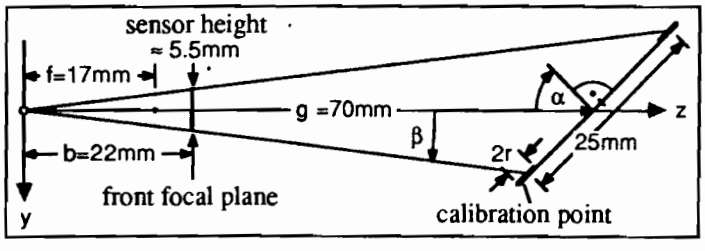


Fig. 4. Close-range videometric setup to evaluate systematic errors due to calibration point deformation by central perspective imaging and/or radial lens distortion. Both errors increase with the square of point image size 'a'.

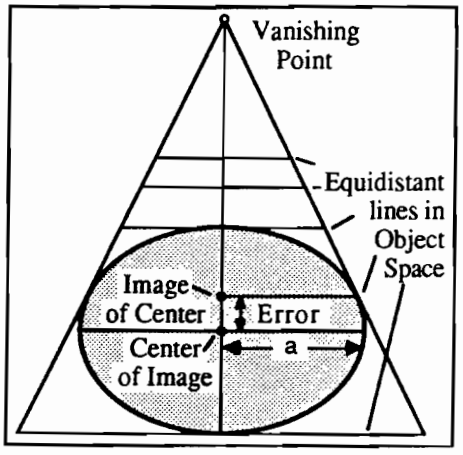


Fig. 5. For a given size 'a' of the image of a calibration point the systematic error due to perspective imaging is inversely proportional to b.

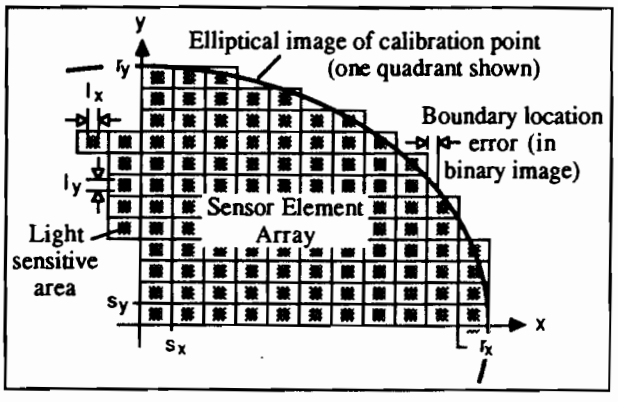


Fig. 6. Spatially quantized image of a calibration point, used to derive the relationship between boundary estimation RMSE σ_b and center estimation RMSE σ_c as a function of calibration point size. For optimally thresholded binary images, the boundary location error is evenly distributed between $\pm 0.5\, s_x$ (s_y).

$\alpha = 45°$ and $\tan\beta = 1/8$), the error amounts to 3.2 μm and is by no means negligible.

The radial difference $r_S - r_{sv}$ due to 3rd order radial lens distortion between the principal point distance r_{sv} [see (3)] of the circle-center and center of gravity r_S of the imaged circle is approximately given by:

$$r_S - r_{sv} \approx 3 r_{sv} k_3 a^2 \tag{18}$$

with 'a' as in (17).

For standard TV-lenses with large k_3 (-0.0017 mm^{-2} was found), this systematic error is in the same order of magnitude as the error in (17), $r_S - r_{sv} \approx -3$ μm in the image corners ($r_{sv} \approx 5$ mm) for the setup in Figure 4.

We will now treat the effect of spatial sensor quantization on calibration point localization accuracy. On the basis of Figure 6, showing the quantized

image of a calibration point, we will derive the mean square measurement error (MSE) σ_x^2 of the center of gravity x-coordinate in Sels, neglecting the difference between Pels and Sels for reasons to be explained later. We assume that mean square boundary location error σ_b^2 is given. For optimally binarized images with evenly distributed errors between $\pm 0.5\, s_x$, σ_b^2 would simply be:

$$\sigma_b^2 = \int_{x=-s_x/2}^{s_x/2} x^2 dx \Big/ \int_{x=-s_x/2}^{s_x/2} dx = s_x^2/12 \qquad \text{or } \sigma_b \approx 0.289 s_x$$

for binary images (19)

In greyvalue images, σ_b^2 is dependent upon optical band-limiting, sensor-element integration area, interpolation kernels, image noise and more, which will be discussed later.

The contribution $\sigma_{c(one)}^2$ of the error σ_b^2 of one boundary element to the error σ_c^2 of the x center estimate, calculated by dividing the first order moment in x by the total area of the ellipse, is dependent upon its x-coordinate:

$$\sigma_{c(one)}^2(x) = (x\sigma_b s_y/\pi r_x r_y)^2 = x^2(\sigma_b/\pi r_x n_y)^2 \tag{20}$$

where $n_y = r_y/s_y$ is the number of sensor elements from center to border in y-direction. 4 n_y boundary elements with the average x^2-coordinate x_{ave}^2:

$$x_{ave}^2 = \int_{y=0}^{r_y} (r_x^2 - y^2(r_x/r_y)^2)dy \Big/ \int_{y=0}^{r_y} dy = \tfrac{2}{3} r_x^2 \tag{21}$$

contribute to σ_c^2. For circular shaped objects, where the boundary line is effectively *uncorrelated* with the Sel raster according to investigations by Hill (1980) the *variances* $\sigma_{c(ave)}^2$ are added:

$$\sigma_c^2 = 4n_y \sigma_{c(ave)}^2 = 4n_y \tfrac{2}{3} r_x^2 (\sigma_b/\pi r_x n_y)^2 = \frac{8}{3n_y}\left(\frac{\sigma_b}{\pi}\right)^2 \text{ or } \sigma_c \approx \frac{0.15 s_x}{\sqrt{n_y}}$$

for binary images (22)

Thus the RMS-error σ_c in x-direction is inversely proportional to the square root of the number of sensor elements n_y in y-direction (and vice versa). For $n_y = 12$ as in Figure 6, $\sigma_c \approx 0.04\, s_x$. If the boundary line of the calibration 'point' is strongly *correlated* with the Sel raster (as may be the case for plumb-line calibration, with Reseau-Grids or rectangular calibration points aligned with the Sel raster) the *standard deviations* $\sigma_{c(ave)}$ might add up and, under unfortunate circumstances, increasing calibration object size may result in no accuracy gain at all. Boundary extraction schemes in greyvalue images, where the main source of error may not be spatial quantization but of some other truly random nature will gain from increasing object size, even if its boundary

is strongly correlated with the Sel raster. In order to estimate the boundary location error σ_b in greyvalue images, we have to analyze the signal transfer characteristics.

3 THE SIGNAL TRANSFER MODEL

In Figure 7, the signal transfer model is given, together with some sources of noise. A point imaged by a lens onto the sensor is degraded (spread) by diffraction, defocussing and lens errors.

3.1 Optical transfer function

If, for reasons of simplicity, a square-shaped aperture aligned with the sensor array is assumed, x- and y-axes become separable. Then, the modulation transfer function MTF (u,λ,F) as function of the spatial frequency u, wavelength λ and F-number due to diffraction of incoherent light will be:

$$\begin{aligned} &\mathrm{MTF}(u,\lambda,F) = 1 - |u|\lambda F \quad &&\text{for } |u| < 1/\lambda F \\ &\mathrm{MTF}(u,\lambda,F) = 0 &&\text{elsewhere} \end{aligned} \tag{23}$$

Defocussing is equivalent to convolving the image with a scaled version of the aperture (here a square of sidelength Δx is assumed) and leads to the modulation transfer function:

$$\mathrm{MTF}(u,\Delta x) = \int_{-\Delta x/2}^{\Delta x/2} e^{-j2\pi ux}\,dx \Big/ \int_{-\Delta x/2}^{\Delta x/2} dx = \frac{\sin(\pi\Delta xu)}{\pi\Delta xu} =: \mathrm{Sinc}(\pi\Delta xu) \tag{24}$$

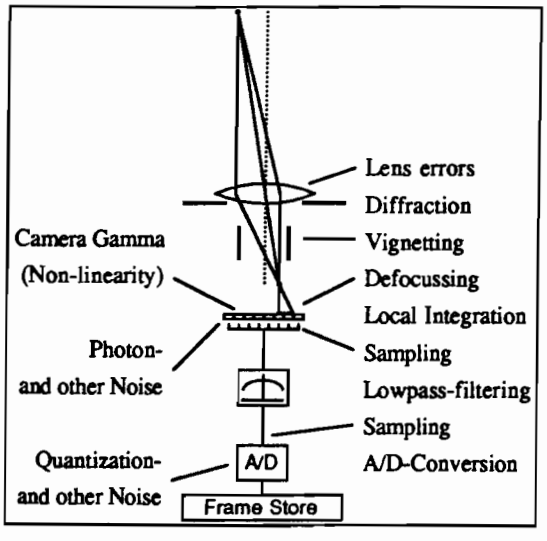


Fig. 7. Some factors affecting the signal transfer function (see the survey of Grün 1987 for more sources of radiometric degradation).

The MTFs for $\lambda=600$ nm, $F=5.6$ ($|u_{max}| \approx 300/\text{mm}$) and $\Delta x=6$ μm [$\sin(\pi \times 6\ \mu\text{m} \times n \times 167/\text{mm})=0$] and their product are shown in Figure 8. These quantities were chosen looking forward to the experiments described later.

3.2 Sensor transfer function

Next, the photons hitting the light-sensitive portion of a sensor element cell are partially converted into photo-electrons ($\approx 50\%$ quantum efficiency, Fairchild, 1984) and integrated in space and time. The temporal integration period is one TV frame time, the spatial integration area was determined by measuring the light sensitivity profile (Fig. 10) within a Sel cell with the setup in Figure 9.

At readout time, after local and temporal integration, the charges are successively sampled and converted into the analog electrical signal. The effect of spatially fixed local integration with $I_x=I_y=6$ μm and sampling can be

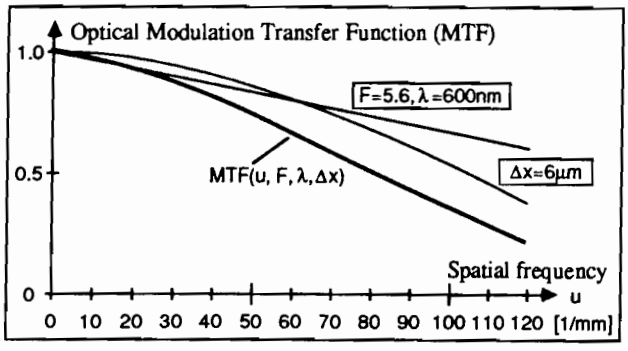


Fig. 8. Theoretical modulation transfer functions due to diffraction and defocussing. MTF attenuation due to phase errors of the lens surfaces is nearly unpredictable and not included in this figure.

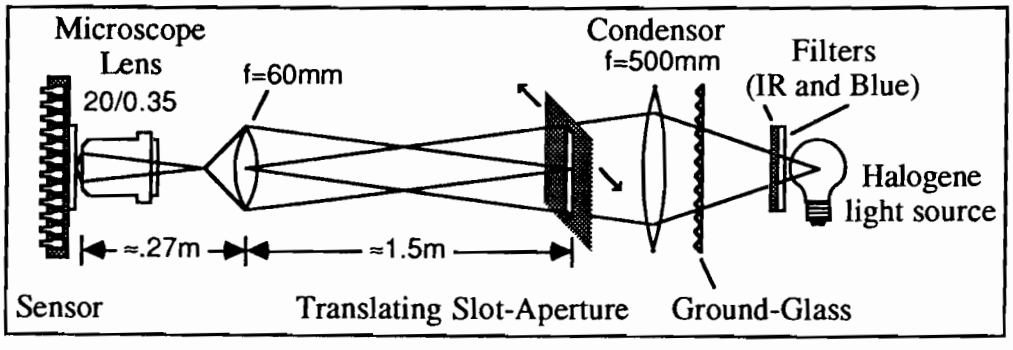
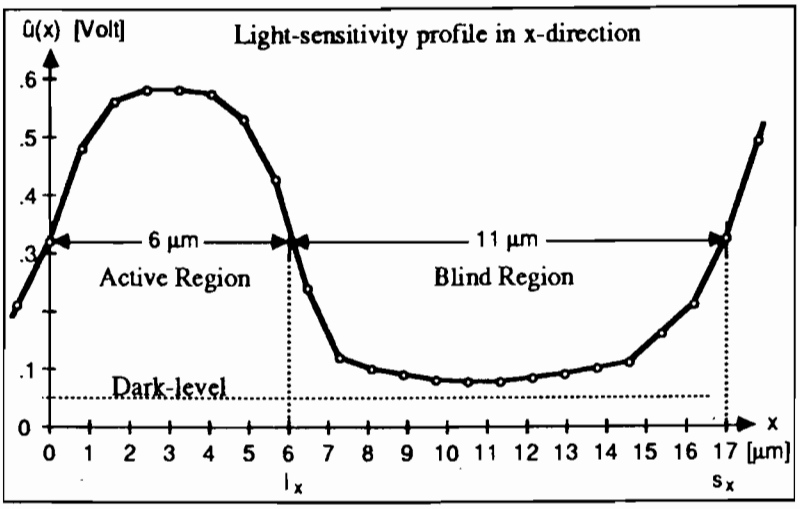


Fig. 9. Experimental setup used to measure the light sensitivity profile, impulse response, Sel-pitch and distortion of a CCD-camera. The slot-aperture consisting of two opposing razor blades (25 mm by 0.25 mm, scaled down by a factor of ≈ 620 to 40 μm by ≈ 2 μm diffraction limited width-equivalent) was aligned with either the x- or y-sensor-axis and translated in increments of 0.5 mm (0.81 μm on the sensor). For pitch and distortion measurements the camera was translated on a micrometer stage as well with the slot serving for fine adjustment.



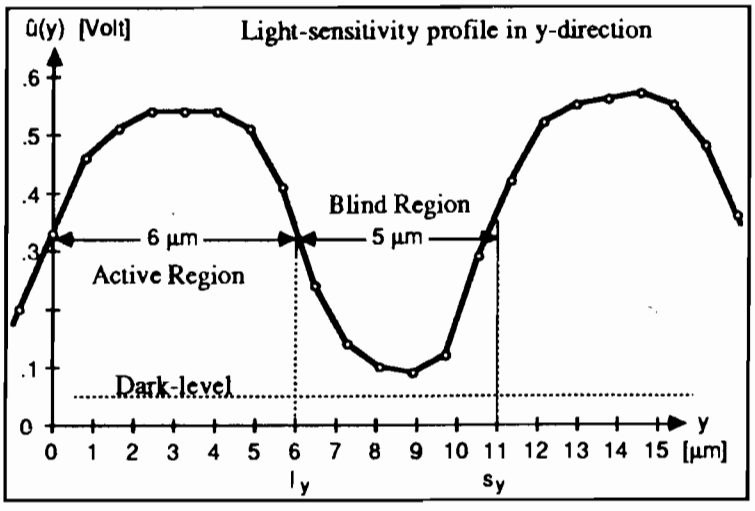


Fig. 10. Measured sensor sensitivity profile in x- and y-direction. Only $l_x = 6\ \mu m$ out of the horizontal pitch $s_x = 17\ \mu m$ and $l_y = 6\ \mu m$ out of $s_y = 11\ \mu m$ of the interline transfer CCD-sensor used in the Panasonic WV-CD50 are light sensitive. This leaves less than 20% sensitive area, whereas frame transfer devices more than to 90%. The gradual transition from sensitive to blind region is probably mostly due to the diffraction limited spatial bandwidth ($\approx 500/mm$) of the illuminating line, since manufacturers take great care to avoid bandlimiting, electro-optical crosstalk between neighbouring Sels, proudly specifying MTFs that have dropped by no more than 25% at the Nyquist-Rate (CCD 3000, Fairchild, 1984). From a videometric point of view this is quite unfortunate, because aliasing effectively *reduces* the accuracy achievable with CCD-sensors.

described as an attenuation of high spatial frequencies identical to defocussing ($\Delta x = 6\ \mu m$ in Fig. 8) and a subsequent repetition of the MTF(u,v) with the rates $1/s_x$ and $1/s_y$, see Figure 11.

3.3 Electrical transfer function

The setup in Figure 9, with a vertical line x-centered in the light-sensitive square of a Sel, was also used to measure the horizontal impulse response of

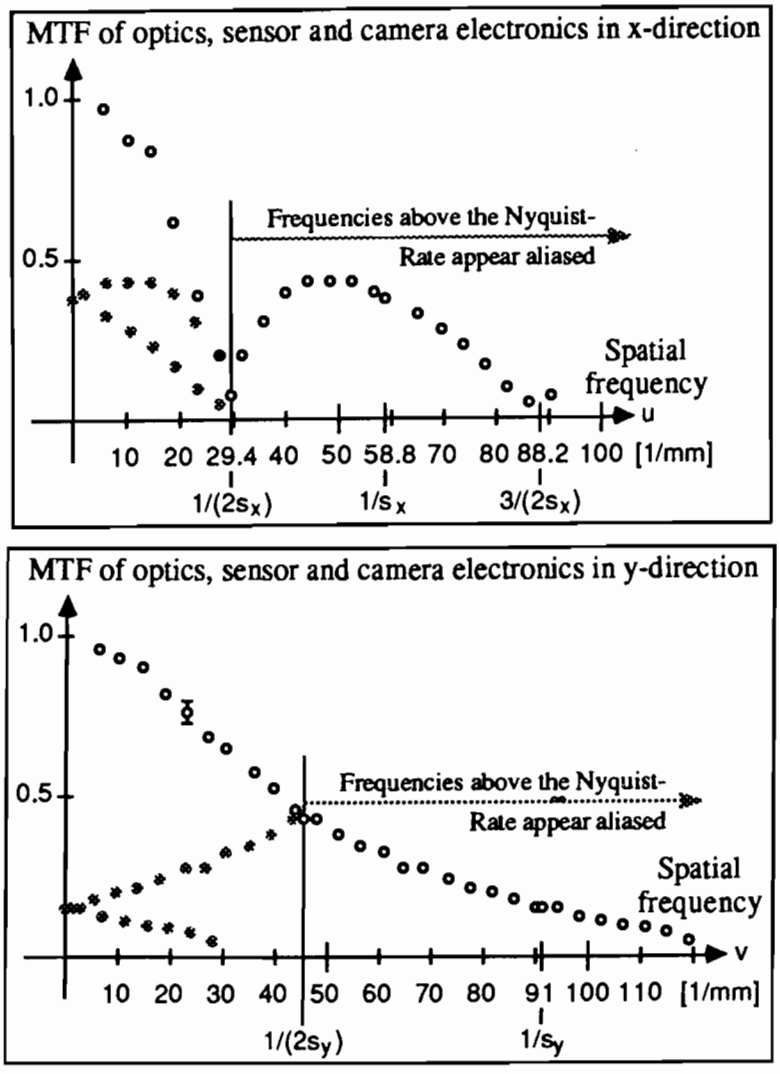


Fig. 11. Modulation transfer functions in x/u and y/v, measured for a line grid (rectangular profile, ≈ 2/mm, duty cycle 50%, harmonics neglected) at varying distances (20 cm to 3 m at increments of ≈ 10 cm) from the camera, equipped with a 50-mm-reproduction lens (@$f/5.6$). The spectral repetition with the rate $1/s_x$, $1/s_y$ leads to aliasing of frequencies above $u=1/(2s_x)$ and $v=1/(s_y)$, which are observed at $u'=1/s_x-u$ and $v'=1/s_y-v$. The electrical lowpass is most effective at odd multiples of $u=1/(2s_x)$, whereas the MTF(v), which is approximately the envelope for MTF(u) due to $I_y=I_x$, remains unaffected. In absence of lens errors, this envelope should correspond to the theoretical curve MTF ($u,\lambda=600$ nm, $F=5.6$, $\Delta x=6$ μm) in Fig. 8.

the WV-CD50 (Fig. 12), resulting from a built-in high order electrical lowpass with a cutoff-frequency of ≈ 4.7 MHz, corresponding to the spatial frequency $u=1/(2s_x)$. Due to the position of the electrical lowpass in the system (see Fig. 7), it cannot avoid aliasing caused by spatial sampling through the sensor — it's already too late. However, it still serves two useful purposes:

(a) All frequencies above the equivalent of $u=1/(2s_x)$ coming from the

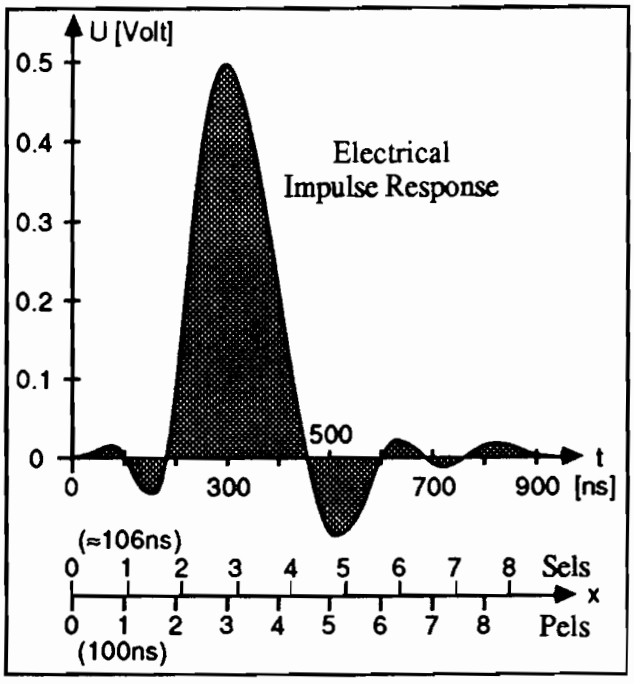


Fig. 12. Electrical response to an optical impulse/point (the dark-level from Fig. 10 is subtracted). It includes the effect of the Sample&Hold circuitry in Fig. 2. Depending upon interpretation, the impulse response in y is either Dirac- or rectangular shaped. Both are periodically space-variant, see text.

sensor can only be noise (caused by amplifiers etc.), which is eliminated by the filter, and

(b) Due to bandlimiting, no further aliasing is introduced by sampling through the following A/D-converter, since $s_x > p_x$ (for our system $f_p = 10$ MHz).

3.4 Periodic space variancy

Unsufficient optical low-pass filtering before spatially fixed sampling gives rise to a periodically spacevariant impulse response in x and y, that is, the electrical impulse response in x (Fig. 12) always appears at the same location relative to the Sel raster, independent of where a sensor element was illuminated within its active region. Therefore, without optical bandlimiting there will necessarily be an ambiguity of $\pm I_x/2 (\pm I_y/2)$ when trying to locate a point source (even if greyvalue interpolation is sued), or even worse, no camera output at all if the blind region of size $(s_x - I_x)$ by $(s_y - I_y)$ is hit.

In more detail, we will now investigate the influence of periodic space-variancy of the impulse response on the localization error of an input step function. Since the camera analog output is sufficiently lowpass filtered before sampling by the A/D-converter and could, in theory, be reconstructed from the digitized image (apart from greyvalue quantization noise, which is almost

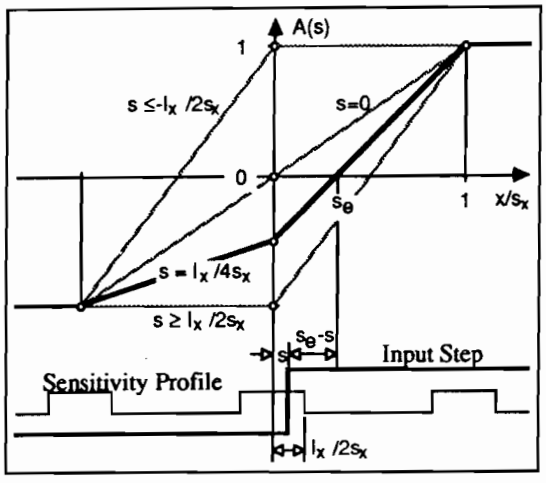


Fig. 13. The error $s_e - s$ (here normalized with respect to s_x) made in estimating the relative location s of an input step function using greyvalue interpolation in a noiseless sampled image depends upon the effective sensitivity profile, obtained by convolving the optical point spread function (PSF) with the sensor sensitivity profile. The worst case, 'perfect' optics (Dirac-PSF) and an 'ideal' sensor profile shown above, leads to a maximum error of at least $(s_x - I_x)/2$, independent of the interpolation algorithm used.

negligible for an 8-bit ADC in comparison to other sources of noise, as shown later), we may again neglect the Pels and only consider Sels, as in the derivation of eqs. 19ff. Let us assume that we have compensated for the delay caused by the lowpass filter [which, in fact, is swallowed by the principal point coordinate c_x in (4)] and, in a two-dimensional array of numbers, have perfectly reconstructed the charges accumulated by each sensor element. For a one-dimensional, only in y varying, image, this would correspond to the camera output voltage (which is then constant within a TV-line) as a function of the integer line number.

The derivation is carried out for the x-coordinate only, but is of course valid for the y-coordinate as well. As shown in Figure 13, the *shift* s of an input step function relative to the center of the light sensitive area is transformed into the output *amplitude* $A(s)$ of the Sel in question.

Without optical bandlimiting and an idealized rectangular sensor sensitivity profile we have:

$$A(s) = -2s/I_x \text{ for } |s| \leq I_x/2,$$

$$A(s) = 1 \text{ for } s < -I_x/2 \tag{25}$$

$$A(s) = -1 \text{ for } s > I_x/2$$

for a normalized step from $A = -1$ to $A = +1$. The difference $s_e - s$ between actual step location s and linearily interpolated intersection s_e with an imaginary threshold at $A = 0$ is a function of s and has a maximum of at least $\pm (s_x - I_x)/2$ at $|s| = I_x/2$ (blind zone ambiguity). Due to the nearly trian-

gular shape of $s_e - s = f(s)$ an integral similar to (19) leads to a boundary estimation RMSE σ_b of

$$\begin{aligned} \sigma_{bx} &= (s_x - I_x)/\sqrt{12} \quad (\approx 3.2\,\mu\text{m}) \text{ for } x, \\ \sigma_{by} &= (s_y - I_y)/\sqrt{12} \quad (\approx 1.4\,\mu\text{m}) \text{ for } y \end{aligned} \tag{26}$$

more or less independent of the interpolation algorithm. A 3rd order cubic spline interpolation using the amplitudes of 4 Sels instead of only 2 comes slightly closer to the bound given by (26), but is more susceptible to noise. Latter is due to the fact, that high-order interpolation polynomials usually have negative coefficients and therefore the sum of their squares (relevant for superposition of uncorrelated noise) is greater than for linear interpolation.

In case of Dirac-sampling (with $I_x/s_x (I_y/s_y)$ approaching zero) there is no advantage obtained by using greyvalue interpolation in comparison to optimally thresholded binary images. In contrast, with proper bandlimiting of spatial frequencies above $1/(2s_x)$ the system would in effect become *space-invariant* and in absence of noise the error $s_e - s$ could be reduced to zero. Unfortunately, adequate optical bandlimiting requires very small apertures ($f/57$ @ $\lambda = 600$ nm for $s_x = 17\,\mu$m), resulting in a prohibitively large loss of light. Nonredundant arrays using rectangles as aperture elements put into a $f/1.4$-lens can in theory achieve $f/57$ in x, $f/37$ in y with an intensity attenuation corresponding to $f/8$, but are somewhat impractical. Buying a very cheap lens with built-in bandlimiting might be the better solution. Single chip consumer color CCD cameras employ an optical low-pass filter based on birefringent crystal plates.

It is interesting to note that the accuracy ratio between x- and y-axis is not given by the ratio $s_x/s_y \approx 1.5$ of the Sel pitch, but rather by the ratio $(s_x - I_x)/(s_y - I_y) \approx 2.2$ of the blind zones. Together with the uncompensatible portion of the line jitter and an asymmetrical impulse response, this will make the x-axis accuracy inferior by a factor of about 3, which is consistent with practical experiences made by the authors with interline transfer cameras. As another rule of thumb one can say, that in high-contrast binarized images the bound given by (22) is nearly reached in y-direction, an improvement of a factor somewhere between 2 and 3 is obtained in both axis using greyvalue analysis and large F-numbers ($F = 11$ or higher).

3.5 Model errors

Another source of error is an asymmetrical impulse response in x-direction, causing e.g. a slightly unsymmetrical shift of the left and right boundary of a calibration point when decision thresholds are varied, see Dähler 1987 and Lenz 1987a for more details. Inhomogeneous illumination and/or lens vignetting have similar effects. The response of the WV-CD50 in Fig. 12 is rather

well phase compensated and therefore nearly symmetrical. One camera we investigated, the Javelin JE2063C, automatically switches to a *different filter* at low light levels (not just Automatic-Gain-Control AGC), making the image ('automatically' shifted to the right by ≈ 1 Sel) look less noisy. Seen from the computer, it therefore has a light level dependent principal point location.

3.6 Linearity

Due to their operating principle, direct conversion of photons to electrons, the linearity of CCD cameras seems to be excellent (if their Gamma correction can be turned off, which electrically 'corrects' their immanent $\gamma = 1$ to $\gamma \approx 0.65$, the inverse of that given by the logarithmic relationship between Wehnelt cylinder voltage and beam current of cathode ray tubes in TV monitors). Negligible deviations from the ideal behavior, probably mostly due to our aged Kodak-Greyscale No. Q-14, were found when testing the WV-CD50 over a dynamic range of 32:1, see Figure 14. Similar results were obtained by Curry 1986 for the CID-camera TN2200 from General Electric.

3.7 Noise

The last source of error treated in this article is noise. It was measured for the WV-CD50 by subtracting defocussed homogeneous images, digitized un-

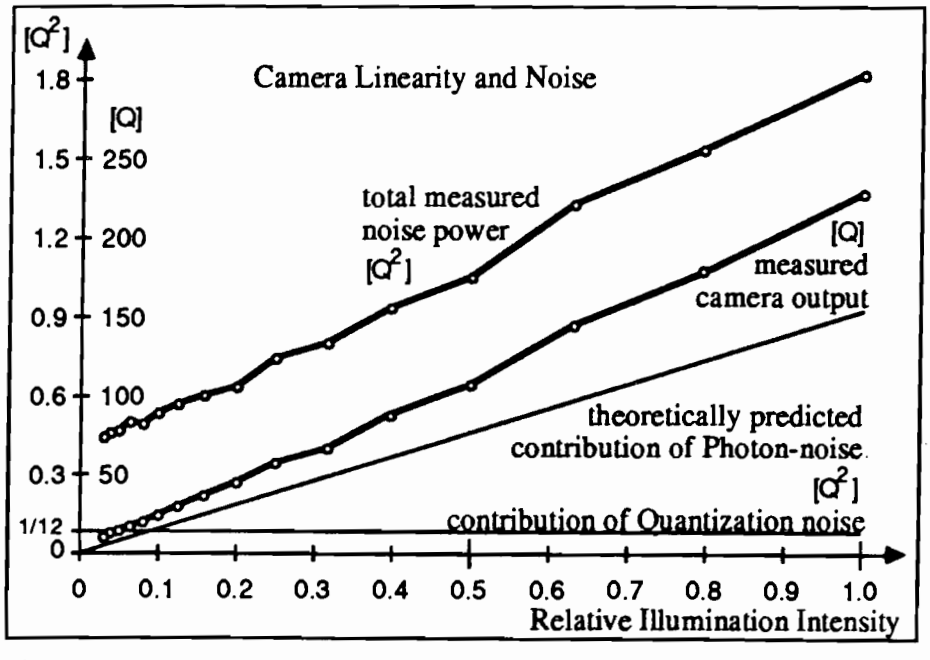


Fig. 14. Camera linearity and noise was measured with the Kodak-Greyscale from optical density 0 to 1.5 steps of -1 dB, illuminated with 300 lux. The linear relationship between intensity and noise power leads to the assumption, that most of the noise is caused by Poisson statistics of the photo-electrons, and only about $(0.4 - 1/12) \approx 0.3\ Q^2$ is intensity-independent, constant background/amplifier noise.

der identical conditions. In contrast to Vidicon cameras, where the Signal/Noise ratio is primarily limited by thermal noise in the input impedance of the first high-bandwidth amplifier stage, CCD sensors seem to come close to a theoretical bound, the Poisson statistics of the finite number of photo-electrons gathered by the sensor elements. This is indicated by the fairly large amount of linearily light intensity dependent noise power seen in Figure 14 and supported by theoretical predictions, based on the recommended illumination 300 lux at $f/1.4$ for 100% camera output. The number N_{e^-} of photo-electrons gathered by one Sel during the integration period of 40 msec is:

$$N_{e^-} \approx \frac{\begin{array}{c}\text{Illumination} \times \textit{White paper reflectivity} \times \text{Active Sel area} \times \\ \text{Wavelength} \times \text{Integration time} \times \text{Quantum efficiency}\end{array}}{\begin{array}{c}\text{Photometric radiation equivalent (@555 nm)} \times 4 \times \text{F-number}^2 \times \\ \text{Planck's constant} \times \text{Speed of light}\end{array}}$$

$$N_{e^-} \approx \frac{\begin{array}{c}300\text{lumen/m}^2 \times 0.5 \times (6\,\mu\text{m} \times 6\,\mu\text{m}) \times 555\text{ nm} \\ \times 40\text{ msec} \times 0.5\text{ e}^-/\text{Photon}\end{array}}{\begin{array}{c}680\text{lumen/W} \times 4 \times 1.4^2 \times 6.6 \times 10^{-34}\text{Wsec}^2/ \\ \text{Photon} \times 3 \times 10^8\text{ m/sec}\end{array}} \approx 57000\text{ e}^- \quad (27)$$

The derivation is again left to the reader, some physical constants and hints came from Vieth (1974).

The illumination was adjusted such that the camera output voltage lead to a digitized value of 230 Q for step 0 on the Kodak-Greyscale ($-$OdB relative reflectivity), yielding a predicted Poisson noise power of $(230\ Q)^2/57000 \approx 0.93\ Q^2$ in terms of quantization steps Q of the digitizing 8-bit A/D-converter. The ADC quantization noise power of $Q^2/12 \approx 0.08\ Q^2$ [derivation as in (19)] is in very good approximation additive to the noise power of a Gaussian process if latter is bigger than $0.1\ Q^2$. (The additivity is not obvious, since two *quantized* images were subtracted in order to measure the noise - the ideal signal is not available.) Thus, in order to determine the camera noise alone, one can simply subtract the quantization noise power from the measured total noise power in Figure 14.

A total noise power of $\approx 1Q^2$ (@$110Q$ with rel. illum. $-3\text{dB} = 0.5$ from Fig. 14) at an assumed greyvalue slope of 50 Q/Sel will lead to a boundary estimation RMSE $\sigma_b \approx 1\ Q/(50\ Q/\text{Sel}) = 0.02\ \text{Sel} \approx (0.3\ \mu\text{m in } x)$ using greyvalue interpolation, small in comparison to errors caused by inhomogeneous illumination, asymmetrical impulse responses, incorrectly chosen thresholds, perodic space-variancy, perspective distortion, line jitter etc.. By choosing large calibration points, the center localization RMSE σ_c can be about one order of magnitude less than σ_b (22). Due to its random nature based on Poisson statistics however, this is a fundamental limit for the accuracy of Videometry.

4 CONCLUSIONS

This paper has shown that CCD area sensors are ideal image aquisition devices in terms of their geometrical accuracy and stability. The problems resulting from their analog signal interfacing to computers are of purely technical nature and should be easily eliminated.

The analysis with linear systems theory indicates their high resolution potential (e.g. 167 lp/mm), which is far beyond the limit (e.g. 29.4 lp/mm) given by the detector element spacing and thus the sampling theorem. This knowledge can be utilized for the development of photogrammetric CCD-cameras with up to 2000 by 1500 picture elements, using sensors which are available today.

REFERENCES

Beyer, H.A., 1987. Some aspects of the geometric calibration of CCD-cameras. Proc. ISPRS Int. Comm. Conf. Fast Processing of Photogrammetric Data, Interlaken, pp. 68–81.

Curry, S. et al., 1986. Calibration of an array camera. Photogram. Eng. Remote Sensing, 52: 627–636.

Dähler, J., 1987. Problems in digital image acquisition with CCD-cameras. Proc. ISPRS Int. Comm. Conf. Fast Processing of Photogrammetric Data, Interlaken, pp. 48–59.

Fairchild, 1984. Fairchild Charge Coupled Device (CCD) Catalog - 1984.

Grün, A., 1987. Towards real-time photogrammetry. Photogrammetria, 42: 209–245.

Hill, J. et al., 1980. Machine intelligence research applied to industrial automation. 10. Report to the National Science Foundation, SRI Project 8487, pp. 75–105.

Lenz, R.K. and Tsai, R.Y., 1986. Techniques for calibration of the scale factor and image center for high accuracy 3D machine vision metrology. IBM Research Report RC 54867.

Lenz, R.K., 1987a. High accuracy feature extraction using chain-code in greyvalue images. IBM Research Report RC 56811.

Lenz, R.K., 1987b. Lens distortion corrected CCD-camera calibration with co-planar calibration points for real-time 3D measurements. Proc. ISPRS Int. Comm. Conf. Fast Processing of Photogrammetric Data, Interlaken, pp. 60–67.

Luhmann, T., 1987. On geometric calibration of digitized video images of CCD arrays. Proc. ISPRS Int. Comm. Conf. Fast Processing of Photogrammetric Data, Interlaken, pp. 35–47.

Tsai, R.Y., 1985. A Versatile Camera Calibration Technique for High Accuracy 3D Machine Vision Metrology using Off-the-Shelf TV Cameras and Lenses. IBM Research Report RC 51342.

Vieth, G., 1974. Meßverfahren der Photographie. Focal Press, London.