On the Accuracy of Videometry

Reimar Lenz, Institute for Telecommunications *
Dieter Fritsch, Institute for Photogrammetry
Technical University of Munich, Arcisstraße 21, D - 8000 München 2
Federal Republic of Germany

Abstract: For accurate computer vision based on standard video signals, the term **Videometry** is introduced (PLATZER 87). 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 (15msec 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 non-linear optimization (self-calibration with bundle adjustment) using the results from the linear approach as initial guess. In order to obtain a feature localization error well below sensor element resolution, rather large calibration points for boundary averaging and a special chain code operating in greyvalue images are used (LENZ 87a).

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 17μm and 582Sels vertically, spaced at 11μm), digitized with several frame grabbers (Imaging Technologies AP512, Kontron IBAS II and Matrox PIP-1024A, all with 512x512 Picture Elements (Pels)). In detail, the following was investigated:

Geometrical Camera Model
Exterior Parameters: Rotation, Translation
Inner Parameters: Principal Distance, Lens Distortion, Principal Point, Scaling Factors
Analyzed Model Errors:
Line Jitter, Spatial Quantization, Center Offset caused by Perspective Imaging & Lens Distortion, Sensor distortion
Neglected Model Errors:
5th Order Radial Lens Distortion, Tangential Lens Distortion, Thermic Camera Instability

Signal Transfer Model
Optical Transfer Function: Diffraction, Defocussing, Phase Errors of Lens Surface
Sensor Transfer Function: Local Integration, Sampling, Linearity
Electrical Transfer Function (x-direction only): Sample & Hold, Lowpass-Filtering, Sampling
Random Noise: Photon Noise, Amplifier Noise, Quantization Noise
Fixed Noise: Sensor Noise, A/D-Converter Noise, Computer Noise
Analyzed Errors: Periodically Space-Variant & Asymmetrical Impulse Response

The Geometrical Camera Model

Calibration of the geometrical camera model: First, the model in Fig.1 will be briefly described. It originates in work from TSAI 85 and was modified by LENZ 87b to allow for a fully linear calibration algorithm capable of real-time performance.

The geometrical imaging process may be subdivided into four steps (here, parameters are printed bold):

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

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

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

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

* This work was partially funded by Kontron Bildanalyse GmbH in Eching near Munich, FRG.
Early research results were obtained during a visit at the IBM Research Labs, Yorktown Heights, NY, USA.

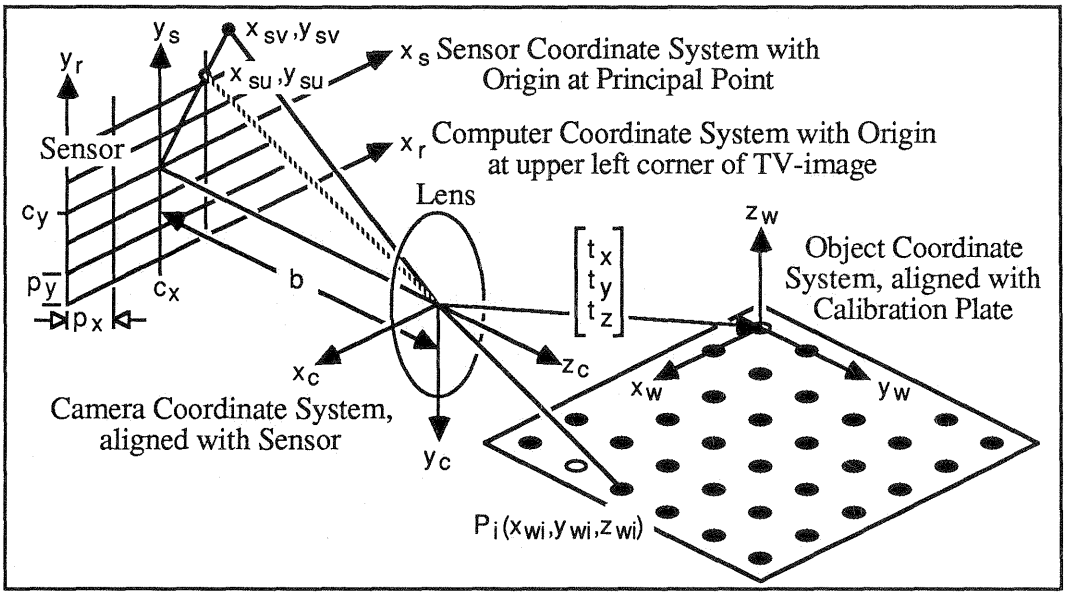

y_s
x_{sv},y_{sv}
y_r
x_{su},y_{su}
x_s Sensor Coordinate System with Origin at Principal Point
Sensor
x_r Computer Coordinate System with Origin at upper left corner of TV-image
Lens
c_y
p_y
c_x
p_x
b
t_x
t_y
t_z
z_w
Object Coordinate System, aligned with Calibration Plate
x_c
z_c
x_w
y_w
Camera Coordinate System, aligned with Sensor
y_c
$P_i(x_{wi},y_{wi},z_{wi})$

Fig. 1: Geometrical Camera Model with 6 exterior and 6 inner Parameters

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

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

$$x_{sv} = 2x_{su}/(1+[1-4k_3\, r_{su}^2]^{1/2}); \quad y_{sv} = 2y_{su}/(1+[1-4k_3\, r_{su}^2]^{1/2}); \quad r_{su}^2 = x_{su}^2 + y_{su}^2 \qquad (3b)$$

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

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

$$x_r = x_{sv}/\mathbf{p}_x + \mathbf{c}_x\,; \qquad y_r = y_{sv}/\mathbf{p}_y + \mathbf{c}_y\,; \qquad (4a)$$

$$x_{sv} = (x_r - \mathbf{c}_x)\cdot \mathbf{p}_x\,; \qquad y_{sv} = (y_r - \mathbf{c}_y)\cdot \mathbf{p}_y\,; \qquad (4b)$$

($\mathbf{c}_x$, $\mathbf{c}_y$) are the frame buffer or computer coordinates of the principal point, $\mathbf{p}_x$ ($\mathbf{p}_y$) is the distance on the image sensor between two horizontally (vertically) adjacent frame buffer (**not** sensor) Pels. These last four intrinsic parameters are assumed to be already known for the camera calibration procedure described in the following. TSAI 85 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}:

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

When using a *coplanar* set of calibration points P_i with known coordinates (x_{wi},y_{wi},z_{wi}), the object 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)

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

of a homogeneous system of equations $\mathbf{Av} = \mathbf{0}$ which is solved in a least squares sense for the non-trivial 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 3x3 orthonormal matrix, all elements of **v** must be scaled with a common constant such that

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

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

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

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

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

with $\quad x_{ci} = x_{wi}\, r_{xx} + y_{wi}\, r_{xy} + t_x\,; \quad y_{ci} = x_{wi}\, r_{yx} + y_{wi}\, r_{yy} + t_y\,; \quad r_{svi}^2 = x_{svi}^2 + y_{svi}^2\,;$

Because b and t_z must be positive, we can now resolve all sign-ambiguities by multiplying

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

Obtaining r_{xz}, r_{yz}, r_{zz} with the outer product completes the calibration, using only physically meaningful, inpendent 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 σ =0.18µm were reached.

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 84 specifies ±5ppm cumulative pitch error for the electron beam written mask, ±0.016µm between any two adjacent Sels and -0.23% to -0.46% isotropic contraction due to subsequent high temperature processing steps. Because latter affects x- and y- direction in the same fashion, the - more important - ratio between horizontal and vertical Sel-pitch is known to ±10ppm. 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%±.04% in x-direction by measuring a distance of 6798µm between two sensor elements 400Sels apart, the camera in Fig.8 mounted on a micrometer stage with ±3µm measurement error. (Measuring the distance between 500Sels 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 500Sels•17µm/Sel = 8.5mm to 8.25mm (8.8mm were specified). Falsely specifying important features (Sel# and 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 non pixel-synchronized, TV standard based systems, the Sels with the pitch s_x are read out and sampled&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

$$p_x = s_x \cdot f_s / f_p \approx s_x \cdot \text{(Number of Sels per Line / Number of Pels per Line)} \qquad (11)$$

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 512Pels, 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. 455cycles/line for the Fairchild CCD3000 with 380Sels/line), the average ratio f_s/f_p is fixed for such systems (e.g. 455/640≈.711, which is only approx. #Sels/#Pels=380/512≈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 87) have a continuously running crystal oscillator with $4f_p$ 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 $\Delta f_s/f_s \approx 2\%$ (Aqua-TV HR600), fully affecting the horizontal scale factor p_x, has been observed in a thorough investigation carried out by DÄHLER 87. 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 GRUEN 87 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 2ppm/°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 & TSAI 86. 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 if $2f_s > f_p$). 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 (≈20mV) was added with a bandpass filter, see Fig.2. A n=512-point one-dimensional Fourier analysis of a vertically averaged, low light-level blank image ($I_{av}(x_r) = \Sigma$ 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

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

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

$$\underline{U}'(m) = u_a \cdot e^{-j[\pi(m-f_a) + \varphi_a]} \cdot \sin[\pi(m - f_a)] / [2\pi(m - f_a)(m - f_a - 1)(m - f_a + 1)], \qquad j = \sqrt{-1} \qquad (13)$$

of the discrete complex FFT-spectrum $\underline{U}(m, 0 \leq m < n)$ 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 and to solve for f_a:

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

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

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

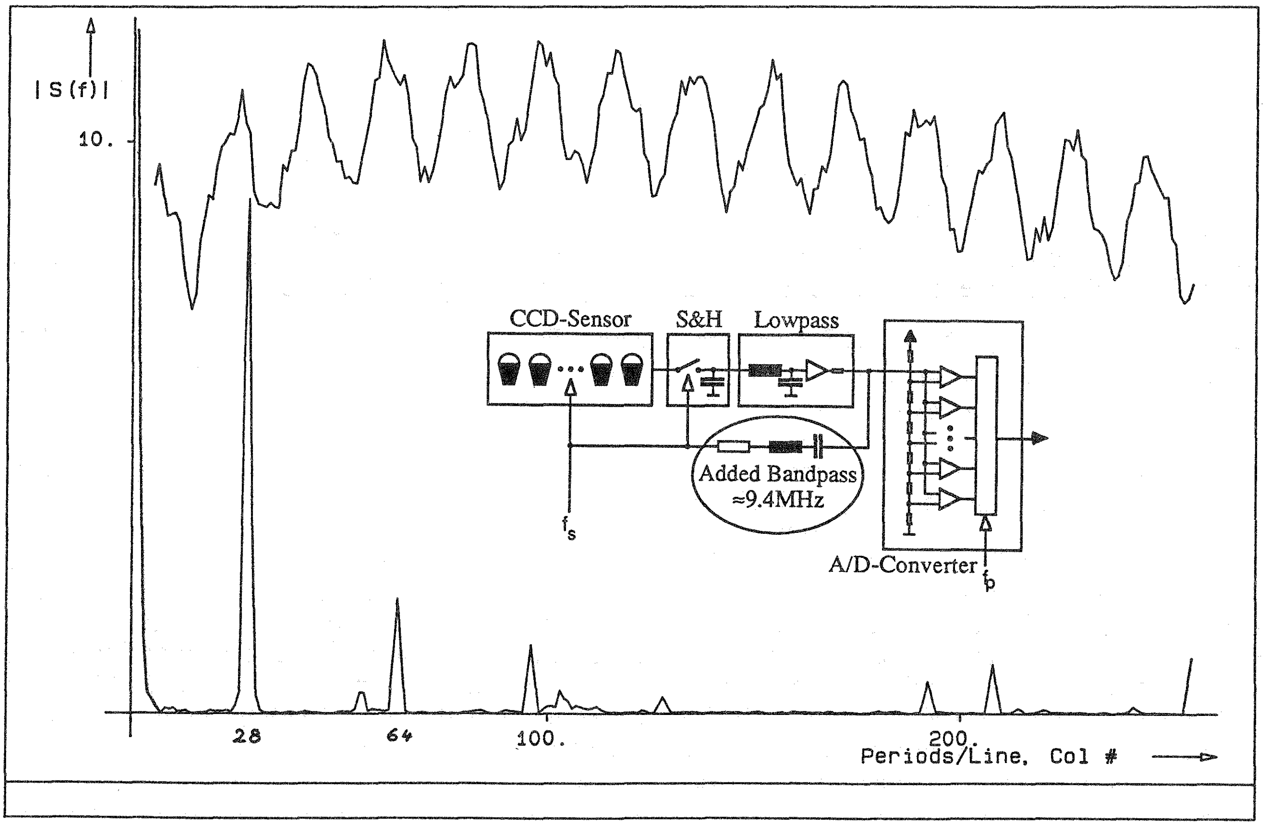


Fig. 2: The Sensor element (Sel) clock signal, added as reference to the video signal with the circuitry shown, can be observed aliased in a TV-line digitized with the Picture element (Pel) clock rate. The horizontal scale factor and line jitter are determined by phase and amplitude analysis of the peak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

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

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

In (16), f_a is taken real-valued from (14), not integer. Again, raised cosine weighting is used. The phase $\varphi(y_r)$ should be accumulated from line to line (both TV-fields individually) to cope with the ambiguity of 2π, which corresponds to a shift of 1Pel. Fig.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 1/4Pel ($\pi/2$). PLL control oscillations similar to those observed by LUHMANN 87 using optical line synchronization can be seen. For the camera and frame grabber analyzed, the behavior in Fig.3 seems to be long- and short-term stable within 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 $\Delta\varphi(y_r)/\Delta y_r$. The relative global variation from top to bottom is $\approx(\pi/2)/(256\text{lines/field}\cdot 2\pi\cdot 640\text{cycles/line})<2\text{ppm}$ 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 $\Delta p_x/p_x \approx 1/40\text{Pel/line}/(640/\text{line}) \approx 40\text{ppm}$ was found). Systems with free-running oscillators are not subject to frequency modulation.
Other sources of noise can be detected in the spectrum in Fig.2: the peaks at 64,128,256periods/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. 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.

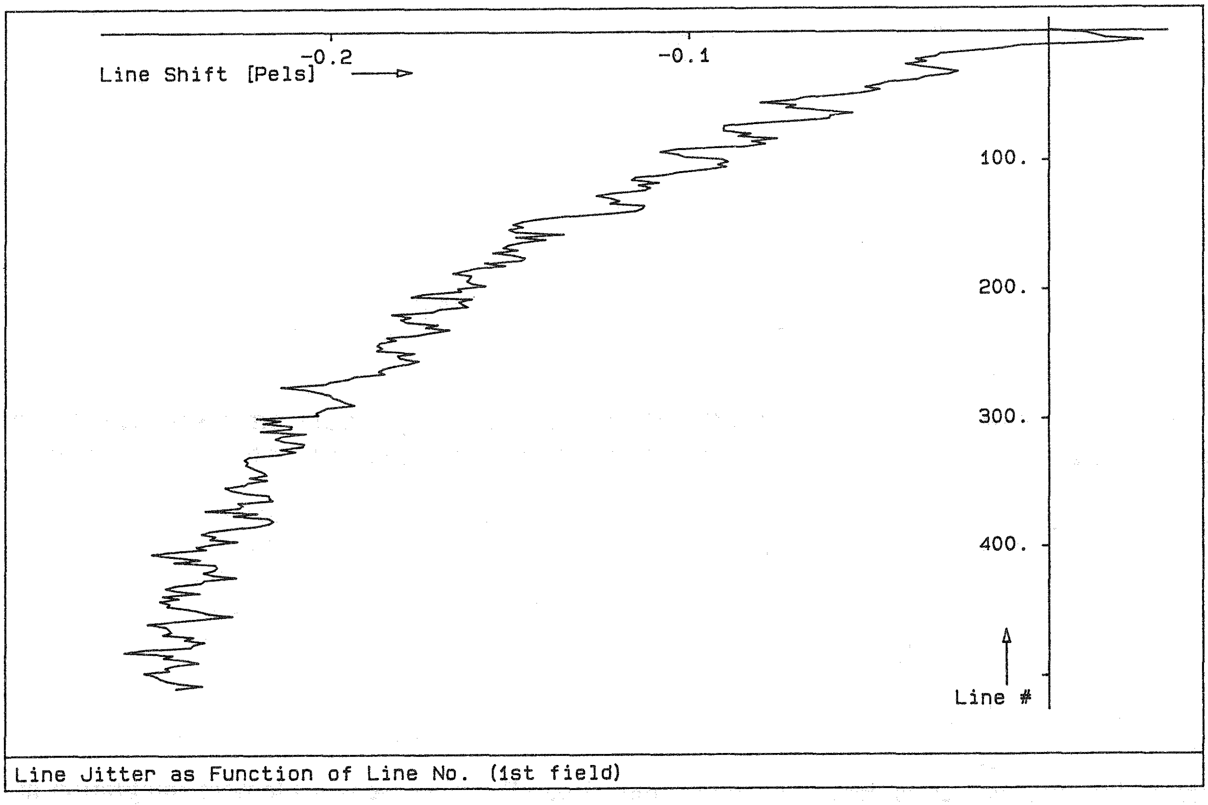
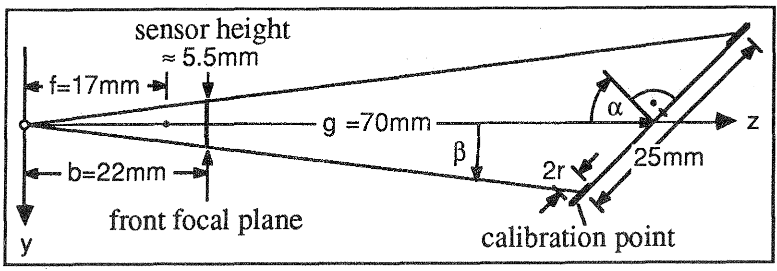


Fig. 3: Line jitter due to imperfect synchronization between camera and frame grabber, 1st field. Most errors are caused by the loss of perfect sync during the vertical blanking period, much less result from PLL control oscillations.

More 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 Fig.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 Fig.5 becomes approximately

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

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 Fig.4 (r=1mm, α=45° and tanβ=1/8), the error amounts to 3.2µm and is by no means negligible.



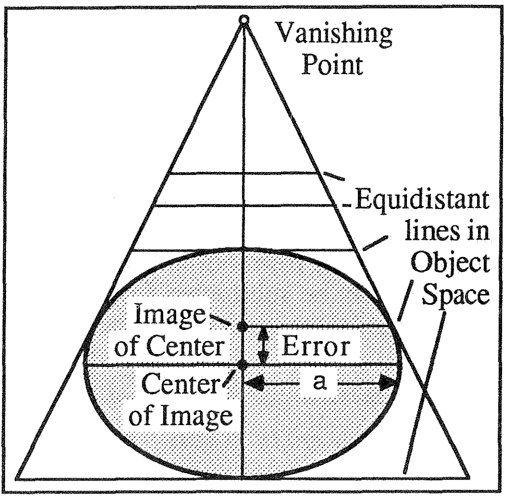


Fig. 4: Close-range videometric setup to evaluate systematic errors due to calibration point deformation by central perspective imaging and/or radial lens distortion. Both errors increase with the square of point image size 'a'.

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.

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

$$r_S - r_{sv} \approx 3\, r_{sv} k_3 a^2 \qquad \text{with 'a' as in (17)} \qquad (18)$$

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

Some non-systematic errors: We will now treat the effect of spatial sensor quantization on calibration point localization accuracy. On the basis of Fig.12, showing the quantized image of a calibration point, we will derive the mean square measurement error (MSE) σ_c^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 locaction 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

$$\sigma_b^2 = \int_{x=-s_x/2}^{s_x/2} x^2 dx \Big/ \int_{x=-s_x/2}^{s_x/2} dx = s_x^2/12 \qquad \text{or} \qquad \sigma_b \approx 0.289\, s_x \quad \text{for binary images} \qquad (19)$$

In greyvalue images, σ_b^2 is dependent upon optical bandlimiting, 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:

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

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

$$x^2_{ave} = \int_{y=0}^{r_y} (r_x^2 - y^2 (r_x / r_y)^2)\, dy \Big/ \int_{y=0}^{r_y} dy = \frac{2}{3}\, r_x^2 \qquad (21)$$

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 80, the *variances* $\sigma_{c(ave)}^2$ are added:

$$\sigma_c^2 = 4 n_y\, \sigma_{c(ave)}^2 = 4 n_y \frac{2}{3}\, r_x^2\, (\sigma_b / \pi\, r_x\, n_y)^2 = \frac{8}{3 n_y} \left(\frac{\sigma_b}{\pi}\right)^2 \qquad \text{or} \qquad \sigma_c \approx \frac{0.15 s_x}{\sqrt{n_y}} \quad \text{for binary images} \qquad (22)$$

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 Fig.12, $\sigma_c \approx 0.04 s_x$. 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.

The Signal Transfer Model

In Fig. 6, 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.

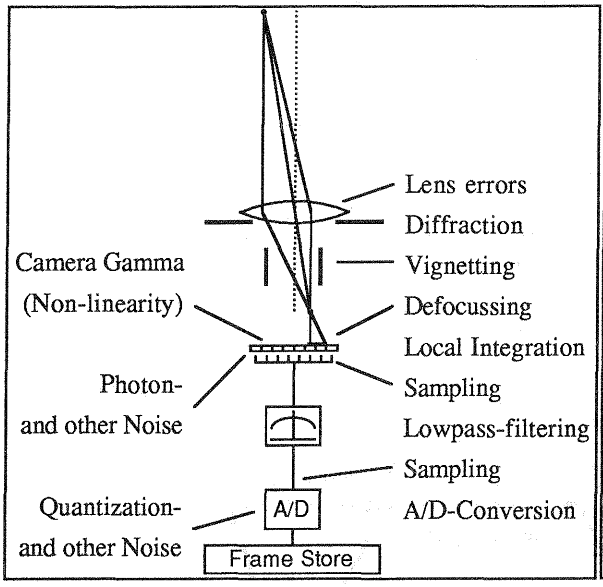


Fig. 6: Some factors affecting the signal transfer function (see the survey of GRUEN 87 for more sources of radiometric degradation)

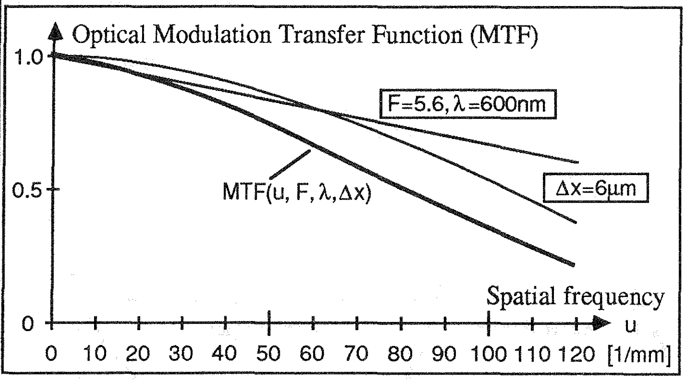


Fig. 7: Theoretical modulation transfer functions due to diffraction and defocussing. MTF attenuation due to phase errors of the lens surfaces is nearly unpredictable and not included in this figure.

If, for reasons of simplicity, a square shaped aperture aligned with the sensor array is assumed, x- and y-axis 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

$$\mathrm{MTF}(u,\lambda,F) = 1 - |u|\cdot\lambda\cdot F \quad \text{for } |u| < 1/\lambda F, \text{ else} \quad \mathrm{MTF}(u,\lambda,F) = 0 \tag{23}$$

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

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

The MTFs for λ=600nm, F=5.6 ($|u_{max}|\approx 300/\mathrm{mm}$) and Δx =6μm ($\sin(\pi\cdot 6\mu m\cdot n\cdot 167/\mathrm{mm})=0$) and their product are shown in Fig.7. These quantities were chosen looking forward to the experiments described later.
Next, the photons hitting the light sensitive portion of a sensor element cell are partially converted into photo-electrons (≈50% quantum efficiency, FAIRCHILD 84) 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.9) within a Sel cell with the setup in Fig.8 designed by PLATZER 88.
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 $l_x=l_y=6\mu m$ and sampling can be described as an attenuation of high spatial frequencies identical to defocussing (Δx=6μm in Fig.7) and a subsequent repetition of the MTF(u,v) with the rates $1/s_x$ and $1/s_y$, see Fig.10.
The setup in Fig.8, with a vertical line x-centered in the light sensitve square of a Sel, was also used to measure the horizontal impulse response of the WV-CD50 (Fig.13), resulting from a built-in high order electrical lowpass with a cutoff-frequency of ≈4.7MHz, corresponding to the spatial frequency $u=1/(2s_x)$. Due to the position of the electrical lowpass in the system (see Fig.6), it cannot avoid aliasing caused by spatial sampling through the sensor - it's already too late. However, it still serves two useful purposes:
1) All frequencies above the equivalent of $u=1/(2s_x)$ coming from the sensor can only be noise (caused by amplifiers etc.), which is eliminated by the filter, and 2) 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=10MHz).

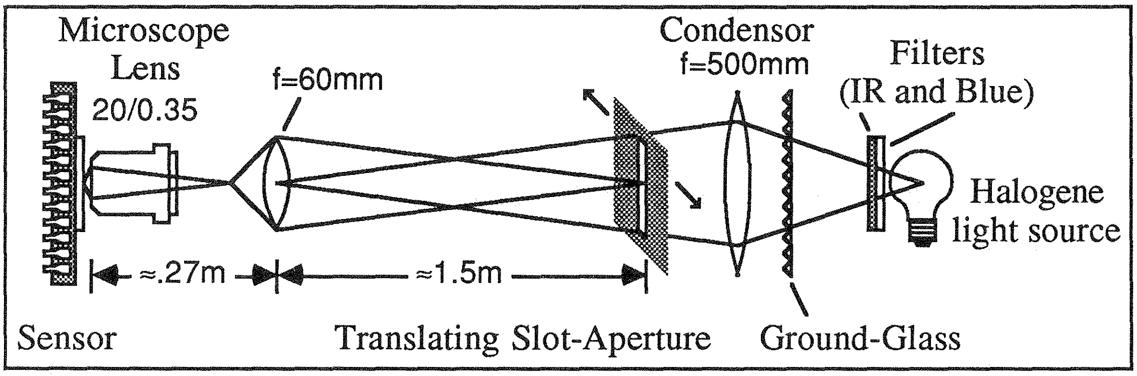


Fig. 8: Experimental setup used to measure the light sensitivity profile, impulse response, Sel-pitch and distortion of a CCD-camera (designed by PLATZER 88). The slot-aperture consisting of two opposing razor blades (25mm by 0.25mm, scaled down by a factor of ≈620 to 40µm by ≈2µm diffraction limited width-equivalent) was aligned with either the x- or y-sensor-axis and translated in increments of 0.5mm (0.81µm 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.

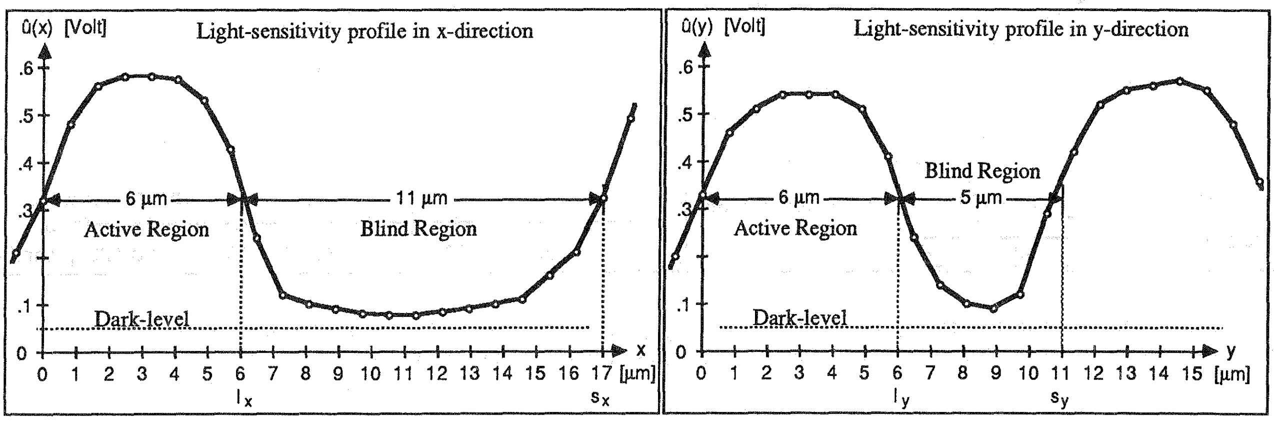


Fig. 9: Measured sensor sensitivity profile in x- and y-direction. Only l_x=6µm out of the horizontal pitch s_x =17µm and l_y = 6µm out of s_y=11µm of the interline transfer CCD-sensor used in the Panasonic WV-CD50 are light sensitive. This leaves less than 20% sensitive area, whereas frame transfer devices reach more than 50%. The gradual transition from sensitive to blind region is probably mostly due to the diffraction limited spatial bandwidth (≈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 84). From a videometric point of view this is quite unfortunate, because aliasing effectively *reduces* the accuracy achievable with CCD-sensors.

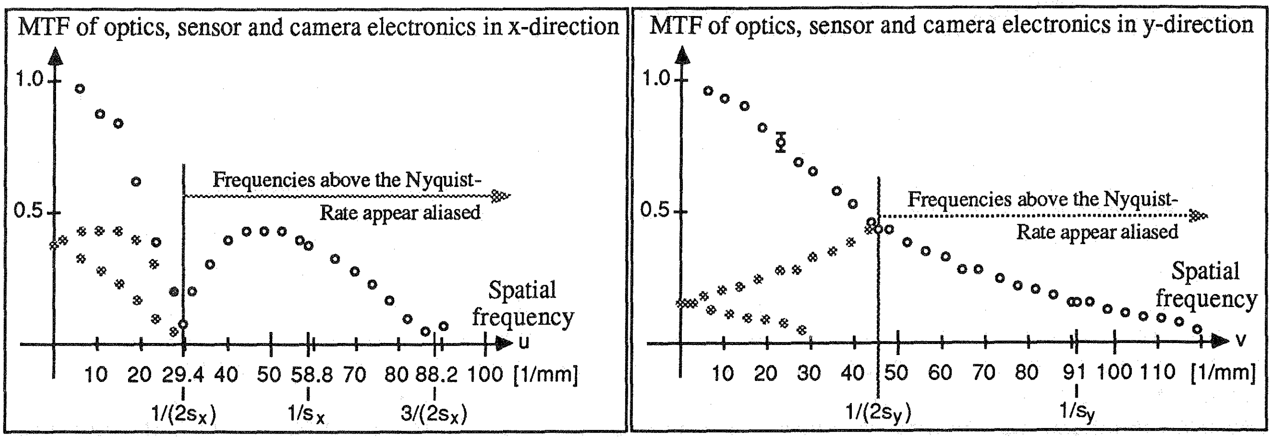


Fig.10: Modulation transfer functions in x/u and y/v, measured for a line grid (rectangular profile, ≈2/mm, duty cycle 50%, harmonics neglected) at varying distances (20cm to 3m at increments of ≈10cm) from the camera, equipped with a 50mm-reproduction lens (@ f/5.6). The spectral repetition with the rate $1/s_x$, $1/s_y$ leads to aliasing of frequencies above u=$1/(2s_x)$ and v=$1/(2s_y)$, 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 aproximately the envelope for MTF(u) due to l_y=l_x, remains unaffected. In absence of lens errors, this envelope should correspond to the theoretical curve MTF(u,λ=600nm,F=5.6,Δx=6µm) in Fig.7.

Unsufficient optical lowpass filtering before spatially fixed sampling gives rise to a *periodically space-variant* impulse response in x and y, that is, the electrical impulse response in x (Fig.13) 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 $\pm l_x/2$ ($\pm l_y/2$) when trying to locate a point source (even if greyvalue interpolation is used), or even worse, no camera output at all if the blind region of size (s_x - l_x) by (s_y - l_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 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-dimensionally, 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 Fig.11, 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.

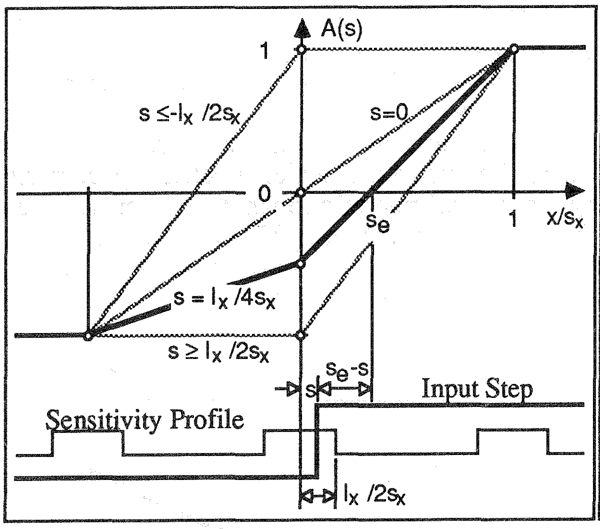


Fig.11: The error s_e-s (here normalized with respect to s_x) made in estimating the relative location s of an input step function using greyvalue interpolation in a noiseless sampled image depends upon the effective sensitivity profile, obtained by convolving the optical point spread function (PSF) 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 $(s_x - l_x)/2$, independent of the interpolation algorithm used. With proper bandlimiting through optics, photo-electron diffusion processes on the sensor or both, the camera would in effect become a space-*invariant* system, reducing the estimation error to the theoretical limits given by the ratio of signal to noise.

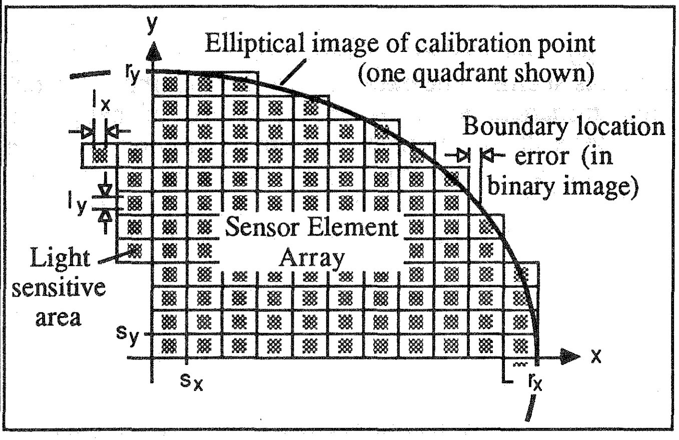


Fig.12: Spatially quantized image of a calibration point, used to derive the relationship between boundary estimation RMSE σ_b and center estimation RMSE σ_c as a function of calibration point size. For optimally thresholded binary images, the boundary location error is evenly distributed between $\pm 0.5 s_x$ (s_y).

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

$$A(s) = -2s/l_x \text{ for } |s| \le l_x/2, \qquad A(s) = 1 \text{ for } s < -l_x/2 \qquad \text{and} \qquad A(s) = -1 \text{ for } s > l_x/2 \qquad (25)$$

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 $\pm(s_x - l_x)/2$ at $|s| = l_x/2$ (blind zone ambiguity). Due to the nearly triangular shape of $s_e - s = f(s)$ an integral similar to (19) leads to a boundary estimation RMSE σ_b of

$$\sigma_{bx} = (s_x - l_x)/\sqrt{12} \ (\approx 3.2\mu m) \text{ for } x, \qquad \sigma_{by} = (s_y - l_y)/\sqrt{12} \ (\approx 1.4\mu m) \text{ for } y \qquad (26)$$

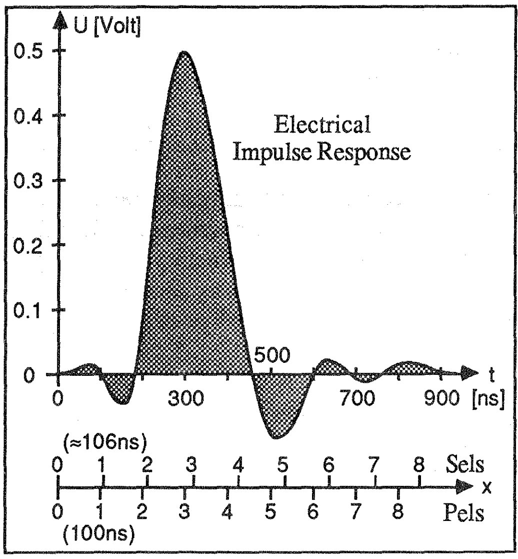
more or less independent of the interpolation algorithm. A 3rd order cubic spline interpolation using the amplitudes of 4 Sels instead of only 2 comes slightly closer to the bound given by (26), but is more susceptible to noise. Latter is due to the fact, that high-order interpolation polynomials usually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

In case of Dirac-sampling (with l_x/s_x (l_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 @ λ=600nm for s_x=17μm), resulting in a prohibitively large loss of light. Non-redundant 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.

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 $(s_x - l_x)/(s_y - l_y) \approx 2.2$ 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).

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 87 and LENZ 87a for more details. Inhomogeneous illumination and/or lens vignetting have similar effects. The response of the WV-CD50 in Fig.13 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 ≈1Sel) look less noisy.

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 γ=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 Fig.14. Similar results were obtained by CURRY 86 for the CID-camera TN2200 from General Electric.



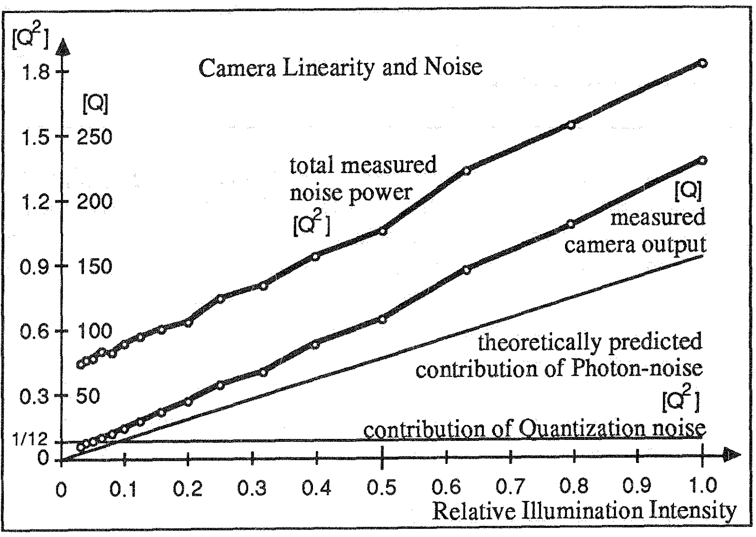


Fig.13: Electrical response to an optical impulse/ point (the dark-level from Fig.9 is subtracted). It includes the effect of the Sample&Hold circuitry in Fig.2. Depending upon interpretation, the impulse response in y is either Dirac- or rectangular shaped. Both are periodically space-variant, see text.

Fig.14: Camera linearity and noise was measured with the Kodak-Greyscale from optical density 0 to 1.5 in steps of -1dB, illuminated with 300lux. 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) ≈ 0.3Q^2 is intensity-independent, constant background/amplifier noise.

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 under 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 Fig.14 and supported by theoretical predictions, based on the recommended illumination 300lux at f/1.4 for 100% camera output. The number N_{e^-} of photo-electrons gathered by one Sel during the integration period of 40msec is:

$$N_{e^-} \approx \frac{\text{Illumination} \cdot \text{White Paper Reflectivity} \cdot \text{Active Sel Area} \cdot \text{Wavelength} \cdot \text{Integration Time} \cdot \text{Quantum Efficiency}}{\text{Photometric Radiation Equivalent(@555nm)} \cdot 4 \cdot \text{F-Number}^2 \cdot \text{Planck's Constant} \cdot \text{Speed of Light}}$$

$$N_{e^-} \approx \frac{300\text{lumen/m}^2 \cdot 0.5 \cdot (6\mu\text{m} \cdot 6\mu\text{m}) \cdot 555\text{nm} \cdot 40\text{msec} \cdot 0.5\text{e}^-/\text{Photon}}{680\text{lumen/W} \cdot 4 \cdot 1.4^2 \cdot 6.6 \cdot 10^{-34}\text{Wsec}^2/\text{Photon} \cdot 3 \cdot 10^8\text{m/sec}} \approx 57\,000\ \text{e}^- \qquad (27)$$

The derivation is again left to the reader, some physical constants and hints came from VIETH 74.
The illumination was adjusted such that the camera output voltage lead to a digitized value of 230Q for step 0 on the Kodak-Greyscale (-0dB relative reflectivity), yielding a predicted Poisson noise power of $(230Q)^2/57000 \approx 0.93Q^2$ in terms of quantization steps Q of the digitizing 8bit A/D-converter. The ADC quantization noise power of $Q^2/12 \approx 0.08Q^2$ (derivation as in (19)) is in very good approximation additive to the noise power of a Gaussian process if latter is bigger than $0.1Q^2$. (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 Fig.14.
A total noise power of $\approx 1Q^2$ (@110Q with rel. illum. -3dB=0.5 from Fig.14) at an assumed greyvalue slope of 50Q/Sel will lead to a boundary estimation RMSE $\sigma_b \approx 1Q/(50Q/Sel) = 0.02Sel \approx (0.3\mu m$ 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.

References:

BEYER, H.A., 1987: Some Aspects of the Geometric Calibration of CCD-Cameras, *Proceedings of the ISPRS Intercommission Conference on "Fast Processing of Photogrammetric Data"*, Interlaken, June 2-4, pp. 68-81

CURRY, S. et al., 1986: Calibration of an Array Camera, *Photogrammic Engineering and Remote Sensing*, Vol. 52, May 1986, pp. 627-636

DÄHLER, J., 1987: Problems in Digital Image Acquisition with CCD-Cameras, *Proceedings of the ISPRS Intercommission Conference on "Fast Processing of Photogrammetric Data"*, Interlaken, June 2-4, pp. 48-59

FAIRCHILD, 1984: Fairchild Charge Coupled Device (CCD) Catalog - 1984

GRUEN, A., 1987: Towards Real-Time Photogrammetry, *Invited Paper to the 41. Photogrammetric Week* Stuttgart, September 14-19

HILL, J. et al., 1980: Machine Intelligence Research applied to Industrial Automation, *Tenth Report to the National Science Foundation*, SRI Project 8487, pp. 75-105, Nov.

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*, Oct. 8

LENZ, R.K., 1987a: High Accuracy Feature Extraction using Chain-Code in Greyvalue Images, *IBM Research Report RC 56811*, Mar. 27

LENZ, R.K., 1987b: Lens distortion corrected CCD-camera calibration with co-planar calibration points for real-time 3D measurements, *Proceedings of the ISPRS Intercommission Conference on "Fast Processing of Photogrammetric Data"*, Interlaken, June 2-4, pp. 60-67
In German: Linsenfehlerkorrigierte Eichung von Halbleiterkameras mit Standardobjektiven für hochgenaue 3D-Messungen in Echtzeit, *Informatik-Fachberichte 149, Proc.9.DAGM-Symposium 1987*, Braunschweig, Sep.29 - Oct.1, Springer Berlin ISBN 3-540-18375-2, pp. 212-216

LUHMANN, T., 1987: On Geometric Calibration of Digitized Video Images of CCD Arrays, *Proceedings of the ISPRS Intercommission Conference on "Fast Processing of Photogrammetric Data"*, Interlaken, June 2-4, pp. 35-47

PLATZER, H., 1987 & 1988: Personal Communication

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*, May 8

VIETH, G., 1974: Meßverfahren der Photographie, *Focal Press*, London, ISBN 3-486-39611-0, p. 9