K. Linkwitz U. Hangleiter (Eds.)

High Precision Navigation

Integration of Navigational and Geodetic Methods

With 289 Figures

Springer-Verlag Berlin Heidelberg New York
London Paris Tokyo Hong Kong

Proceedings of an International Workshop

Stuttgart and Altensteig, May 1988,
organized by the Sonderforschungsbereich 228
(Special Collaborative Programme) of the
Deutsche Forschungsgemeinschaft

Precision of Geometric Features Derived from Image Sequences

Wolfgang Förstner
Institut für Photogrammetrie, Universität Stuttgart

Summary:
The paper discusses the accuracy potential of mono and stereo image sequences. Specifically treated are the effect of image blur onto the precision of image features, the precision obtainable for relative position, orientation and speed of the sensor platform with respect to a given coordinate frame or to other moving objects. The results can be used for designing a multisensor-navigation system.

0. Introduction

Sensing and analysing changing environment is a basic requirement in navigation and is the reason for the development of vision based navigation systems. Here vision sensors are understood to be all types of sensors which produce a one, two or three dimensional geometric projection or image of the world. In contrast to "zero"dimensional sensors, which only determine the position or orientation at single points, possibly over time, they are not blind to changes in the environment, do not necessarily need a map and, in case they use maps, do not require them to accurately represent reality during the movement of the vehicle. These restrictions of sensors used e. g. in the Global Positioning System (GPS) or in Inertial Navigation Systems (INS) have their counterpart in the higher efficiency of these sensors for determining the orientation which in general is a magnitude higher than the efficiency of orientation determination of vision-type sensors, both with respect to accuracy and speed.

On the other hand vision-type sensors derive geometric properties of the observed scene out of the images in any case, either as a prerequitite for scene interpretation, e. g. to determine free space, or as a byproduct, e. g. for monitoring the own trajectory and/or the movement of other vehicles as a basis for the planning of collision free paths.

This paper wants to discuss which basic geometric properties can be derived from sequences of optical images, taken e. g. with

a CCD-camera and what accuracies are obtainable solely from image sequences with the aim to get information what type of other sensors may be necessary to support vision based systems for high precision navigation. Specifically we want to discuss the effect of image blur onto the precision of image features, the precision obtainable for relative position, orientation and speed of the sensor platform with respect to a given coordinate frame or to other moving objects. Though the derivations are basic and general, the questions and examples are taken from the environment of a road vehicle.

1. On the Geometric Precision of Image Features

The derivation of the own trajectory or of the geometry of the observed scene has to be based on elementary geometric features which have to be found in a sequence of images. The precision of the derived geometric quantities in 3D depends on the accuracy with which the image features can be extracted and on the geometric relations between the objects and the trajectory. In this section we only deal with the precision of the extracted image features.

As we are interested in the potential of image sequence analysis for reconstruction 3D geometry we have to ask for the geometric precision inherent in the digital images. This precision is influenced by various factors, e. g. by

- the quality of the interior orientation of the camera, including principle point, focal length and lens distortion,
- the geometric quality of the sensor, mainly depending on the electronic properties of the camera and the frame grabber,
- the properties of the resolution of the sensor in terms of a modulation transfer function of simply the pixel distance,
- the electronic noise,
- the image blur induced by the non-zero integration time,

to name the main error sources.

Most of these effects can be reduced by proper calibration procedures and thus taken into account in the anlysis, especially the geometric properties of the camera. The resolution of images taken from moving sensors mainly depend on the image blur, which in this type of application because of the varying distance of

the object points to the sensor is inhomogeneous and anisotropic.

The following analysis therefore neglects all effects except for the image content, including texture, blurr and noise. It is closely related to the theoretical investigations on the accuracy of image matching (cf. FÖRSTNER 1982, 1984). This is because selecting optimal points for image matching using cross correlation is identical to selecting optimal feature points, like corners or centres of circular features, for image analysis (cf. FÖRSTNER 1986, FÖRSTNER/GÜLCH 1987).

1.1 The geometric precision of image features

The precision of locating a known object, or determining a feature point in a digital image is given by:

$$\sigma_x = \frac{1}{\sqrt{n}} \cdot \frac{\sigma_n}{\sigma_{f_x}} \tag{1}$$

where n is the number of pixels involved

σ_n is the noise standard deviation and

σ_{f_x} is the standard deviation of the first derivative of the given or the true signal $f(x)$.

There are other versions of the same relation, which give rise to various interpretations and which are useful in our context. With the signal to noise ration SNR

$$SNR = \sigma_f \; / \; \sigma_n \; , \tag{2}$$

the correlation coefficient ϱ between the measured and the given signal f and with the effective bandwidth b_x of the signal with power spectrum $P_f(u)$

$$b_x^2 = \int u^2 \, P_f(u) \, du \quad / \quad \int P_f(u) \, du \tag{3}$$

we obtain the relations:

$$\sigma_x = \frac{1}{\sqrt{n}} \cdot \frac{1}{SNR} \cdot \frac{\sigma_f}{\sigma_{f_x}} = \frac{1}{\sqrt{n}} \cdot \sqrt{\frac{1-\varrho}{\varrho}} \cdot \frac{\sigma_f}{\sigma_{f_x}} \tag{1a,b}$$

$$\sigma_x = \frac{1}{\sqrt{n}} \cdot \frac{1}{SNR} \cdot \frac{1}{2\pi b_x} = \frac{\sigma_n}{\sum_1^n f_{x_i}^2} \tag{1c,d}$$

If two images are matched the theoretical precision for the parallax can be determined from

$$\sigma_p = \sqrt{2} \cdot \sigma_x \quad . \tag{4}$$

Examples:
A few examples want to demonstrate the usefulness of these relations, specifically how they compare to empirical findings.

a. Phase Measurements using narrow banded signals with average wavelength $l = 1 / b_x$:
For GPS measurements the SNR of the recovered carrier is appr. 16 dB. This leads to

$$N = 1,\ SNR = 40,\ \Rightarrow \sigma_x = 1 / (40 * 2 * \pi) \approx 0.004\ l$$

and with a wave length of $l = 20$ cm we obtain a theoretical precision of 0.7 mm. This has to be compared with empirical findings for the accuracies of short distances in the range of a few mm.

b. Area based image matching using cross correlation or least squares matching:
The power spectrum in aerial images can be approximated by $P(u) \cdot \exp(-a|u|)$. In good images $a = 0.2$ mm (cf. HELAVA 1976) which leads to a bandwidth of $b = \sqrt{2}/a = 7\ mm^{-1}$. Using windows of 12 x 12 pixels and assuming SNR = 5 this leads to a theoretical precision of the parallaxes of $\sigma_p = 0.54$ µmm which for a pixel size of $dx = 20$ µm is equivalent to 1/40 dx. The empirical findings range from 1/10 to 1/50 dx (cf. PERTL 1984, VOSSELMAN/FÖRSTNER 1988).

c. Extraction of feature points with an interest operator (cf. PADERES et al. 1984, FÖRSTNER/GÜLCH 1986):
The extraction of feature points is usually based on small windows containing e. g. only 2 x 4 values. Under the same conditions as in the previous example (SNR = 5, $b = 7\ mm^{-1}$) this leads to a positional accuracy of appr. $\sigma_x = 1.6$ µm. It has to be compared to the empirical precision of 3 µm. (from HAHN/ FÖRSTNER 1988, rms differences of two independent height measurements 0.7 ‰ of the flying height).

d. Matching of two Digital Elevation Models (DEM):
EBNER and MÜLLER (1986) report on using a DEM as vertical and horizontal control for aerial triangulation with three line scanner imagery. In their example the precision of the height is given by appr. 0.10 m. The standard deviation of the slopes of the hilly terrain can be approximated by 10 %. They use 100 control points which would lead to a theoretical precision of the horizontal position of the DEM of $\sigma_p = 0.14$ m (neglecting the geometric complex geometric relations). This has to be compared with their empirical finding of $\sigma_p = 0.20$ m.

Obviously the theoretical accuracies indicate the potential of the methods which could be reached if all other effects can be taken care of by proper calibration. ▪

1.2 The effect of image blur onto the geometric precision

We will now investigate the effect of image blur onto the precision of feature extraction using eq. (1c).

Let the unblurred and noiseless (onedimensional) image f(x) be given with power spectrum $P_f(u) = P_f(0) \cdot \exp(-a|u|)$. The bandwidth of this image is (cf. 1.1b) $b = \sqrt{2}/a$. Blurring of f(x) can be approximated by convolution with a box filter h(x) with a width being the integration path δx. In order to derive the precision of feature extraction analytically in dependency of δx, we approximate h(x) by a Gaussian with the same standard deviation $\sigma_h = \delta x/\sqrt{12}$. Its Fourier spectrum H(u) also is a Gaussian with standard deviation $\sigma_H = 1/(2\pi\sigma_h) = \sqrt{3}/(\pi \cdot \delta x)$ (cf. CASTLEMAN 1979), which again can be approximated by an ideal low pass filter with upper frequency $u_o = 3/(\pi \cdot \delta x)$. We therefore assume that blurring with path δx can be approximated by an ideal low pass filter h(x) with this upper frequency. The blurred image

$$g(x) = h(x) * f(x) \tag{5}$$

then has the effective bandwidth

$$b_g^2 = \int_0^{u_o} u^2 \, P_f(u) \, du \; / \int_0^{u_o} P_f(u) \, du \tag{6}$$

$$= \frac{1}{a^2} \; \frac{2 - (z^2 + 2z^2 + 2)\, e^{-z}}{1 - e^{-z}} \tag{7}$$

with $z = a\, u_o$, which for large values $\delta x \gg a$ leads to

$$b_g^2 = \frac{3}{\pi^2 \, \delta x^2} \; . \tag{8}$$

With eq. (1c) we finally obtain the precision of feature location under (strong enough) image blur δx

$$\boxed{\sigma_x(\delta x) = \frac{1}{\sqrt{12\,n}} \cdot \frac{1}{SNR} \cdot \delta x} \tag{9}$$

Discussion:

a. The precision is independent of the texture of the image measured by the parameter a. This is reasonable as only the low frequencies of the spectrum are relevant for the location after image blur.
b. The precision is proportional to the blurring path δx, which is to be expected.

c. The precision increases with the squareroot of the number of pixels. As in all versions of eq. 1 n is the number of pixels really contributing to the location, i. e. only those with gradient not zero.

d. For SNR = 1 and interpreting δx as pixelsize we obtain the theoretical precision of the location of a straight line of height n in a noiseless binary image (cf. FÖRSTNER 1986). It indicates that the standard deviation of the noise at an ideal binary edge can be approximated by the standard deviation of the signal. The derivation for this theretical result is based on the spirograph theory of DORST and DUIN (1984). This coincidence of relations derived on totally different theories is remarkable as such.

Example:
Let the pixel size be dx, the blur be δx = 10 dx, N = 10, and SNR = 3. Then the image feature (say an edge) in spite of image blur can be located with a precision of 1/3 dx. This may be e. g. achieved by local approximation of the signal by a polynomial of low degree (say 3) and determining the inflection point.▪

We are now ready to model the accuracy of feature location in blurred images. We express it as a standard deviation of the direction Φ from the sensor to the point in object space in dependency on

- the pixel size dx
- the focal length f
- the relative velocity v of the object with respect to the sensor
- the effective integration time t_I
- the distance s from the sensor to the object
- the angle α betwen the viewing direction and the velocity vector
- the assumed precision $k_1 \cdot dx$ of feature detection in the unblurred image. k_1 can be derived from eq. (1) and the pixel size: $k_1 = \sigma_x / dx$ and will lie in the range $0.02 < k_1 < 0.5$,
- the assumed precision $k_2 \cdot \delta x$ of feature detection in the blurred image. k_2 can be derived from eq. (9): $k_2 = 1/(\sqrt{12n} \cdot SNR)$ and will also lie in the range $0.02 < k_2 < 0.5$

We can reasonably approximate the precision of the direction Φ by

$$\sigma_\Phi^2 = \left[\frac{k_1 \, dx}{f} \right]^2 + \left[k_2 \cdot \frac{v \, t_I \, \sin\alpha}{s} \right]^2 \qquad (10)$$

For v = 0 we obtain the theoretical precision in the unblurred image, while for large blur the second term is dominating.

A worst case analysis can be based on the following assumptions: dx = 20 µm, f = 20 mm, k_1 = 1/2, k_2 = 1/4 (in case image blur is taken into account), v = 20 m/sec and t_I = 1/100 sec. Neglecting the sin-term this leads to the model

$$\sigma_\Phi^2 \approx (0.031 \text{ g})^2 + (0.05 \text{ m / s [m]})^2 \qquad (11)$$

which has a structure commonly used for the variance of direction measurements in geodetic networks and which is used in the example in section 2.3.

Exploiting the potential of feature detection and using a longer focal length may result in standard deviations which are a factor 3 to 20 better.

2. Precision of Geometric Properties derived from Image Sequences

This section wants to discuss the precision of various geometric properties of fix and moving objects as they can be derived by tracking features in image sequences. Specifically we are interested in relative distances and velocities between the sensor and possibly moving objects, and in the time the sensor is expected to meet the object, e. g. when being overtaken by another vehicle. We will treat the case of mono and stereo sequences. Finally we are interested in the precision with which one is able to track the own position and orientation from continuous observation of targets. To simplify the derivations we restrict the discussion to a two dimensional environment. Due to the nature of the mensuration process, namely measuring the geoemtry of bundles of rays, i. e. angles, the results can easily be transferred to three dimensions.

The section is organized along three case studies

- estimating the distance to an object from a mono image sequence

- estimating the distance, the velocity and the meeting time from a stereo image sequence,
- estimating the own trajectory from image sequences.

The examples are based on two different sensor systems:

System A (low accuracy; focal length 20 mm, pixel size 20 µm, $k_1 = 1/2$) obtains directions with standard deviation of **1/2000 = 0.03 gon.**

System B (high accuracy; focal length 50 mm, pixel size 10 µm, $k_1 = 1/4$) obtains directions with standard deviation of **1/20000 = 0.003 gon,** thus with a 10 times higher precision than system A.

2.1 Estimating the distance to an object from a mono image sequence

Let us assume the target point P(x,y) is observed from a moving platform over a short path of length l_m measuring the direction α_i n times (cf. Fig. 1). The trajectory of the platform is assumed to be known. The object has an average distance of approximately s from the sensor and is seen under

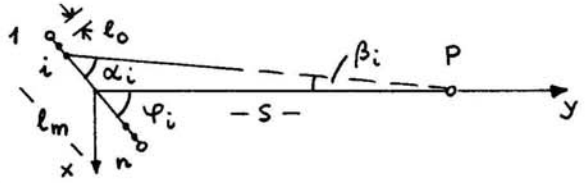

Fig. 1 Distance estimation from short image sequence

the aspect angle Φ. The parallactic angle $\beta_i = (\Phi - \alpha_i)$ can be written in dependency from the sequential number i of the directions as

$$\beta_i = (i - (n+1)/2) \cdot l_o \cdot \sin\Phi / s \qquad (12)$$

and can be used in the n observation equations for the overconstrained problem

$$(\Phi - \alpha_i) + e_i = dx / s + \beta_i / s \cdot dy \qquad w_i = 1 \qquad (13)$$

using the approximations $s_i \approx s$, $\sin\beta_i \approx \beta_i$, $\cos\beta_i \approx 1$. The directions are assumed to have equal weight $w_i = 1$. From the normal

equation matrix

$$N = 1/s^2 \begin{bmatrix} n & 0 \\ 0 & \sum_{i=1}^{n} \beta_i^2 \end{bmatrix} \tag{14}$$

one obtains the standard deviations for the coordinates

$$\sigma_x = \frac{s}{\sqrt{n}} \sigma_\alpha \tag{15}$$

and

$$\sigma_y = \frac{s^2}{l_o \sin\Phi} \cdot \sqrt{\frac{12}{n \cdot (n^2-1)}} \cdot \sigma_\alpha = \frac{s^2}{l_m \sin\Phi} \cdot \sqrt{\frac{12 \cdot (n-1)}{n \cdot (n+1)}} \cdot \sigma_\alpha \tag{16}$$

which for large n can be approximated by

$$\sigma_y = \frac{s^2 \sqrt{12}}{l_m \sin\Phi} \cdot \frac{\sigma_\alpha}{\sqrt{n}} = \frac{s^2 \sqrt{12}}{l_o \sin\Phi} \cdot \frac{\sigma_\alpha}{n^{3/2}} \tag{17}$$

Discussion:

1. The x-coordinate just is the mean over all individual estimates, and therefore its precision increases with $\sqrt{n}$ and is proportional to the distance.
2. The precision of the distance decreases with the square of the distance, independent of the other parameters, which is known for the case of only two rays.
3. The precision of the distance depends on the angle Φ between the viewing direction and the direction of the movement. As to be expected, the distance cannot be estimated if the object is in the direction of the trajectory.
4. The precision of the distance can be related either to the path length l_m or to the distance l_o between two neigbouring sensor positions. In case the precision is related to a constant path length, the precision increases with $\sqrt{n}$. On the other hand, if the velocity v and the time intervall t_o between the observations is fixed, thus $l_o = v \cdot t_o$ is fixed, then the precision increases with $n^{3/2}$, thus an increased number of frames can be very effective.

Example:
Let a non moving object have a distance of approximately 100 m and be seen under an angle of $\Phi = 1/25$, thus lying 4 m aside the momentary driving direction. The directions of camera system A have a standard deviation of 1/2000 = 0.03 gon. Let the own velo-

city be v = 20 m/s = 72 km/h and t_o = 1/25 s, thus l_o = 0.8 m. Then the distance can be measured with a standard deviation of $86/\sqrt{n \cdot (n^2-1)}$ [m] with system A. If only two consecutive frames are used (n=2) the precision is appr. 35 m or 35 %. If the object is observed for 1/2 s, thus when using n=13 frames, the precision is 0.6 m or 0.6 %. With camera B already with 2 frames a relative precision for the distance of 3.5 % can be achieved. ▪

Though the own trajectory has been assumed to be known, this information is not difficult to obtain from image sequences over short time ranges (see below). However, a good relative accuracy for the distance obviously is obtainable only for point type objects lying far enough aside the momentary path, if a certain integration over time takes place or if the camera system yields a good enough precision. Aside from these restrictions no information on the relative speed to a moving object can be derived, as parallaxes due to distance and due to speed can not be separated.

2.2 Estimating geometric properties from stereo image sequence

The disadvantages of mono image sequences do not occur in stereo sequences. We want to develop the relations in several steps starting with the estimation of the distance from one image pair. With the base line b, lying perpendicular to the trajectory, and the parallactic angle δ the distance is given by

$$s = \frac{b}{2} \cot \frac{\delta}{2} \tag{18}$$

or for $b \ll s$ approximately

$$s = \frac{b}{\delta}. \tag{19}$$

The variance of the distance then is given by

$$\sigma_s^2 = \frac{1}{\delta^2} \left(s^2 \, \sigma_\delta^2 + \sigma_b^2 \right) = \frac{s^4}{b^2} \, \sigma_\delta^2 + \frac{s^2}{b^2} \, \sigma_b^2 \tag{20}$$

which clearly shows that the uncertainty of the baseline is negligible for larger distances. The first part in eq. (19) can be obtained from eq. (15) with n = 2 and Φ = 100 gon, l_o = b and $\sigma_\delta = \sqrt{2} \, \sigma_\alpha$.

Example:
Let the following values be given σ_b = 0.01 m, $\sigma_\delta = \sqrt{2}/2000$ = 0.02 gon (system A), b = 1 m. A distance of 10 m can be estimated with 0.12 m or 1 %, and a distance of 100 m will have a standard deviation of 7 m, leading to a relative accuracy of 7 %, which

may be fully sufficient for certain tasks. These accuracies can be obtained instantaneously. With camera system B the relative accuracy of a distance of 300 m is still 21 %. ▪

A similar relation is obtained for the velocity derived from two image pairs. The difference s_m of two distances s_1 and s_2 with $\delta = \sqrt{\delta_1 \cdot \delta_2}$ and $\delta_m = \delta_1 - \delta_2$ approximated by

$$s_m = s_2 - s_1 = b \left(\frac{1}{\delta_2} - \frac{1}{\delta_1}\right) = \frac{b\,\delta_m}{\delta^2} = \frac{s\,\delta_m}{b} \qquad (21)$$

and leads to the variance for the measured distance difference s_m

$$\sigma_{s_m}^2 = \left[\frac{\delta_m}{\delta^2}\right]^2 \sigma_b^2 + \left[\frac{b}{\delta^2}\right]^2 \sigma_{\delta_m}^2 \qquad (22)$$

the standard deviation of the velocity then is given by

$$\sigma_v = \sigma_{s_m} / t_m \qquad (23)$$

Here t_m is the time difference between the two measurements.

Example:
Under the same conditions, as in the previous example the standard deviation of the distance difference for system A is σ_{sm} = 10 m, for system B σ_{sm} = 1 m. The influence of the uncertainty of the basis is negligible. Two consecutive frames cannot be used for estimating the velocity, as their standard deviation is σ_v = 250 m/s and 25 m/s resp. . Even if one uses two frames which are 1 s apart, the accuracy is still only 10 m/s = 36 km/h for an object which is 100 m far (system A). The situation, however, is much improved when using sytem B. ▪

These relations directly can be used for estimating the time t until object and sensor meet. It is given by

$$t = s/v = \delta \cdot t_m / \delta_m \qquad (24)$$

and has the relative accuracy

$$\frac{\sigma_t}{t} = \frac{\sigma_{\delta_m}}{\delta_m} \qquad (25)$$

The time is independent of the baselength used in the stereo sequence. Obviously the bias in the baselength influences distance and velocity in the same manner, specifically the relative syste-

matic error in the baselength is identical to the relative systematic erros in distance, distance difference and velocity.

Example:
In case the relative velocity would be 20 m/s and two frames with a delay of 1 s would be used for estimating the meeting time t, which is approximately 5 s, the relative accuracy would only be 50 % for system A, however 5 % for system B. ▪

We now want to investigate the effect of a larger number of stereo measurements onto the precision of distance, velocity and time. From the individual distances $l_i = b/\delta_i$ we can derive the average distance s_o and the velocity v from the overconstrained system

$$l_i + e_i = s_o + t_i \cdot v \,, \qquad \sigma_{l\,i} = \sigma_l = b \,/\, \delta^2 \cdot \sigma_\delta \tag{26}$$

In a similar way as in section 2.1 we can build up the normal equation matrix and obtain the theoretical values for the standard deviation of the average distance and the speed:

$$\sigma_s = \frac{\sigma_l}{\sqrt{n}} = \frac{s^2}{b} \cdot \frac{\sigma_\delta}{\sqrt{n}} \tag{27}$$

and

$$\sigma_v = \frac{s^2}{t_o\ b} \cdot \sqrt{\frac{12}{n \cdot (n^2-1)}} \cdot \sigma_\delta = \frac{s^2}{t_m\ b} \cdot \sqrt{\frac{12 \cdot (n-1)}{n \cdot (n+1)}} \cdot \sigma_\delta \tag{28}$$

which for large n can be approximated by

$$\sigma_v = \frac{s^2\ \sqrt{12}}{t_m\ b} \cdot \frac{\sigma_\delta}{\sqrt{n}} = \frac{s^2\ \sqrt{12}}{t_o\ b} \cdot \frac{\sigma_\delta}{n^{3/2}} \tag{29}$$

The precision for the time t = s/v until object and sensor meet is given by

$$\sigma_t^2 = \left[\frac{s^2}{b\ v}\, \sigma_\delta \right]^2 \cdot \left[\frac{1}{n} + \left\{\frac{t}{t_m}\right\}^2 \frac{12\ n}{n^2-1} \right]^2 \tag{30}$$

For observation times $t_m < t$ and large n the 1/n term in the brackets, which represents the influence of the distance estimate onto the time, is negligible. This leads to the relations for the relative accuracy of the time estimate

$$\frac{\sigma_t}{t} = \frac{t}{t_m} \cdot \sqrt{\frac{12}{n}} \cdot \frac{\sigma_\delta}{\delta} = \frac{s^2}{v\ b} \cdot \sqrt{\frac{12}{n}} \cdot \sigma_\delta \tag{31}$$

It is interesting to note that the relative accuracy of the time t does not depend on the velocity nor on the basis, but only on the relative observation time t_m/t, the number of frames and the relative accuracy of the parallactic angle. Referring to a constant precision of the directions, only the parallactic angle $\delta = b / s$ is really influencing the relative precision.

Example:
If under the same conditions as in the last example 25 frames are taken for the estimation of the time t, it can be estimated with a relative accuracy of 25 % with system A and 2.5 % with system B, which is an increase of only a factor 2 compared to the result with only 2 frames. With camera system B a relative precision of 10 % would be reached for objects with a distance up to 200 m. ▪

The advantages of stereo sequences are obvious: They allow an instantaneous estimate of distances, achieve useful estimates for the absolute speed, when applied to fix objects or for the relative speed to other vehicles. In contrast to mono sequences they need no information on their own trajectory, epecially no orientation. On the other hand the results of this subsection can be used also in mono image sequences in case pairs of points with a constant distance between them are tracked. In this case the point pair can be treated as the base line and the parallactic angle δ is directly measured in one camera. As the relative error of the baselength usually is negligible or in the case of time estimates even of no influence, the discussed accuracies can also be achieved in mono sequences, again not requiring an absolute frame. Such point pairs could be the front or the rear lights of a car, which within a class of vehicles show only a limited variation in distance.

2.3 The precision of the trajectory from image sequences

In case the trajectory of the sensor is long compared to the distance to the object points, which connect the individual sensor positions, the accuracy of the trajectory follows the same rules as a strip of images in aerial triangulation.

The accuracy properties of strips have intensively investigated by ACKERMANN (1965). Of special interest was the dependency of the precision of the connecting tie points on the length of

the strip. The result can be reduced to the following relations:

$$\sigma_X = k_X \cdot n^{3/2}, \quad \sigma_Y = k_Y \cdot n^{3/2}, \quad \sigma_Z = k_Z \cdot n^{3/2} \qquad (32)$$

where n is the number of images in the strip and the ki are apropriate constants. These relations were found to hold for all types of calculation schemes, suggesting that the structure of the strips dominates all other effects, and differences in algorithm and control point density only influence the constants.

There are some differences of photogrammetric strips and image sequences taken from a vehicle:

- The viewing direction in image sequences usually is not 90° across the direction of the movement. This leads to much higher overlap in image sequences. In extreme cases single points may be trackable through hundrets of images. This locally may lead to extremely stable connections between the images especially with respect to the orientation parameters.
- The distances to the tie points in image sequences varies to a large extent. Smallest distances may be a few meters, long distances a few kilometers. This leads to a high inhomogeneity of the connections and makes realistic simulations difficult to achieve.
- in navigation one primarily is interested in the precision of the parameters for the sensor platform, as the positions of the object points can be determined with high enough accuracy.

The differences will be overruled by the strip-nature of the image sequence in the long run, thus, if only one sensor is used, leading to the same $n^{3/2}$ behaviour of the standard deviations of the positions of the sensor platform. On short distances the behavior may however be of totally different nature.

The orientation parameters were not investigated in this study. The principle of double summation which lead to the $n^{3/2}$ rule for the position can be transferred to obtain the the dependency of precision of the orientation parameters on the number of images. As a single summation takes place for the orienation angles, the standard deviations of them only increases with $\sqrt{n}$ (cf. FÖRSTNER 1983, p. 8).

Thus we can summarize the general structure of the precision of a long homogeneous trajectory from mono image sequences:

$$\sigma_i = k_i \cdot n^{3/2}, \quad i = x, y, z; \qquad \sigma_i = k_i \cdot n^{1/2}, \quad i = \Omega, \Phi, K \tag{33}$$

Investigations have to find out the constants ki in dependency on the main parameters describing an image sequence.

Example:
In order to get an impression of the absolute accuracy of the trajectory derived from image sequences we simulated a short sequence. It consists of a point field with 100 points randomly distributed in a rectangle of 100 x 200 m². The trajectory consists of 50 points with a distance of 1 m each thus had a length of 50 m. It is situated in the centre along the major axis of the point field. The first two points are assumed to be known. All 100 points are assumed to be visible from each of the 50 points of the trajectory. This gives rize to 50 sets of direction. The orientation of the first two sets was also assumed to be known. The angular error is assumed to be 1/2000 = 0.03 gon with an additional a term 0.05/s[m] simulating the effect of image blur (cf. sect. 1.2). With an average of a distance of 40 m this leads to a standard deviation of the direction of 0.0013 = 0.085 gon. The 50 sets of bundles have been used for a joint estimation of all unknown parameters, including the coordinates of the tie points and the trajectory.

The result which reflects the last step of a rigorous Kalman-Filter is the following:

$$\sigma_x = 0.68 \text{ m}, \quad \sigma_y = 0.031 \text{ m}, \quad \sigma_\Omega \approx 0.008 \text{ gon}$$

The result is remarkable in two ways:

- The longitudinal error (x-direction) is only 1/2 m, thus 1 % of the travelled distance.
- The traversal error (y-direction) is even better. The trajectory is only off the starting direction by 0.08 % = 0.038 gon.
- The standard deviation of the orientaion angle has been estimated from the precision of the tie points and with $\sigma_\alpha/\sqrt{nt}$ very well can be predicted from the precision of the individual directions.

The result has been achieved with the camera system A including the effect of image blur. An increase of the accuracy can be expected if a more sophisticated feature detector is assumed. ▪

For stereo image sequences (cf. MATTHIES 1988) the geometry is much more stable in the direction of the trajectory. As each image pair has the same scale, which is fixed by the baseline, the longitudinal error results from a single summation, while the other error laws stay unchanged in structure. Thus the general structure for the precision of long homogeneous trajectories from

image sequences is given by

$$\sigma_x = k_x \cdot n^{1/2}, \quad \sigma_i = k_i \cdot n^{3/2}, \quad i = y, z$$
$$\sigma_i = k_i \cdot n^{1/2}, \quad i = \Omega, \Phi, K \qquad (34)$$

Again the faktors k_i have to be determined in dependency of representative parameters of the image sequence.

Summarizing the study clearly demonstrated which precision can be obtained from image sequences. It is up to the designer of a system to integrate the geometric information coming from a visual sensor with other sensors which may support it. In any case the visual sensor will have its main application in the interpretation of the environment.

I want to thank Matthias Neureither for his support in the simulations.

References:

IAP International Archives of Photogrammetry and Remote Sensing

IEEE T-PAMI IEEE Transactions on Pattern Analysis and Machine Intelligence

SIPUS Schriftenreihe des Instituts für Photogrammetrie der Universität Stuttgart

ACKERMANN F. (1965): Fehlertheoretische Untersuchungen über die Genauigkeit photogrammetrischer Streifentriangulationen, DGK C 87, München 1965

EBNER H., MÜLLER F. (1986): Processing of Digital Three Line Imagery Using a Generalized Model for Combined Point Determination, IAP 26-3/1, Rovaniemi, 1986, pp. 212-222

CASTLEMAN K. R. (1979): Digital Image Processing, Prentice Hall, 1979

DORST L., DUIN R. P. W. (1984): Spirograph Theory: A Framework for Calculations on Digitized Straight Lines, IEEE T-PAMI, Vol. 6, No. 5, 1984, pp. 632-639

FÖRSTNER W. (1982): On the Geometric Precision of Digital Correlation, IAP Vol. 24-3, Helsinki, 1982, pp. 176-189

FÖRSTNER W. (1983): On the Morphological Quality of Digital Elevation Models, Int. Colloquium on "Mathematical Aspects of Digital Elevation Models", Stockholm 1983

FÖRSTNER W. (1984): Quality Assessment of Object Location and Point Transfer Using Digital Image Correlation techniques, IAP 25-A3a, Rio de Janeiro, 1984, pp. 197-219

FÖRSTNER W. (1986): Prinzip und Leistungfähigkeit der Korrelation und Zuordnung Digitaler Bilder, SIPUS 11, Stuttgart 1986, S. 69-90

FÖRSTNER W., GÜLCH E. (1987): A Fast Operator for Detection and Location of Points, Corners and Circular Features, Intercommission Workshop on "Fast Processing of Photogrammetric Data", Interlaken 1987

HAHN M., FÖRSTNER W. (1988): On the Applicability of a Feature Based and a Least Squares Matching Algoritm for DEM Aquisition, IAP 27-3, Kyoto 1988

HELAVA U. V. (1976): Digital Correlation in Photogrammetric Instruments, IAP 23-2, Helsinki 1976

MATTHIES L. (1988): Depth from Stereo Image Sequences, Int. Workshop "High Precision Navigation", Stuttgart-Altensteig-Warth 1988

PADERES F. F., MIKHAIL E. M., FÖRSTNER W. (1984): Rectification of Single and Multipleof Satellite Scanner Imagery using Points and Edges as Control, NASA-Workshop on Mathematical Pattern Recognition and Image Analysis, Houston, 1984

PERTL A. (1984): Digital Image Correaltion with the Analytical Plotter Planicomp C 100, IAP 25-3, Rio de Janeiro 1984

VOSSELMAN G., FÖRSTNER W. (1988): The Precision of a Digital Camera, IAP 27-1, Kyoto, 1988