

RANGE AND VELOCITY ESTIMATION OF OBJECTS AT LONG RANGES USING MULTI-OCULAR VIDEO SEQUENCES

N. Scherer, R. Gabler

FGAN-FOM Research Institute for Optronics and Pattern Recognition, Gutleuthausstraße 1, 76275 Ettlingen, Germany
(scherer, gabler)@fom.fgan.de

Intercommission WG V/III

KEY WORDS: Stereoscopic, Infrared, Sequences, Tracking, Resolution, Accuracy

ABSTRACT:

Automatic warning systems for protection against airborne threats require a high detection probability at long ranges. Infrared Search and Track (IRST) systems which are designed as monocular and passive systems are able to fulfil this requirement. However, due to clutter and distortions by other objects, like birds or clouds, these systems do not achieve the required low false alarm rate. For a reduction of the false alarm rate the exploitation of features like shape, size, texture and intensity is not reliable, because objects at long ranges appear as points. Range and velocity are more robust features of the point like objects. These features can be obtained by stereo vision from image sequences of multi-ocular systems.

In this paper we describe investigations about the accuracy and reliability concerning the three-dimensional position and velocity of the objects. Unavoidable uncertainties in the measurement of the two-dimensional object position in the sensor focal plane lead to large errors in the estimated distance. We present a quantitative analysis of this issue, which results in fundamental restrictions for velocity estimation of objects. These considerations of accuracy and reliability are important for the design of multi-ocular IRST systems.

The theoretical analysis is compared with the result of a processed IR stereo image sequence recorded at a measurement campaign with real objects. It is shown that data processing considering the fundamental restrictions lead to robust results for estimation of the spatial position and velocity. This information can be effectively used to reduce the false alarm rate.

KURZFASSUNG:

Automatische Warnsensoren zum Schutz gegen Bedrohungen aus der Luft müssen auch bei großen Zielentfernungen hohe Detektionswahrscheinlichkeiten sicherstellen. Mit einem monokularen passiven IRST(Infrared Search and Track)-System kann diese Forderung erfüllt werden. Bei Störungen durch Clutter und Objekte wie z.B. Vögel oder Wolken wird jedoch nicht die geforderte geringe Falschalarmrate erreicht. Merkmale wie Größe, Form, Textur oder Intensität der abgebildeten Objekte sind aufgrund der großen Entfernungen und damit der Punktförmigkeit der Objekte nicht zuverlässig genug, um zur Reduktion der Falschalarmrate genutzt zu werden. Entfernung und Geschwindigkeit der Objekte, die durch Auswertung multiokularer Stereo-Bildfolgen bestimmt werden können, lassen sich dagegen erfolgreich als Merkmale zur Trennung zwischen Ziel- und Störobjekten einsetzen.

In diesem Beitrag werden Untersuchungen zur Genauigkeit und Verlässlichkeit der durch Stereoauswertung bestimmten dreidimensionalen Objektposition und -geschwindigkeit vorgestellt. Unvermeidbare Fehler in der Bestimmung der Position der abgebildeten Objekte in den Bilddaten führen zu großen Entfernungsfehlern für die berechnete dreidimensionale Objektposition. Wir zeigen, dass eine quantitative Untersuchung dieses Problems zu prinzipiellen Einschränkungen in der Genauigkeit der berechneten Objektgeschwindigkeit führt. Diese Überlegungen sind grundlegend für den Entwurf eines multiokularen IRST-Systems.

Die theoretische Untersuchung wird mit der Auswertung von Stereobildfolgen, die bei einer Messkampagne mit echten Flugobjekten gewonnen wurden, verglichen. Wir können zeigen, dass man unter Berücksichtigung der prinzipiellen Einschränkungen zuverlässige Objektpositionen und -geschwindigkeiten erhält. Diese können effektiv zur Reduzierung der Falschalarmrate genutzt werden.

1. INTRODUCTION

Modern automatic warning systems for protection against airborne threats are of increasing interest worldwide. The use of a monocular Infrared Search and Track (IRST) system for this purpose can fulfil the requirements for high detection probability at long ranges as well as for all day vision capability (Campana 1993). However, the desired low false alarm rate (FAR) could not be achieved by these systems while keeping the high detection probability of objects at long ranges. The reason for this can be traced back to the fact that target objects at long ranges as well as clutter and distortions by other objects like birds or parts of clouds appear as points in the images. As a consequence features like shape, size, texture and even intensity

are not reliable for the separation of real alarms from false alarms. More robust features for discrimination are the three-dimensional object position and the three-dimensional object velocity. To obtain this information about the objects the system may be augmented by active sensors like radar or laser, but to maintain the passive operation of the system the use of stereo vision from image sequences of multi-ocular sensor systems is investigated.

Our approach to reconstruct three-dimensional information about the observed objects from multi-ocular image sequences is composed of different steps. First, a precise sensor calibration (inner and outer orientation) is carried out to get the high

accuracy needed for measurements of objects at long ranges. In each image of a sequence, objects are detected and compatible objects are connected to tracks within the sequence. Additionally, the correspondence problem between objects in sequences of different sensors will be solved for each time step. Then, the three-dimensional position of a potential object will be determined by resection in space. Last, the object positions of each time step are transformed into a trajectory within a space-time cube and from this, three-dimensional vectors of velocity will be calculated.

In this paper we describe investigations about the accuracy and reliability of this approach, mainly concerning the last two steps (the three-dimensional position and the velocity of the objects). Unavoidable uncertainties in the measurement of the two-dimensional object position in the sensor focal plane lead to rather large errors in the estimated distance, which in turn affect the accuracy of velocity extraction from the sequence.

We present a quantitative analysis of this issue, resulting in statements about fundamental restrictions for the velocity estimation of objects. These considerations of accuracy and reliability are important for the design of multi-ocular IRST systems.

A measurement campaign was carried out to capture image sequence data with real objects using IR sensors. It will be shown that by considering the fundamental restrictions an adaptive processing leads to more robust results for the estimation of the spatial position and velocity. This information can be effectively used to reduce the FAR.

2. POSITION ACCURACY

To get reliable information about the three-dimensional object position and velocity it is necessary to ascertain the accuracy of these values. Since the accuracy of the three-dimensional object velocity depends on the accuracy of the three-dimensional object positions calculated for each image-pair in the sequence, we first discuss the fundamental limitations in the accuracy of the object position and use this result to obtain the accuracy of the calculated object velocity.

In order to find fundamental limitations in position and velocity accuracy, we assume everything being as perfect as possible. For example, the sensors are exactly identical with the same focal length and pitch (distance between centres of two adjacent detector elements) and can be approximated by pinhole cameras. In addition, the two sensors are aligned exactly parallel. Figure 1 shows a sketch of this situation and introduces the coordinate system with the z-axis along the viewing direction of the sensors and the x-axis along the baseline. The baseline is given by the distance of the sensors perpendicular to their alignment.

Also shown in Figure 1 are the images of an object for both sensors. From one image the range of the object is not known, but the object position in the image of one sensor (e.g. left) together with the pinholes of the two sensors define a plane, which intersects the focal-plane of the other sensor (right) in the so called epipolar line. In the ideal case the image of the object in this (right) sensor lies exactly on this epipolar line. The difference of the object positions on this epipolar line, which is in our case the difference in the horizontal positions x_{img1} and x_{img2} , is called disparity and measured in units of the pitch.

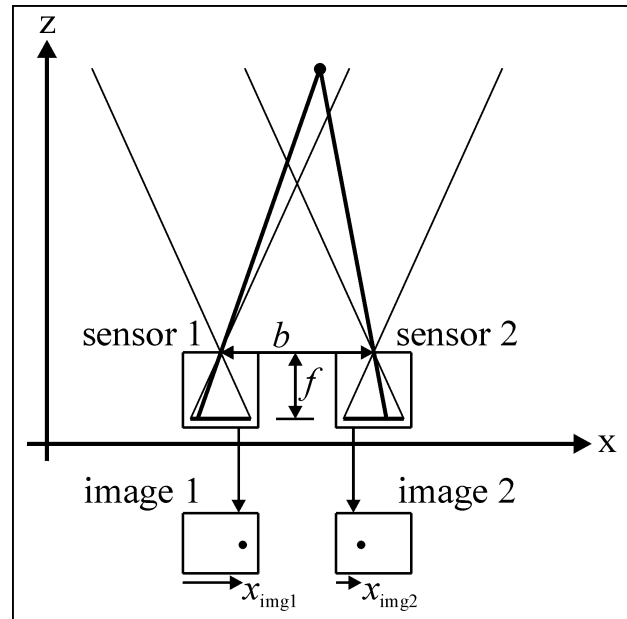


Figure 1. Arrangement of the stereoscopic system with focal-length f and baseline b and the two sensor images of an object with horizontal positions x_{img1} and x_{img2} .

The three-dimensional position of the object is given either by the direction (two angles) and range or the three Cartesian coordinates (x , y and z). Both representations are equivalent and can be transferred easily into each other. The direction is given by the two-dimensional object position in the image of one sensor. The range can be calculated with the z -component of the three-dimensional object position. The value of the z -component is given with the baseline (b), focal-length (f), pitch (a) and disparity (d) by the simple equation (1):

$$z = \frac{b \cdot f}{a \cdot d} = \frac{z_{max}}{d} \quad (1)$$

The quantity z_{max} is defined by the three parameters baseline, focal-length and pitch of the stereoscopic system. It has the dimension of length and introduces a natural measure of length for a given system. In addition, z_{max} is the greatest distance distinguishable from infinity for the given system.

Since we want to look at objects at long ranges the size of the physical image of the object in the focal plane is usually smaller than the pixel size. This means that the minimal uncertainty of the direction of the object is given by the instantaneous field of view (IFOV) of the concerning pixel. The intersection of the two IFOVs from the two sensors leads to a volume in space which defines the uncertainty of the object position. In Figure 2 a horizontal section through this volume is shown.

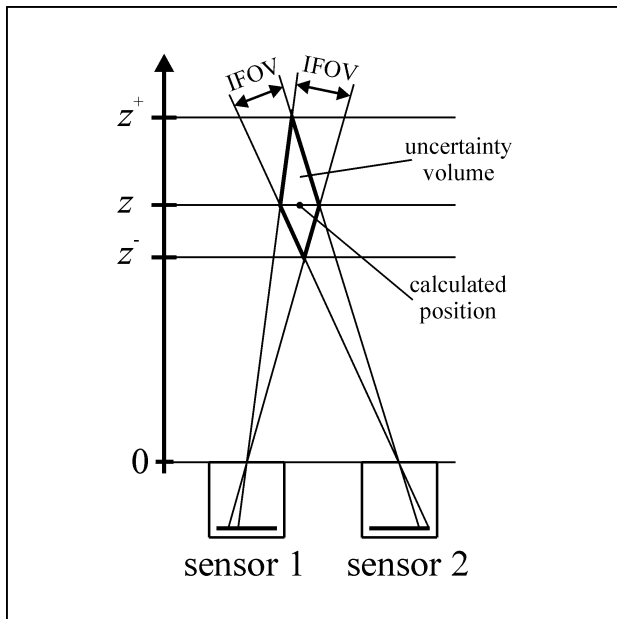


Figure 2. Position uncertainty volume with upper (z^+) and lower limit (z^-) for the calculated position with coordinate z .

For each object position in the two images the position uncertainty is $\pm \frac{1}{2}$ pixel. This leads to an uncertainty of ± 1 for the disparity. For small deviations of the disparity the deviation in the z -component of the calculated three-dimensional position is given by equation (2):

$$\Delta z = \left| \frac{\partial z}{\partial d} \right| \Delta d = \frac{z^2}{z_{\max}} \Delta d \quad (2)$$

It shows that for a fixed uncertainty in disparity, i.e. $\Delta d = \pm 1$, the uncertainty of the z -component (Δz) increases with quadratic order in z . Equation (2) is a good approximation for the position uncertainty at short ranges, but at long ranges the uncertainty of ± 1 for the disparity is not small compared to the absolute value of the disparity. There we use the exact values for the z -component of the farthest point (z^+) and the z -component of the nearest point (z^-) of the uncertainty volume given by equation (1) but decreasing and increasing the measured disparity by one unit, respectively, as shown in equation (3).

$$\begin{aligned} z^+ &= \frac{z_{\max}}{d-1} = \frac{1}{1/z - 1/z_{\max}} \\ z^- &= \frac{z_{\max}}{d+1} = \frac{1}{1/z + 1/z_{\max}} \end{aligned} \quad (3)$$

It is important to note that this uncertainty is not a statistical error, but a systematic error. This implies that it can not be decreased by subsequent independent measurements. To explain this fact, we assume that we have found the same disparity in several subsequent stereo-image pairs. With that the

probability for an object being near the average position is not higher than being somewhere else in the uncertainty volume, which means that the uncertainty volume does not decrease compared to the uncertainty volume for a single stereo-image pair. We have to remind this fact when we discuss the uncertainty for the velocity estimation in the next chapter.

3. VELOCITY ACCURACY

In this chapter we want to discuss the fundamental limitations for the accuracy of the object velocity based on the accuracy of the object position given in chapter 2. Since for objects at long ranges the position uncertainty in the z -component is always much greater than the position uncertainty perpendicular to it, we restrict our discussion to the z -component of the velocity and use for that component the variable v .

For the extraction of the object velocity we need subsequent position determinations for the object at different points of time. When an object is moving in space, it successively hits the same or other uncertainty volumes, which means that the calculated position located in the middle of each uncertainty volume jumps from time to time by rather large amounts. Figure 3 shows an example of this situation. All three tracks in Figure 3 lead to alternating values of the disparity of 3 and 4. Even though the tracks differ in the ratio of the number of these disparity values, for long ranges this ratio is affected by disturbance of the atmosphere and not quite reliable. Since the difference in disparity of ± 1 unit can be caused by approaching objects as well as by departing objects, this difference yields no information about the sign of v .

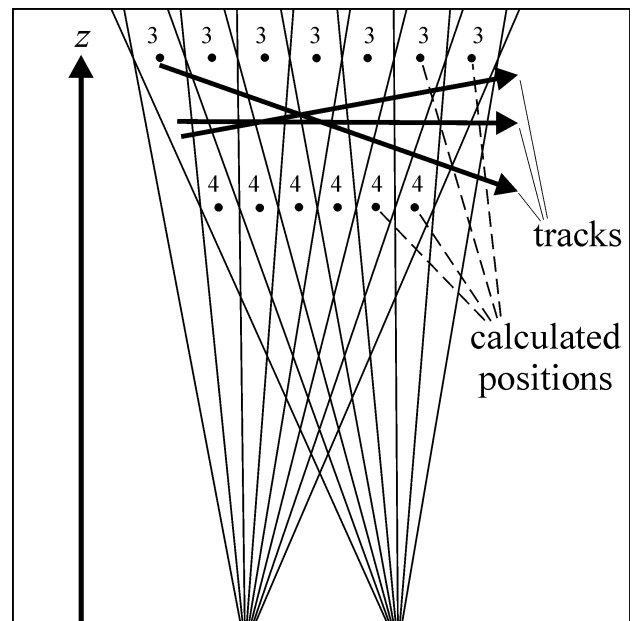


Figure 3. Different object tracks together with their corresponding uncertainty volumes. The calculated quantized positions and disparity values are indicated.

Only after the disparity reaches a difference of $+2$ or -2 units it makes sense to calculate a velocity value. Of course, the uncertainty of the velocity at this time is rather large, but it improves while the absolute difference of the disparity increases. Figure 4 shows how the position and the upper and

lower uncertainty boundary, z^+ and z^- according to equation (3), evolve with time for an object approaching the sensors with constant speed. The track starts at time t_0 with a value for the z -component of z_0 and reaches at the current time t the value z . For this diagram we waive the quantization of the position since for increasing difference in disparity this will be less important, but we have to keep the uncertainty range of ± 1 unit for the disparity. All possible trajectories with constant velocity which are consistent with the measured data are defined by the set of all straight lines which lie totally inside the boundaries z^+ and z^- . The extreme cases, upper (v^+) and lower (v^-) boundary for the calculated velocity, just touch the position boundaries as shown in Figure 4. There is no reduction of the velocity uncertainty depending on the number of measured positions as known from regression due to the fact, that the position uncertainty is not a statistical error as discussed at the end of chapter 2.

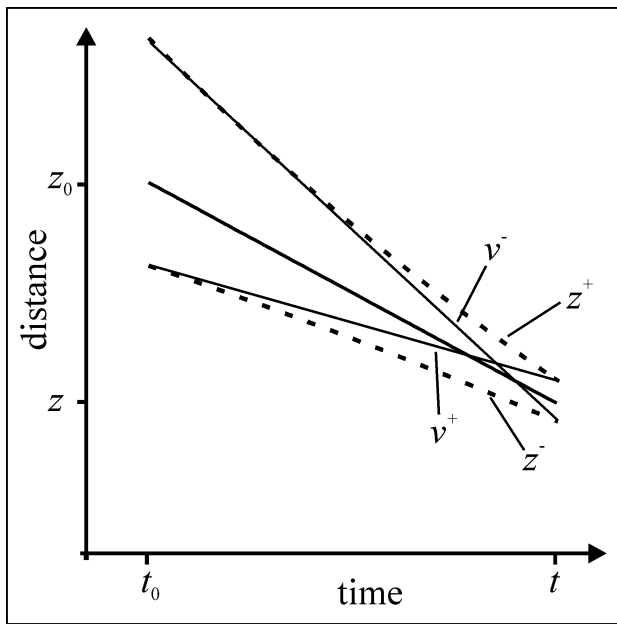


Figure 4. Distance over time, position uncertainties (z^+ and z^-), and limits for the fitted average velocity (v^+ and v^-) for an object approaching with constant speed.

Obviously the difference between v^+ and v^- decreases while the object approaches the sensors and therefore the position uncertainty at the current time t decreases. As a measure for the uncertainty of the velocity we introduce the relative velocity error $\Delta v/v$ given by equation (4):

$$\begin{aligned} \frac{\Delta v}{v}(z_0, z) &= -\frac{v^+ - v^-}{v^+ + v^-} \\ &= \frac{z_0}{z_{\max}} \cdot \frac{1 - \left(\frac{z}{z_{\max}}\right)^2 + \left(\frac{z}{z_0}\right)^2 \left(1 - \left(\frac{z_0}{z_{\max}}\right)^2\right)}{1 - \left(\frac{z}{z_{\max}}\right)^2 - \left(\frac{z}{z_0}\right)^2 \left(1 - \left(\frac{z_0}{z_{\max}}\right)^2\right)} \end{aligned} \quad (4)$$

The minus sign in the first line of equation (4) is to get positive values for the relative velocity error since the velocities v^+ and v^- have negative values for approaching objects. The values z and z_0 in equation (4) have the same meaning as in Figure 4, except z_0 is not necessarily the start-value of the track, but only some earlier and therefore greater value than z . The difference between z_0 and z is the track-length taken into account for the velocity calculation, which can be the whole or only part of the track. Inspecting equation (4), we find that the best choice regarding a small value for the relative velocity error is to use for z the smallest possible distance, which is the current distance, but to vary z_0 up to a maximum value which is set by the finite track-length. We expect an improvement of the relative velocity error with increasing difference between z and z_0 . This is only true until z_0 reaches a certain distance z_{opt} which is given by equation (5):

$$z_{\text{opt}}(z) = \frac{1}{\sqrt{2 \cdot \left(1/z^2 - 1/z_{\max}^2\right)} - 1/z} \quad (5)$$

In fact, further increasing of z_0 worsens the relative velocity error. The reason for this is that the boundaries for the position uncertainty, z^+ and z^- , are of second order in z but the fitted average velocity is represented by a linear line. This means that the choice of $z_0 = z_{\text{opt}}$ (if the track is long enough to allow this) leads to a minimal uncertainty for the object velocity achievable for the current distance z .

In summary, we found two fundamental limitations for the velocity determination. First, for an object at a current distance z we need a minimal track-length which must be in conformity with a difference of 2 units of disparity and second, the relative velocity uncertainty is limited by a minimal value achieved by selecting the optimal track-length ($z_{\text{opt}}-z$) for velocity extraction. In Figure 5 these results are shown in a diagram. The lower solid line shows the minimal track-length, which is necessary for velocity extraction, depending on the current distance given by the normalized value z/z_{\max} . The track-length is given by the normalized value $(z_0-z)/z_{\max}$. The upper solid line shows the optimal track-length, which is necessary to reach the minimal velocity uncertainty. The dashed lines show the relative velocity uncertainty obtained by choosing a track-length between the two limits.

Of course, the track-length taken into account for velocity extraction is limited by the length of the whole track. For example, this limit is shown in Figure 5 by the dotted line for an object detected first at $z/z_{\max} = 0,27$. As this object approaches the track-length increases. At the distance $z/z_{\max} = 0,17$ the whole track-length is long enough to recognize that the object approaches. While the object further approaches we are free to select any track-lengths for velocity extraction between the minimal track-length and the maximal track-length which is either given by the whole track-length or the optimal track-length. For example we can choose the whole track-length, which increases, until it exceeds the optimal track-length, which then decreases. This yields the best velocity uncertainty achievable at every current distance of the object but at the expense of a rather long track-length taking into account for the averaged velocity, what means that the velocity is averaged also over a rather long time interval. Or we can choose a fixed velocity uncertainty and after reaching this

velocity uncertainty move on the corresponding dashed line in Figure 5. This decreases the used track-length and therefore also the time interval for averaging much faster but on the expense of a constant perhaps rather poor velocity uncertainty. There are a lot of different alternative possibilities of choosing the track-length depending on the application.

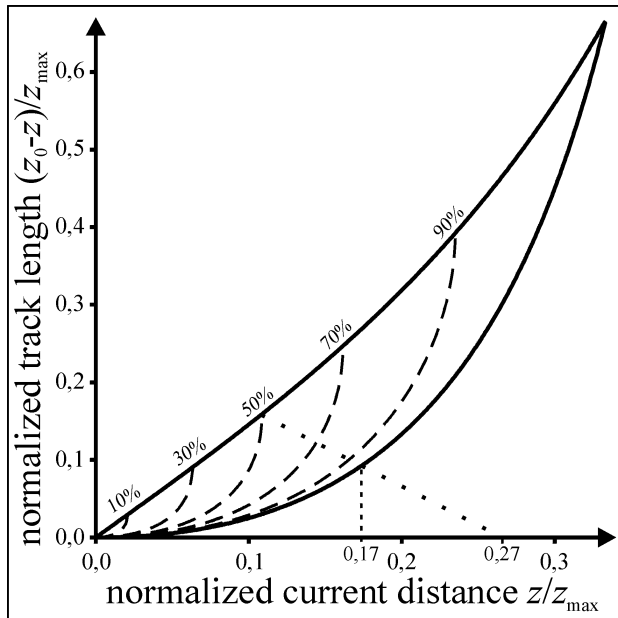


Figure 5. Minimal track-length (lower solid line), optimal track-length (upper solid line) and relative velocity uncertainties for different track-lengths (dashed lines) as function of the normalized current distance. The dotted line shows, as example, the whole track-length for an object detected first at $z/z_{max} = 0,27$.

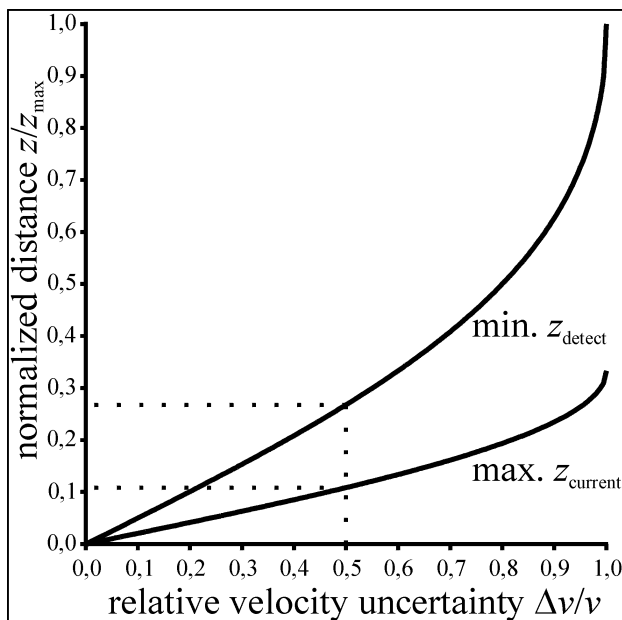


Figure 6. Maximal normalized current distance and minimal detection distance as function of desired relative velocity uncertainty.

It can be seen from Figure 5 that velocity extraction with a certain velocity uncertainty is only possible at distances less than an upper bound distance. This upper bound distance is shown in Figure 6 by the lower line. For the example of 50% uncertainty, this value is reached only below $z/z_{max} = 0,11$. In addition, to reach this value at this distance the track-length must exceed a certain value, what means that the first detection of the object must be farer than a lower bound distance shown by the upper line in Figure 6. In the given example the first detection must then be farer than $z/z_{max} = 0,27$.

In summary, for velocity extraction with an acceptable value for the velocity uncertainty at a certain range the stereoscopic system must be designed to resolve a much greater distance, which is given by the value of z_{max} .

4. DATA EVALUATION

For the data evaluation we used IR image sequences taken in November 2001 at the coast of Eckernförde in North Germany with various airborne objects approaching the sensors. We used two sensors from AIM in Germany with a Field of View (FOV) of $8.8^\circ \times 6.6^\circ$, 640×480 pixels and a focal length of 100 mm. The ground-truth positions of the objects were recorded using Differential Global Positioning System (DGPS).

Our approach to the evaluation of three-dimensional object position and velocity from bin-ocular image sequences can be subdivided into different, consecutive steps as follows: Each of the two image sequences is processed individually by an IRST algorithm which is composed of the tasks image pre-processing to correct sensor-specific inhomogeneities, motion-compensating temporal image integration to increase the signal-to-noise-ratio, non-linear spatial filtering to detect point-like as well as extended objects, segmentation of objects and spatio-temporal tracking of potential objects to create two-dimensional tracks in each image sequence.

For each stereo-image pair the positions of the objects which built the two-dimensional tracks are combined with the position and orientation of the sensors to reconstruct the three-dimensional positions of the objects by resection in space. These three-dimensional positions in consecutive image-pairs are linked together to three-dimensional tracks, using the two-dimensional track information. The three-dimensional velocity extraction starts when the difference of the maximum and minimum z -component of the track exceeds the minimal necessary track-length given by the lower solid line in Figure 5. After that the track-length for optimum velocity uncertainty is used for all inbound objects to extract the velocity until the relative velocity uncertainty falls below 50%. Then only the track-length to achieve 50% relative velocity error is used further. The used track-lengths are shown in Figure 7 as dots, together with the theoretical curves from Figure 5.

The dots in Figure 7 occur only at certain distances as expected from the quantization of the calculated three-dimensional positions. The positions are slightly smeared out due to the fact, that the object while moving occurs at different positions in the images and the sensors were not exactly aligned. The track-length chosen for velocity extraction is systematically greater than given by the 50% line. The reason for that is, that the calculated track-length to achieve 50% velocity error is transferred to an averaging time interval using not the averaged velocity but the lower bound of the velocity range calculated

one time step before. For high velocity uncertainties this leads to higher track-lengths and longer averaging time intervals and therefore to more robust results but reaches the correct value for decreasing velocity uncertainty. In our case, of course, the velocity uncertainty is kept fixed and therefore the chosen track-length is always a little bit greater than necessary for 50% uncertainty.

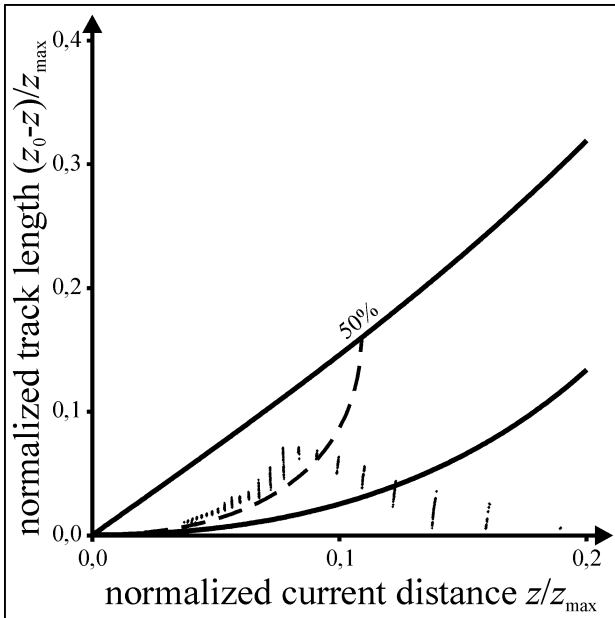


Figure 7. Detail of Figure 5 together with results of the algorithm (dotted points) evaluating a real stereo image sequence.

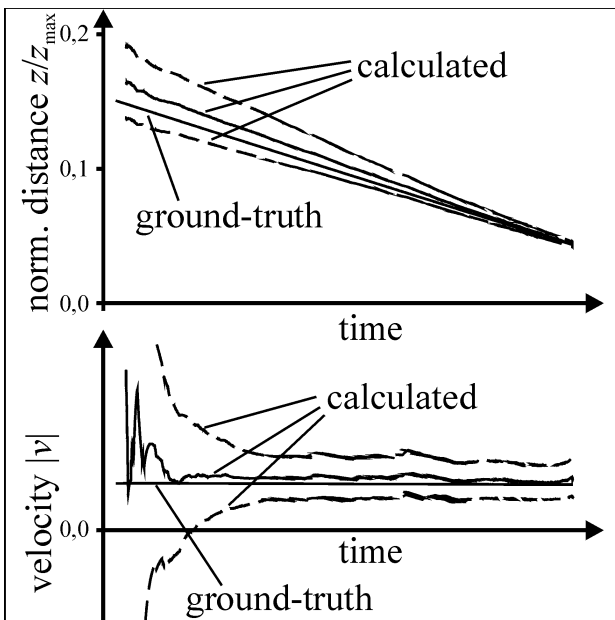


Figure 8. Comparison of evaluated positions (upper part) and velocities (lower part) with ground-truth data. The calculated data consist of averaged values and upper and lower boundary of uncertainty.

The calculated distances and velocities, which are averaged over the according time intervals, are shown in Figure 8 in the upper and lower part together with their uncertainty ranges as

function of time. The ground-truth distances and velocities are also shown.

As expected the distance uncertainty decreases constantly over time, while the velocity uncertainty decreases only until it reaches 50% uncertainty.

The very good agreement of the calculated positions and velocities with the ground-truth data within the calculated uncertainty ranges shows that the calculated values are fully reliable and can be used as robust features for any decision.

5. CONCLUSION

The quantitative analysis of the accuracy and reliability of three-dimensional position and velocity extraction for objects at long ranges from multi-ocular image sequences shows that fundamental restrictions limit the range for velocity extraction with a chosen uncertainty range much more than the simple geometric resolution of the multi-ocular system. Nevertheless, by adequate selection of the sensor arrangement and by taking into account the dependency between accuracy and the essential system parameters it is possible to derive reliable position and velocity values for objects even at long ranges when other features such as shape and intensity cannot be derived reliably.

ACKNOWLEDGEMENTS

The described work has been supported by BMVg (Bundesverteidigungsministerium, German Ministry of Defence) and BWB (Bundesamt für Wehrtechnik und Beschaffung, German Office for Armament and Procurement). The assistance of WTD 71 (Wehrtechnische Dienststelle für Schiffe und Marinewaffen) for preparation and realization of the measurement campaign at Surendorf in 2001 is gratefully acknowledged.

REFERENCES

Campana, S. B., 1993. Passive Electro-Optical Systems. In: *The Infrared & Electro-Optical Systems Handbook*, Ann Arbor, Michigan USA and Bellingham, Washington USA, Vol. 5, Chapter 4, pp. 209-344.