EVALUATION OF CAMERA CALIBRATION APPROACHES FOR VIDEO IMAGE DETECTION SYSTEMS

S. Bauer, A. Luber, R. Reulke

German Aerospace Center, Institute for Transportation Systems, Rutherfordstr. 2, 12489 Berlin, Germany
(sascha.bauer, andreas.luber, ralf.reulke)@dlr.de

Commission I, WG I/1

KEY WORDS: Calibration, Multisensor, Fusion, Orientation, Camera, Georeferencing

ABSTRACT:

In modern traffic management Video Image Detection Systems (VIDS) are becoming increasingly important as traffic sensors. They are getting more affordable and don't require any road construction like commonly used induction loops. Furthermore, due to the fact that they are able to monitor a wide area, they potentially offer the derivation of a whole new set of traffic parameters. Good examples are the derivation of source-destination relations, queue-length, travel-times or general event detection like untypical movements, accidents, blockages and congestions. Additionally, by using more than one camera the surveillance area can be enlarged or the detection accuracy can be increased due to redundancy of observations. However, in order to take advantage of a multiple camera system, the observations from different cameras have to be fused. In the setup that will be presented a geometric fusion is proposed by projecting the observations into a combined geo-referenced coordinate frame. The basic requirement for this transformation is the knowledge of the interior and exterior orientation of every camera. Three different approaches for determining the exterior orientation have been implemented, namely a Newton method, a least squares adjustment based on ground control points and a method based on line features. Furthermore, direct linear transformation and minimum space resection are applied to calculate initial estimates. These algorithms are subject to an in depth evaluation in respect to their application as a traffic monitoring sensor.

1. INTRODUCTION

1.1 Motivation

The task of modern traffic management is to utilize the limited resources in transportation infrastructure as efficient as possible. In order to meet the requirements of this challenging task, a precise and up-to-date knowledge of the traffic situation is needed. Nowadays, a wide variety of traffic sensors that are commercially available can be applied as a source of such information (Klein et al., 1997). Induction loops and microwave radar systems are the most commonly used detectors. They typically provide traffic parameters like presence, speed and length of a vehicle as well as time gap between vehicles. Since these parameters offer only a rudimentary description of the traffic situation a great deal of research has concentrated on new detectors capable of providing more complex traffic parameters.

Video image detection systems (VIDS) constitute an important group of traffic detectors (Michalopoulos, 1991; Wigan, 1992; Kastinaki et al., 2003). In contrast to the typical focus on a single location they are able to monitor a wide area, and hence, they potentially provide a whole new range of traffic parameters (Datta et al., 2000; Harlow and Wang, 2001; Setchell and Dagless, 2001; Yung and Lai, 2001). Such parameters can be source-destination relations, queue length, travel times or general event detection like untypical movements, accidents, blockages and congestions. Furthermore, a combination of several cameras is often beneficial to enlarge the surveillance area or to increase the detection accuracy due to redundancy of observations. However, in order to take advantage of a multiple camera system, the observations from different cameras have to be fused in a way that allows their subsequent comparison.

Information can be fused on different levels. Combining observations on a geometric level is a common approach for object detection by multi camera systems. This is done by projecting the observations into a combined coordinate frame. In general, a geo-referenced frame is preferred (Ernst et al., 2005). The knowledge of the interior and exterior orientation of every camera is an essential requirement for this transformation. The interior orientation can be deduced before camera installation using a well know test field. Different strategies can be applied to compute the unknown parameters of the exterior orientation of the final setup. The following approaches have been implemented:

1. Direct Linear Transformation (DLT)
2. Space Resection

for determining initial estimations and

3. Newton Method
4. Adjustment using Gauss-Markov and point features
5. Adjustment using Gauss-Markov and line features

to calculate the position and orientation of the cameras. These algorithms differ in their complexity, ease of use and their expected accuracy. The objective of this paper is to give an in-depth comparison in respect to these properties. Especially the relationship between needed scene information and achieved accuracy is highlighted with regard to the requirements of modern traffic surveillance.

1.2 Outline

The following chapter presents an example implementation of a multi-camera VIDS. The utilized processing chain is briefly presented in order to emphasize the need for precise knowledge of the exterior orientation. Chapter 3 introduces the different implemented approaches to determining the exterior orientation of the cameras. This is followed in chapter 4 by an evaluation of the approaches concerning their accuracy and usability in respect to the given task. The last chapter summarises and concludes the results of the evaluation.

2. APPROACH FOR A MULTI-CAMERA VIDS

2.1 Processing Approach

A multi-camera setup has been installed using three cameras to observe the traffic intersection Rudower Chaussee/ Wegedornstrasse, Berlin (Germany). The cameras cover overlaid or adjacent observation areas. Thus, the same road user can be observed by different cameras from different positions and angles (Figure 1). Using image processing methods the objects of interest can be found in the image data.

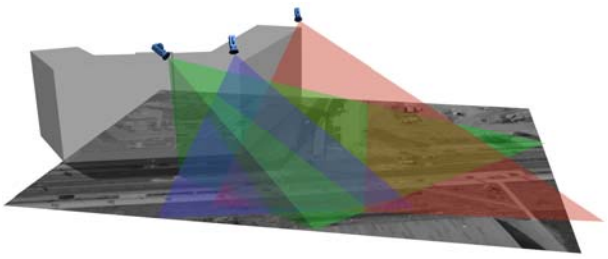

Figure 1. Visualisation of the multi-camera-setup

In order to enable the tracking and fusion of detected objects in the observation area the image coordinates of these objects are converted into a common world coordinate system. In case of poor quality of the orientation parameters, the same objects are observed from different positions. To avoid misidentification of objects derived from different cameras, a high precision transformation of their image coordinates into the object space coordinates is required. Therefore, a very exact calibration (interior orientation) as well as knowledge of the position and the view direction (exterior orientation) of the cameras is necessary.

The approach presented here can be separated into three main steps. Firstly, all moving objects have to be extracted from each frame of the video sequence. Next, these traffic objects have to be projected onto a geo-referenced world plane. Afterwards, these objects are tracked and associated to trajectories. This can be utilized to assess comprehensive traffic parameters and to characterize trajectories of individual traffic participants. These steps are described more precisely below.

2.2 Video Acquisition and Object Detection

In order to receive reliable and reproducible results, compact digital industrial cameras with standard interfaces and protocols (IEEE1394) are used.

To extract moving objects from an image sequence the image processing library OpenCV was utilized. The algorithm is based on a background estimator, which adapts to the variable background and extracts the desired traffic relevant objects. The extracted objects are then grouped using a cluster analysis combined with additional filters to avoid object splitting by infrastructure at intersections and roads.

The dedicated image coordinates as well as additional parameters like area, volume, color and compactness can be computed for each extracted traffic object.

2.3 Coordinate Transformation and Camera Calibration

The employed tracking concept is based on extracted objects, which are geo-referenced to a world coordinate system. This concept allows the integration or fusion of additional data sources as long as their observations can be transferred to the same coordinate system.

Therefore, a transformation between image and world coordinates is necessary for a multi-camera system. Using the collinearity equations (1), the world coordinates X, Y, Z can be derived from the image coordinates x', y':

$$X = X_0 + (Z - Z_0)\frac{r_{11}(x'-x_0) + r_{21}(y'-y_0) - r_{31}c}{r_{13}(x'-x_0) + r_{23}(y'-y_0) - r_{33}c} \qquad (1)$$

$$Y = Y_0 + (Z - Z_0)\frac{r_{12}(x'-x_0) + r_{22}(y'-y_0) - r_{32}c}{r_{13}(x'-x_0) + r_{23}(y'-y_0) - r_{33}c}$$

where X, Y = world coordinates (to be calculated)
Z = Z-component in world coordinates (to be known)
X_0, Y_0, Z_0 = position of the perspective center in world coordinates
r_{11}, r_{12}, …, r_{33} = elements of the rotation matrix
x', y' = uncorrected image coordinates
x_0, y_0 = coordinates of the principle point
c = focal length

The Z-component in world coordinates can be deduced by appointing a dedicated ground plane. Additional needed input parameters are the interior and exterior orientation of the camera. The interior orientation (principal point, focal length and additional camera distortion) can be determined using a well known lab test field. The 10 parameter Brown camera model was used for describing the interior orientation (Brown, 1971). The parameters can be determined by a bundle block adjustment as described in (Remondino and Fraser, 2006).

In order to calculating the exterior orientation of a camera, hence determining its location and orientation in a well known world coordinate system, different approaches can be applied. An important set of these approaches are presented and evaluated in the following chapters.

2.4 Tracking and Trajectory Creation

The tracking algorithm is supposed to provide object data information combined in a so-called state vector with respect to time. The state of an object can be described as position, velocity and acceleration in X-, Y- and Z-direction. Features like form, size and color can be added. The first task is the object identification in a video sequence by its predicted state

vector. This is done by observation-object-association (Kumar et al., 2005; Luo and Bhandarkar, 2005).

The tracking of every object was realized using a Kalman-filter (Anderson and Moor, 1979; Blackmann, 1986). It estimates the state of an object for the time stamp of the following picture, hence allows to compare the estimated state and the observed object data. If both are located within a certain feature space distance they can be associated to the same object. A considerable problem is initialization of the Kalman-filter.

The resulting trajectories are submitted to the analysis module as soon as they are created for the derivation of traffic parameters.

3. EXTERIOUR ORIENTATION

The collinearity equations (1) require the parameters of the exterior orientation of every camera. The following sections present two general approaches to determine these parameters based on different input sets of scene knowledge. The first algorithms use point correspondences between image points and measured points in the surveillance area. A differential GPS can be applied to acquire geo-referenced ground control points with a standard derivation usually below 2 cm. Other features that can be used are straight lines. Lines are a very common feature in urban environments. In contrast to ground control points, lines have the advantage of being easier to match to their correspondences in the image. Furthermore, this implies if these features are already geo-referenced on a floor-plane or in an orthophoto, the entire process of determining the exterior orientation could be automated.

The approaches that will be presented in the subsections 3.2 and 3.3 depend on initial values for the adjustment of the exterior orientation. With prior normalised images the values can be computed in advance by one of the following techniques.

3.1 Initial values

The *direct linear transformation* (DLT) method is based on the collinear equations which are extended by an affine transformation of the image coordinates (Abdel-Aziz and Karara, 1971; Kwon, 1989). Using these equations a system of linear equations can be set up and solved via well known methods. It results in 11 DLT parameters which define the exterior orientation, the focal length and the principal point. This method cannot detect erroneous measurements hence it relies on well measured image and world coordinate points. Another disadvantage is the liability to singularities if all control points are in a common plane. At least 6 measured points correspondences are needed.

An alternative approach is the *minimum space resection* (Fischler and Bolles, 1981). Given three points in object space and the projection center of the camera, a tetrahedron is defined. Knowing the 3 angles (derived from focal length and principal point) simple geometric dependencies can be established. By solving the resulting quartic equation the length of the three sides can be determined. The orientation of the camera is deduced by determining the intersection points of three spheres constructed using the object points as centers and the edge lengths as radius. This method requires 3 control points, the focal length and principal point. By taking into account a fourth point the ambiguous result is dissolved.

Using automated methods for determining control points or by taking into account the human factor, it is always an adequate approach to assume having unreliable ground control data. To exclude erroneous control points it is advised to apply the above procedures to minimal subsets of points. The final value will be the median of randomly chosen subset results. The number of subsets used depends on the amount of errors expected.

3.2 Adjustment using Control Points

Given the interior orientation and initial values for the exterior orientation the following algorithms can be applied to determine the exterior orientation (Luhmann et al. 2006, McGlone et al. 2004):

The *Newton method* is a common mean for retrieving the roots of a polynomial function. Thus it can easily be adapted for retrieving the parameters of the collinearity equations. After having set up the design matrix that is a least–squares estimator of a linearized model, singular value decomposition can be applied to solve the system of linear equations. This approach renders the detection of singular values, i.e. from planar control points, possible.

A *general least squares adjustment* based on a Gauss - Markov method computes the adjusted parameters of the exterior orientation. This method uses a system of normal equations. The dissolving of this system leads to the cofactor matrix of the unknowns as the inverse of the matrix of normal equations times the absolute term:

$$x = A^T P A^{-1} \cdot A^T P l \tag{2}$$

Hereby, the matrix P represents a stochastic model which can exclude erroneous points. The usage of trigonometric functions for setting up the necessary rotation matrix includes the usual ambiguity. Even though the geometric interpretation is difficult it is advisable to use quaternions for the definition of rotation. Despite the more difficult geometric interpretation quaternions are the appropriate mean to disambiguate the rotation.

3.3 Adjustment using straight lines

The collinearity equations can easily be extended for using straight lines as control data:

$$\begin{aligned} a &= \frac{-(\beta Z_0 + \delta - Y_0) r_{11} + (\alpha Z_0 + \gamma - X_0) r_{21} + (\alpha(\delta - Y_0) - \beta(\gamma - X_0)) r_{31}}{(\beta Z_0 + \delta - Y_0) r_{12} - (\alpha Z_0 + \gamma - X_0) r_{22} - (\alpha(\delta - Y_0) - \beta(\gamma - X_0)) r_{31}} \\ b &= c \frac{(\beta Z_0 + \delta - Y_0) r_{13} - (\alpha Z_0 + \gamma - X_0) r_{23} - (\alpha(\delta - Y_0) - \beta(\gamma - X_0)) r_{33}}{(\beta Z_0 + \delta - Y_0) r_{12} - (\alpha Z_0 + \gamma - X_0) r_{22} - (\alpha(\delta - Y_0) - \beta(\gamma - X_0)) r_{31}} \end{aligned} \tag{3}$$

$$\begin{aligned} g &: y = a \cdot x + b \\ G &: X = \alpha \cdot Z + \gamma \wedge Y = \beta \cdot Z + \delta \end{aligned} \tag{4}$$

Having an image line g and an object line G these equations can be substituted for the equations of control points. Hence the determination of exterior orientation is carried out in the same way as it would be done for points as control data.

4. EVALUATION

4.1 Test Setup

The implemented algorithms have been tested using simulated data as well as the described traffic intersection setup (TIS). The interior orientation parameters of the simulated and the real traffic camera can be found in Table 2:

Parameter	SIM	TIS
	in mm	in mm
x_0	-0.001008	0.848510
y_0	0.005275	-0.875919
c	6.950257	8.276964
K1	2.54744e-003	3.16630e-003
K2	-7.83759e-005	-2.17610e-005
K3	1.86310e-006	-1.04590e-006
P1	3.08255e-005	4.04520e-005
P2	-6.92956e-006	3.37660e-005
B1	-1.26747e-004	4.03890e-004
B2	-1.08686e-004	-1.15540e-004
Pixel size	0.00675	0.0065
Resolution	768x488 px	1024x768 px

Table 2. Interior orientation parameters of used cameras

In order to test the implementations with simulated data an array of 100 predefined control points is projected onto a virtual camera sensor. Hence the position of the projection center is predefined as well all values of this simulation setup are known with absolute accuracy. To derive an error behavior noise has been applied to the projected image points and defective control points are consecutively added to data space. The results are given as median values of 100 trials each.

A camera of the VIDS setup at the traffic intersection Rudower Chaussee/ Wegedornstrasse has been used to test the algorithms in their designated environment. 48 ground control points were taken via DGPS. Their corresponding image points were clicked using a designed tool with sub-pixel accuracy. Two ground control points were subsequently used to define lines in the scene to aid comparability. Figure 3 shows an example image of the chosen camera:

Figure 3. Image of a camera from the multi-camera VIDS

4.2 Initial values

When adding noise to the virtual image points, both DLT and the minimum space resection show a fast increasing rout mean square error (RMSE) when back-projecting the ground control points onto the virtual projection plane using the determined exterior orientation (Figure 5). The exterior orientation is unfeasible but the results of the minimum space resection suffice as initial values while the DLT seems too unstable.

Sampling over random minimal point sets and taking the median value proves efficient when adding erroneous control points to the initial data set (Figure 6). When using 150 trials, both approaches prove resilient to the false input up to a certain percentage. The DLT can cope with up to 8% of erroneous points and the minimum space resection with up to 16% before the results are compromised.

Applying both approaches to the traffic intersection setup emphasizes the simulation results. The DLT results with an RMSE of 16 pixel and fails completely with inaccurate input data while the minimum space resection yields an exterior orientation with a resulting RMSE of less then a pixel, even if 40% of the input data set consist of erroneous control points (Figure 7). Table 4 shows the resulting estimations of the exterior orientation. While the calculated position of the DLT varies up to half a meter and the rotation angles up to 2° from the final result of an adjustment approach, the results of the minimum space resection are remarkably close. Never the less, both estimations suffice as initial values.

	DLT	MSR	GMM
X	399899.02m	399899.51m	399899.50m
Y	5809757.70m	5809757.80m	5809757.80m
Z	92.12 m	92.10 m	92.10 m
ω	33.81°	35.68°	35.67°
φ	69.58°	70.62°	70.62°
κ	56.28°	55.26°	55.26°
RMSE	16,23 px	0.36 px	0.31 px

Table 4. Exterior orientations deduced by DLT, minimum space resection and the Gauss Markov method

4.3 Exterior Orientation

The results of the application of the algorithms to the simulated data set and subsequently adding noise to the image control points are visualised in Figure 8. The approaches based on control points yield a RMSE of less than half a pixel even under high noise. The Newton and the Gauss Markov approach, both using trigonometric functions have the same results. Because both apply the same method to solve an overdetermined system of equations they perform equally. The fundamental difference of Gauss-Markov in contrast to the Newton approach is the ability to detect and exclude erroneous points which were absent in this scenario. Using quaternions results in a slight decrease of accuracy. This can be explained by the fact that due to this point no constraints were defined during the adjustment to ensure the orthonormality of the rotation matrix. This leads to a small deviation of the rotation angles derived from the matrix. Using lines as an input feature results in a rapidly decreasing accuracy. Further analysis points towards the problem of line representation using slope-intersect-form. The discrepancy between expected and true slope of a line increases with its steepness. Thus, steep slopes are unproportionally

weighed higher in the adjustment, which distorts the result. A solution could be to rotate the coordinate axes for slopes higher then 45° (Schwermann, 1995).

The results of determining the exterior orientation of a chosen camera from the VIDS are shown in Figure 9. The RMSE is given in meter in object space to emphasize the maximum accuracy when using the camera as a traffic detector. All approaches based on control points converge to the same local minimum. Their RMSE is less than 0.05m up to a ground distance of 80m and less than 0.15m up to a ground distance of 140m. This theoretical ground sampling accuracy is more than sufficient for a traffic sensor. It potentially allows a correct lane assignment of detected vehicles over the whole observation area. The accuracy is higher then the expected localisation accuracy of traffic relevant objects in the scene yielded by image processing. Nevertheless, the Gauss Markov approach based on lines as input doesn't properly converge to the same local minimum, for the same reasons mentioned above. The RMSE is less than 0.8m up to a ground distance of 80m and less than 1.5m up to a ground distance of 140m.

Altering five of the input ground control points erroneous still leads to similar results for the algorithms based on adjustment. The worst case boundaries don't change (Figure 10). The Newton method performs out of scale because of the absence of a statistical model.

Figure 11 shows the results of the Gauss Markov approach for points using trigonometric functions and varying the amount of control points used. Surprisingly, this has no noticeable effect on the overall accuracy when using at least 10 control points.

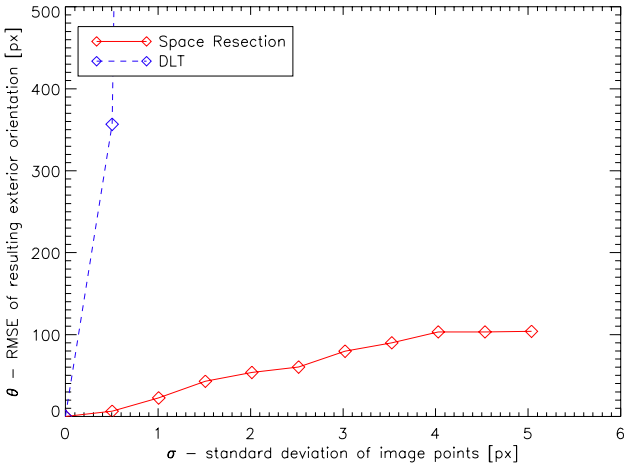

Figure 5. RMSE when adding noise to control points in simulated setup

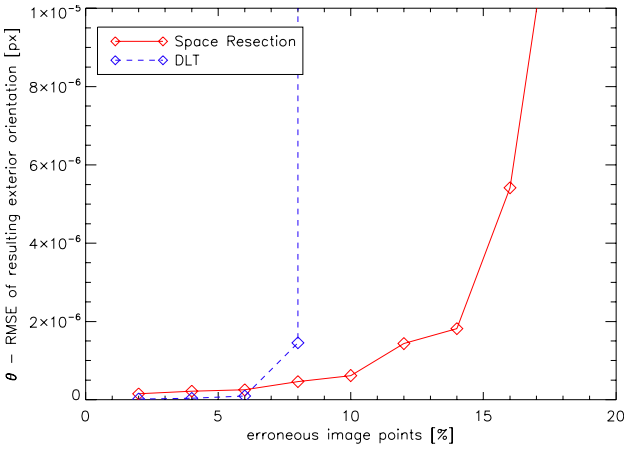

Figure 6. RMSE when adding erroneous image points in simulated setup

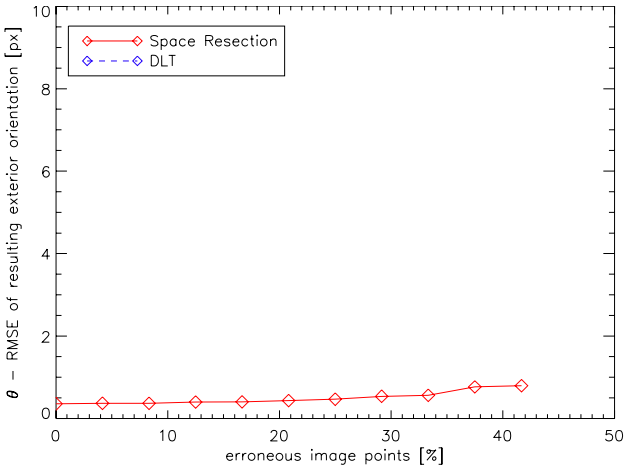

Figure 7. RMSE when adding erroneous image points in traffic intersection setup

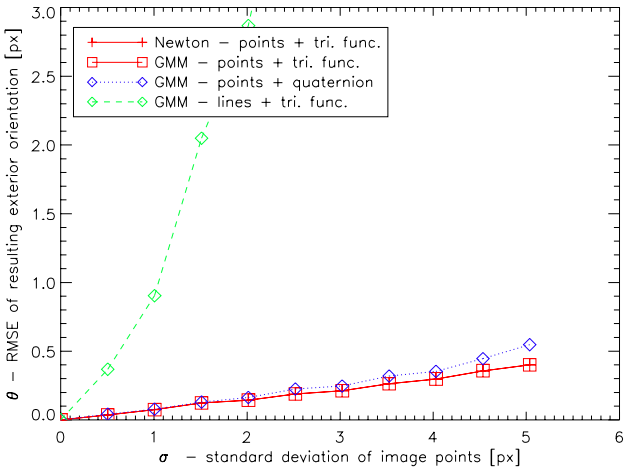

Figure 8. RMSE when adding noise to control points in simulated setup

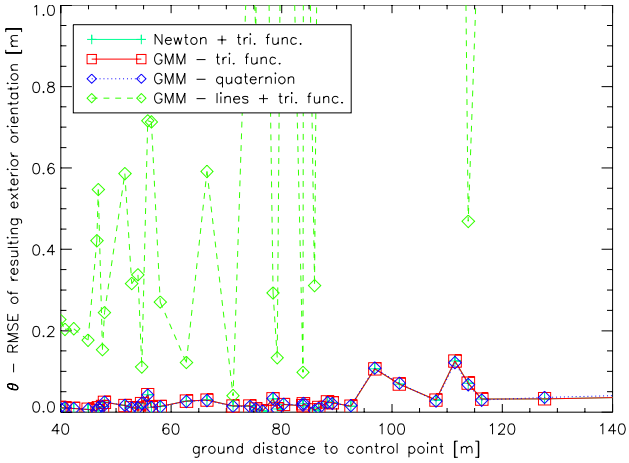

Figure 9. RMSE on ground level as a function of ground distance to the camera

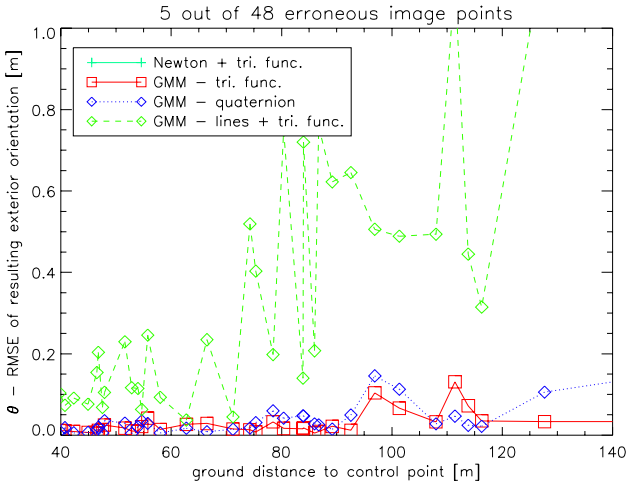

Figure 10. RMSE on ground level as a function of ground distance to the camera with 5 erroneous image points

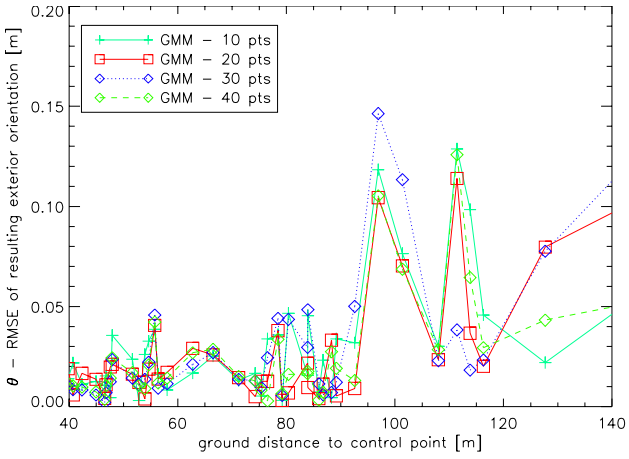

Figure 11. Gauss Markov adjustment for points with a varying amount of control points

5. CONCLUSION

Three different approaches for determining the exterior orientation of cameras for VIDS have been presented and thoroughly tested. The minimum space resection proved to be a very accurate and robust approach for determination of initial values and superior to the DLT.

In general the exterior orientations derived from point based approaches result in a more then sufficient accuracy for a traffic monitoring sensor. The RMSE is less than 0.05m in object space up to a ground distance of 80m and less than 0.15m up to a ground distance of 140m. While the Newton method is unable to cope with erroneous control points, the Gauss Markov approach remains widely unaffected. Furthermore, using quaternions avoids possible ambiguities.

Despite the encountered problems, using line features is a promising mean to determine the exterior orientation. Future research will focus on the stability and automation of calibration using line features.

REFERENCES

Abdel-Aziz, Y. and Karara, H.M, 1971. Direct linear transformation from comparator coordinates into object space coordinates in close range photogrammetry. *ASP Symposium on Close-Range Photogrammetry*, pp. 1-18.

Anderson, B. and Moor, J., 1979. *Optimal Filtering*. Prentice-Hall, Inc., Enlewood Cliffs, New Jersey.

Blackman, S.S., 1986. *Multiple-target tracking with radar applications*. MA: Artech House, Dedham.

Brown, D.C., 1971. Close range camera calibration, *Photogrammetric Engineering*, 37, no. 8, pp. 855-866.

Datta, T.K., Schattler, K. and Datta, S., 2000. Red light violations and crashes at urban intersections, *Highway and* Traffic Safety: Engineering, Evaluation, and Enforcement; Trucking and Motorcycles, 1734, pp. 52-58.

Ernst, I., Hetscher, M., Zuev, S., Thiessenhusen, K.U., Ruhé, M., 2005. New approaches for real time traffic data acquisition with airborne systems, *TRB Transportation Research Board*, TRB 2005, pp. 69-73.

Fischler, M.A. and Bolles, R.C., 1981. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartogrphy, *Communications of the ACM* 24 , vol. 6.

Harlow, C. and Wang, Y., 2001. Automated accident detection system, *Highway Safety: Modeling, Analysis*, Management, Statistical Methods, and Crash Location, 1746, pp. 90-93.

Kastrinaki, V., Zervakis, M. and Kalaitzakis, K., 2003. A survey of video processing techniques for traffic applications, *Image and Vision Computing*, vol. 21(4), pp. 359-381.

Klein, L.A., Kelley M.R. and Mills, M.K., 1997. Evaluation of overhead and in-ground vehicle detector technologies for traffic flow measurement. *Journal of Testing and Evaluation*, 25(2): p. 205-214.

Kumar, P., Ranganath, S., Huang, W.M. and Sengupta, K., 2005. Framework for real-time behavior interpretation from traffic video, IEEE Transactions on Intelligent Transportation Systems, vol. 6(1), pp. 43-53.

Kwon, Y.H., 1989. The effects of different control point conditions on the DLT calibration accuracy, http://www.kwon3d.com/theory/dlt/dlt.html, (accessed 05. Mai 2008)

Luo, X.Z. and Bhandarkar, S.M., 2005. Real-time and robust background updating for video surveillance and monitoring, *Image Analysis and Recognition*, vol. 3656, pp. 1226-1233.

Luhmann, T., Robson, S., Kyle, S. and Harley I., 2006. *Close-Range Photogrammetry*, Wiley

McGlone, C., Mikhail E. and Bethel J., 2004. *Manual of Phtogrammetry*, John Wiley & Sons Inc

Michalopoulos, P.G., 1991. Vehicle Detection Video through Image-Processing - the Autoscope System, *IEEE* Transactions on Vehicular Technology, vol. 40(1), pp. 21-29.

Remondino, F., and Fraser, C., 2006. Digital Camera Calibration Methods: Considerations and Comparisons, *ISPRS Commission V Symposium 'Image Engineering and Vision Metrology'*, pp.266-272.

Schwermann, R., 1995. Bildorientierung in Nahbereichs-photogrammetrie, PhD. Thesis RWTH Aachen

Setchell, C. and Dagless, E.L., 2001. Vision-based road-traffic monitoring sensor, *IEEE Proceedings-Vision Image and Signal Processing*, vol. 148(1), pp. 78-84.

Wigan, M.R., 1992. Image-Processing Techniques Applied to Road Problems, *Journal of Transportation Engineering-Asce*, vol. 118(1), pp. 62-83.

Yung, N.H.C. and Lai, A.H.S., 2001. An effective video analysis method for detecting red light runners, *IEEE* Transactions on Vehicular Technology, vol. 50(4), pp. 1074-1084.

Acknowledgements:

We would like to thank Ragna Hoffmann for the support in preparing this paper and Marcel Lemke, Björn Pilz for their support in the installation of the multi-camera VIDS.