

# AUTOMATIC IMAGE SEQUENCE REGISTRATION BASED ON A LINEAR SOLUTION AND SCALE INVARIANT KEYPOINT MATCHING

Z. Shragai, S. Barnea, S. Filin, G. Zalmanson, Y. Doytsher

Department of Transportation and Geo-Information, Technion – Israel Institute of Technology, Israel

{zivs, barneas, filin, zalmanson, doytsher}@technion.ac.il

## Commission III/1

**KEY WORDS:** Aerial Triangulation, Linear solutions, Scale invariant feature transform, Essential matrix, image correspondence

### ABSTRACT:

Automatic registration of image sequences has been a subject of research for many years, both in the photogrammetric and computer vision communities. As part of the automation, linear orientation methods are used to obtain approximations for a subsequent bundle adjustment solution. Linear solutions can be at time "too general" particularly in a sense that they mostly employ uncalibrated cameras, a fact leading to severely unstable results in most photogrammetric problems such as the case for the direct linear transformation (DLT) in a nearly flat terrain. Furthermore, to the best of our knowledge, none of them handle more than two or three images simultaneously without imposing several theoretical constraints that cannot be guaranteed in practical imaging missions. In this paper a sub-optimal linear solution for the exterior orientation parameters of image sequences is developed. The proposed method is demonstrated on an aerial image strip. The paper shows that the method successfully generates reliable and accurate approximations both for the orientation parameters as well as for tie point coordinates. For an automatic extraction of the latter, the Scale Invariant Feature Transform (SIFT) algorithm is applied.

## 1. INTRODUCTION

It is commonly accepted both in photogrammetry and computer vision communities that bundle adjustment is a "golden standard" method for recovering exterior orientation parameters from image sequences (Hartley et al., 2001). A bundle adjustment process requires, however, good initial values for all the six exterior parameter, as well as approximations for the 3D coordinates of the tie points. To avoid the need for approximations, a great deal of effort has been put on developing general algorithms that provide linear solutions to a variety of orientation problems (see e.g., Hartley et al., 2001; Rother and Carlsson, 2001; Carlsson and Weinshall, 1998). Many of them address a general problem in which the entire set of camera intrinsic (calibration) and extrinsic parameters is unknown. These solutions are stable and perform successfully only in cases where no limitations on either the acquisition geometry or the underlying object space are present. However, for typical photogrammetric problems these solutions have not yet proven useful. For example, the solutions proposed by Hartley et al. (2001) and Rother and Carlsson (2001) require a reference plane across any two images in a sequence. Carlson-Weinshall duality algorithm (1998) requires a specific number of points in a given number of images. Fitzgibbon and Zisserman (1998) offer the use of the trifocal-tensor in a close or open sequence. The trifocal-tensor does not suit, however, the photogrammetric process because of its requirement for tie points to appear in three sequential images. In the standard photogrammetric process, with 60 percent overlap between images, applying this model will relate to only 20 percent of each image. Furthermore, most of the works do not refer to the global exterior orientation parameters and produce only a relative solution. Pollefeys et. al (2002a) offer a solution that is

based on sequentially linking and reconstructing image after image, which is then followed by a bundle adjustment.

In this paper a framework for an automated photogrammetric solution is presented. Our objectives are reducing the operator input to a minimum and eliminating the reliance on initial values for the computation of the exterior orientation parameters. The proposed solution requires neither knowing the order of the images nor their overlapping percentage. The only external information required is the ground control points and their corresponding image points. Solutions that follow a similar line can be found in Nistér et al. (2004) where a sequence of video frames is oriented and in Oliensis (1997) where an iterative solution for weak motion (short baselines) image sequences is presented.

As an outline, our solution detects first tie points in image pairs. For this purpose the SIFT strategy (Lowe, 2004; Lowe 1999) is used as described in Section 2. Following the autonomous extraction of the tie point, comes the geometric computation. The proposed geometric framework is founded on the Essential matrix (Hartley and Zisserman, 2003). The Essential matrix between every image pair is calculated and the five relative orientation parameters are extracted. The geometric concept of the pose estimation and the scene reconstruction are given in Section 3. Section 4 presents experimental results and Section 5 concludes the paper.

## 2. EXTRACTION OF CORRESPONDING POINTS

The Scale Invariant Feature Transform - SIFT (Lowe, 2004; Lowe 1999) is a methodology for finding corresponding points in a set of images. The method designed to be invariant to scale,

rotation, and illumination. Lowe (2004) outlines the methodology as consisting of the following four steps:

1. Scale-space extrema detection – using the difference of Gaussian (DoG), potential interest points are detected.
2. Localization – detected candidate points are being probed further. Keypoints are evaluated by fitting an analytical model (mostly in the form of parabola) to determine their location and scale, and are then tested by a set of conditions. Most of them aim guaranteeing the stability of the selected points.
3. Orientation assignment – orientation is assigned to each keypoint based on the image local gradient. To ensure scale and orientation invariance, a transformation (in the form of rotation and scale) is applied on the image keypoint area.
4. Keypoint descriptor – for each detected keypoint a descriptor, which is invariant to scale, rotation and changes in illumination, is generated. The descriptor is based on orientation histograms in the appropriate scale. Each descriptor consists of 128 values.

With the completion of the keypoint detection (in which descriptors are created) the matching process between images begins. Matching is carried out between the descriptors, so the original image content is not considered here. Generally, for a given keypoint, matching can be carried with respect to all the extracted keypoints from all images. A minimum Euclidian distance between descriptors will then lead to finding the correspondence. However, matching in this exhaustive manner can be computationally expensive (i.e.,  $O(N^2)$  with  $N$  the number of keypoints). Common indexing schemes cannot be applied to improve the search here because of the descriptors dimensionality. However, an indexing paradigm, called Best Bin First (BBF), is proposed by Beis and Lowe, (1997). The BBF algorithm reduces the search to a limited number of the most significant descriptors values and then tries locating the closest neighbor with high probability. Compared to the exhaustive matching, this approach improves the performance by up to two orders of magnitude, while difference between the amount of matched points is small. Our proposed solution follows Schaffalitzky and Zisserman (2002) and Brown and Lowe (2003) where all key points from all images are organized in one K-d tree. Once a set of matching points has been generated, another filtering process is applied. This process is based on the RANSAC algorithm (Fischler and Bolles, 1981). The fundamental matrix of the image pairs is calculated and points that do not satisfy the geometric relation are filtered out as outliers. Based on the matching, the order of images within the image sequence is determined. When applying the SIFT method for aerial images the huge image size may lead to the extraction of numerous keypoints. Excess of information is valuable for redundancy; however, it comes with high computational cost. Experiments show, however, that even downscaling the aerial image resolution satisfying amount of keypoints has been provided. In comparative research presented by Mikolajczk and Schmid (2003) the SIFT method has shown superiority over classical methods for interest point detection and matching.

Figure 1 shows the matched keypoints on an extract of two overlapping aerial images. Generally, the algorithm extracted ~4000 keypoints per image, out of them 339 points were matched with less than 5 pixels offset between corresponding points. 146 keypoints have satisfied the geometric model with

less than 1 pixel between corresponding points. It is noted that seven points are needed for computing the Fundamental matrix. Experiments on different images with different characteristics (e.g., vegetation, urban scenes) exhibited similar results.

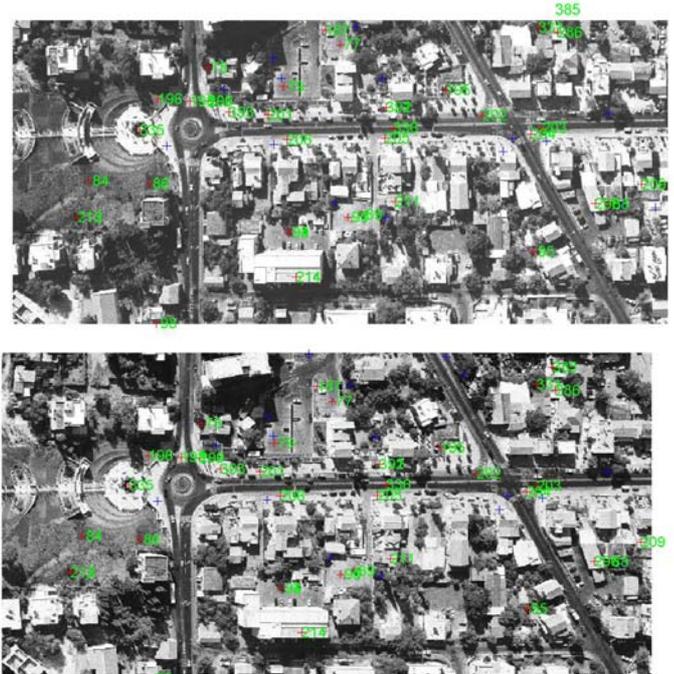


Figure 1. Matched keypoints in an aerial image pair extract

### 3. THE GEOMETRIC FRAMEWORK

The input for the geometric process is a set of matched points for all overlapping images. In addition, the Ground Control Points (GCPs) and their corresponding image points are provided. The solution considers the intrinsic parameters to be known. The process consists of two main steps: first is finding the relative orientation between all image pairs in the sequence. The second is a simultaneous computation of a transformation that takes into account the relative orientations and optionally the control points. This step is performed linearly as a single optimization process.

#### 3.1 Relative Orientation

The first step is the linear computation of the Essential matrix for each of the overlapping image pairs. The minimum number of required tie points ranges between five (Nistér, 2004; Philip, 1996) to seven (Hartley, 1997).

Extraction of the rotation and translation parameters from the Essential matrix can be carried out as proposed by Hartley and Zisserman (2003). We begin with a singular value decomposing of the Essential matrix:  $E=UDV^T$  where  $U$  and  $V$  are chosen such that  $\det(U)>0$  and  $\det(V)>0$ . Assuming that the first camera matrix is  $P = [I | 0]$ , the second camera matrix can be one of four possible choices:

$$P_1' = [UWV^T | u_3], P_2' = [UWV^T | -u_3],$$

$$P_3' = [UW^T V^T | u_3], P_4' = [UW^T V^T | -u_3]$$

with  $u_3$  the third column of  $U$ , and

$$W = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

A reconstructed point  $X$  will be in front of both cameras only in one of the four possible solutions. Thus, testing with a single point to determine if it is in front of both cameras is sufficient for the choice between the four possible solutions of  $P'$  (Hartley and Zisserman, 2003). To fine-tune the relative orientation parameters, a non-linear geometric optimization can now take place.

An important issue to account for is the degeneracy of the Essential matrix which arises in the following cases (Torr et al., 1999):

1. All points and camera centers laying on quadratic surface (e.g., cone, cylinder).
2. There is no translation between the images.
3. All tie points lie on the same plane in object space.

Cases (1) and (2) are also a degeneracy of the bundle adjustment algorithm. Cases (2) and (3) are more common. For these cases there is a simpler geometrical model – the Homography. From a Homography one can retrieve the relative orientation parameters as proposed by (Tsai et al., 1982). To choose between the Essential matrix and the Homography, Torr et al. (1999) proposes a measure they call Geometric Robust Information Criterion (GRIC) that computes scores to the fitness of the geometrical model for a given dataset. This measure is also used by Pollefeys et al. (2002b). An alternative way to avoid the degeneracy as in case (3) is using the five point algorithm (Philip, 1996; Nistér, 2004). However, then a tenth degree polynomial must be solved.

### 3.2 Global Registration

Following the computation of the relative orientation parameters, we are provided with two camera matrices for each image - one, which is fixed (when the image is the first in the pair) and the other, which is relative (when the image is the second). The first and the last images have only one camera matrix. The task of concatenating the relative orientation parameters into one global model is divided into two subtasks: concatenating rotations and concatenating translations. The first subtask can be described by a recursion formula:

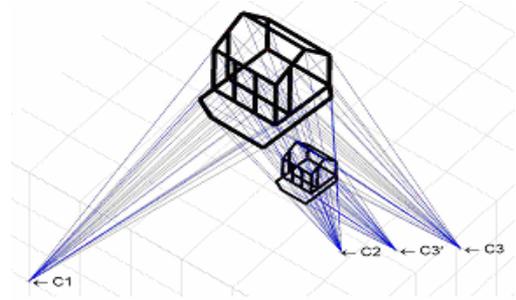
$$R_{i+1} = R_m^{i \rightarrow i+1} R_i \quad \text{Where } R_1 = I_{3 \times 3} \quad (1)$$

where  $R_m^{i \rightarrow i+1}$  is the rotation in the  $m$ -th model between the images  $i$  and  $i+1$ . Concatenating the camera centers (translation) in the sequence (the second subtask) is a more complicated process. Here, similarly to the first subtask, there are two translation vectors for each image in the sequence (apart of the first and last) one is fixed (in the origin) and the other is relative. However, in contrast to the rotations, with the

translation concatenation all vectors are defined up to a scale factor only. The scale ambiguity of each vector affects the size of the reconstructed scene from each image pair, as Figure 2 demonstrates. In Figure 2,  $C_1$  and  $C_2$  are the camera centers of the first and the second images.  $C_3$  is the actual position of image 3, so the scale of the translation vector  $t_{23}$  is correct – the scenes reconstructed from images 1, 2 and images 2, 3 fit. Contrary to  $C_3$ , a camera position in  $C_3'$  leads to reconstructed scenes that differ in scale. The recursion formula of the translation concatenation should, therefore, have the form of:

$$t_{i+1} = t_i + s_m t_m^{i \rightarrow i+1} \quad \text{Where } t_1 = [0,0,0]^T \quad (2)$$

$s_m$  and  $t_m$  are the scale factor and the translation vector of the  $m$ -model between images  $i$  and  $i+1$ .



**Figure 2.** Influence of the translation scale factor on the reconstructed scene.

For solving all the translation scale factors together with the tie point coordinates we now develop a simultaneous and linear solution. The solution is derived from the camera matrix,  $P$  that fulfills the relation  $x=PX$ , with  $X$  the coordinate vector of a point in object space, and  $x$  is the image coordinate vector. Both are given in homogenous coordinates (the last term of  $X$  and  $x$  is set to 1).  $P$  may be decomposed into:

$$P = KR[I | -t] \quad (3)$$

with  $K$  is the camera calibration matrix and  $I$  a 3x3 identity matrix. By substituting (1) and (2) into (3) a recursion formula for the  $P$  matrices can be written as

$$P_{i+1} = K \cdot R_m^{i \rightarrow i+1} R_i \cdot [I | t_i + t_m^{i \rightarrow i+1} \cdot s]$$

leading when inserted into the  $x=PX$  relation to

$$x_{i+1} = K \cdot R_{i+1} \cdot [I | t_i + t_m^{i \rightarrow i+1} \cdot s] \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (4)$$

As  $K$  and  $R_i$  are known  $\forall i$ , they are of no interest. We, therefore, rewrite Equation (4) as follows

$$\hat{x}_{i+1} = [I \mid \sum_{m=1}^i t_m^{i \rightarrow i+1} \cdot s_m] \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (5)$$

with  $\hat{x} = (KR)^{-1}x$ . Equation (5) provides a linear form for the estimation of the point coordinates, X, Y, and Z, and the scale  $s$ . Notice that with this model a point is reconstructed from all its instantiations in all images. Each image point contributes two independent equations. There is still one ambiguity left, namely the scale of the first model. This ambiguity is solved by the absolute orientation (into the object space reference frame). Generally, for each of the components (i.e., tie points and camera matrices) one has to find a similarity transformation,  $X_w = H_s X_m$ , to the object space reference frame via the GCPs, with  $H_s$  of the form:

$$H_s = \begin{bmatrix} R & t \\ 0^T & \lambda \end{bmatrix} \quad (6)$$

and  $\lambda$  as the model scale. Linear solutions to this problem have been offered by several authors, e.g., a quaternion based solution (Horn, 1987), orthogonal matrices (Horn et al., 1988) and the Rodriguez matrix (Pozzoli and Mussio, 2003).

An approach that simultaneously integrates the solution for the scale parameters, tie point coordinates and the absolute orientation parameters is now presented. For a control point that appears in an image, it is possible to use equation (7)

$$x = PH_s^{-1} X_w \quad (7)$$

with  $P$  as any projection matrix in the model space that acquires the point  $X_w$ , and  $H_s$  given in Equation (6). In a simultaneous solution, the scale factor  $\lambda$  in  $H_s$  can be replaced by the scale factor as given in Equation (2) for the first image pair.  $H_s$  becomes now an Euclidian transformation with only six parameters, where  $\lambda = 1$ .

Substituting  $H_s^{-1}$  into equation (4) and multiplying both sides by  $(KR)^{-1}$  will lead to:

$$\hat{x}_{i+1} = [I \mid \sum_{m=1}^i t_m^{i \rightarrow i+1} \cdot s_m] \cdot \begin{bmatrix} R^T & \hat{T} \\ 0^T & 1 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}_{World} \quad (8)$$

with  $H_s^{-1} = \begin{bmatrix} R^T & \hat{T} \\ 0^T & 1 \end{bmatrix}$ . Equation (8) can be rearranged as:

$$\hat{x}_{i+1} = [R^T \mid \hat{T} + \sum_{m=1}^i t_m^{i \rightarrow i+1} \cdot s_m] \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}_{World} \quad (9)$$

Equation (9) provides a linear form for the estimation of the scale factors  $s_m$ , the global translation  $\hat{T}$  and the nine rotation matrix terms. In this representation a 3D affine transformation is solved. This model requires at least four control points. Restricting the solution to a 3D rotation (namely maintaining the orthonormality) can be achieved by using the identity matrix instead of the singular values in the SVD of  $R$ . Using Equation (5) for tie points and (9) for control points, we are provided with a simultaneous and linear solution. This solution allows having the external effect of control points and the internal constrains of the tie points weighted in simultaneously. Furthermore, control points that appear in only one image can also be taken into account. This solution offers an alternative to the two steps procedure. However, it is noted that it is not optimal in the sense of solving nine parameters explicitly instead of an orthonormal rotation matrix. Experiments with this method yield good results only under specific configurations.

## 4. EXPERIMENTAL RESULTS

The proposed method is now investigated using synthetic and real data. The sensitivity of the geometric model to additive Gaussian noise is tested first, followed by an application of the process on a strip consisting of four images.

### 4.1 Synthetic Data

A synthetic configuration that follows typical mapping-mission characteristics was designed with the following parameters, flying altitude, 1700 m, terrain variation ranging between 0-200m, and a focal length of 153 mm. The test set consisted of four images in a sequence with 60 percent overlap. The pitch and roll angles were in the range of  $\pm 2^\circ$ . For each image pair ~50 tie points were provided. Six ground control points were used. To investigate the sensitivity of the proposed to random errors Gaussian noise with zero mean and standard deviation ranging between 0.0 and 0.3 mm has been added to image coordinates of control and tie points. The maximum standard deviation (0.3 mm) is equivalent to an error of 20 pixels for scanning resolution of 15 $\mu$ .

Given this input, fundamental matrices were computed and normalized by the known interior camera parameters to form the Essential matrix. Then, a decomposition of the Essential matrix to the rotation and translation components was carried out, followed by up to five (non-linear) iterations to optimize the computed  $R$  and  $t$  values. The transformation into a global reference frame was computed using Equations (5) and (6). Rodriguez matrices were used to represent rotations. For each noise level 100 trials were performed. Results were evaluated by three measures: the *std.* of the 3D Euclidean distance between the computed object point coordinates and the actual ones, both for tie and control points (Figure 3), the offsets in the camera positions, again in terms of *std.* of the 3D Euclidean distances (Figure 4) and the angular error of the three camera rotation parameters (Figure 5). Results were compared to bundle adjustment solution, as shown in Figures 3-5. The experiments show that even in the presence of a severe noise reasonable and acceptable solutions can be achieved by the proposed geometric model. Indeed, bundle adjustment solution performs better than the sub-optimal solution, which is of no surprise, but the fact that the results obtained using our method

do not fall too far from the optimal solution makes it a good candidate to precede any subsequent optimal solution. Also, the deviations in orientation parameters fairly compare with accuracies obtained with typical GPS/INS systems. Furthermore, under realistic noise level, these results satisfy the requirements of some applications – thus avoiding a subsequent use of bundle adjustment.

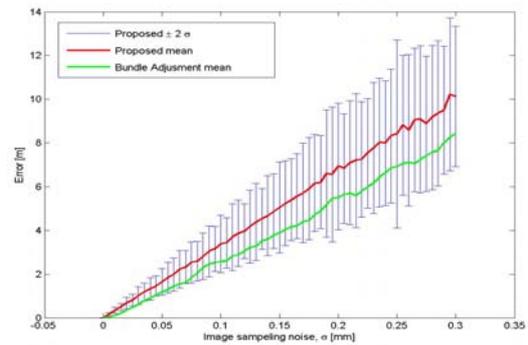
#### 4.2 Real Images

An experiment with a strip consisting of four aerial images with flying altitude of 1800 m, and a focal length of 152 mm is now presented. Eight GCPs were available for this image set. The four images are arranged in an L shape form (see Figure 6); their order is not provided as an input. The image coordinates of the GCPs were manually digitized. Tie points were generated using the SIFT procedure. Globally there were ~1000 matched keypoints. About 300 matched points between images with similar orientation (image pairs 1-2 and 3-4), and about 60 matched points for image pair 3-4. Between image triplets about 10 common points were detected.

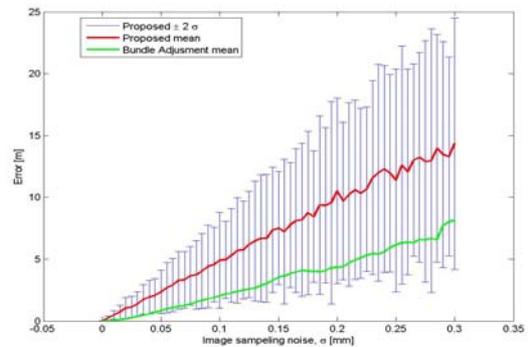
To evaluate the quality of the two-steps method the orientations were computed first by this procedure only, and then using a bundle adjustment solution. For the bundle adjustment solution the parameters originating from the linear procedure were used as initial approximations. To evaluate the difference between solutions we compare the reconstructed tie point coordinates between the two-steps solution and the bundle adjustment. Results show that the mean distance between the two methods is 0.33 m. However, the accuracy estimate of the points achieved by the bundle adjustment procedure is about  $\pm 1$  m. This difference is within the uncertainty range of the tie points coordinates. These results are in agreement with those achieved by the synthetic data experiments in Section 4.1 and indicate that the proposed method can be used as an independent solution when achieving high level of accuracy is not a concern and also as an initial values generator for a bundle adjustment solution.

#### 5. SUMMARY AND CONCLUSIONS

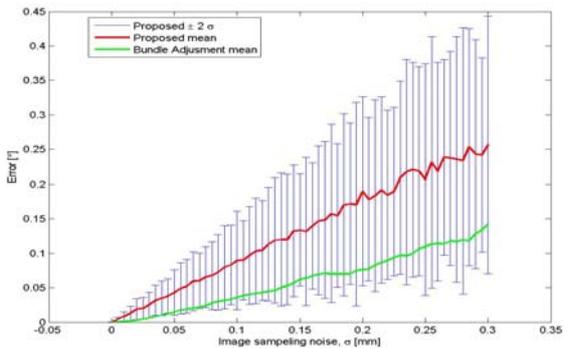
Recent years have seen a significant progress made in automation of registration processes. At the same time advances have been made in the field of multi-view geometry. This paper has demonstrated the integration of these two disciplines. No assumptions on the order of the image sequence have been made to execute the proposed linear solution for estimating the camera parameters. Experiments made have demonstrated robustness and stability of the proposed geometric solution even to severe noise levels. Those with real data showed that even with non-standard image configuration a full automation can be achieved.



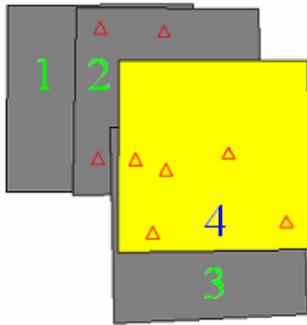
**Figure 3.** Mean error of the reconstructed points. The X-axis is the noise level in millimeters and the Y-axis represent the ground error (distance) in meters. The error bars represent  $\pm 2\sigma$  of the accuracy range as resulted from the trials for each noise level.



**Figure 4.** Mean error of the reconstructed image positions parameters. The X-axis is the noise level [mm] and the Y-axis represents the image positions error (distance) [m]. The error bars represent  $\pm 2\sigma$  of the accuracy range as resulted from the trials for each noise level.



**Figure 5.** Mean error of the reconstructed camera angles. The X-axis is the noise level in mm and the Y-axis represent the angular error [°]. The error bars represent  $\pm 2\sigma$  of the accuracy range as resulted from the trials for each noise level



**Figure 6.** Outline of the aerial image arrangement used for the experiment. Triangles depict control points.

## 6. REFERENCES

- Beis J.S., Lowe D.G., 1997. Shape Indexing Using Approximate Nearest-Neighbour Search in High-Dimensional Spaces. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1000-1006.
- Carlsson S., Weinshall D., 1998. Dual Computation of Projective Shape and Camera Positions from Multiple Images, *International Journal of Computer Vision* 27(3), pp.227-241.
- Fischler M.A., Bolles R.C., 1981. Random Sample Consensus: A paradigm for model fitting with application to image analysis and Automated Cartography. *Communication Association and Computing Machine*, 24(6), pp. 381-395.
- Fitzgibbon A.W., Zisserman A., 1998. Automatic Camera Recovery for Closed or Open Images Sequences. in Proceedings fifth European Conference on Computer Vision (ICCV), 1998, pp. 311-326.
- Hartley R. I., 1997. In defense of the eight-point algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19, pp. 80-93.
- Hartley R., Dano N., Kaucic R., 2001. Plane-based Projective Reconstruction. Proceedings of the Eighth IEEE International Conference on Computer Vision (ICCV), 2001, Vol. 1 pp. 420-427.
- Hartley R., Zisserman A., 2003. *Multiple View Geometry in Computer Vision*. Cambridge University Press, Second Edition.
- Horn B., 1987. Closed-form solution of absolute orientation using unit quaternions. *Journal of the Optical Society of America* 4(4) pp.629-642.
- Horn B., Hilden H.M., Negahdaripour S., 1988. Closed-form solution of absolute orientation using orthonormal matrices, *Journal of the Optical Society of America* vol. 5, no.7, pp. 1127-1638.
- Lowe G. D., 1999. Object Recognition from Local Scale-Invariant Features. Proceedings of the sixth International Conference on Computer Vision (ICCV), Vol. 2, p. 1150
- Lowe G. D., 2004. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision* 60(2) pp. 91-110.
- Mikolajczk, K., Schmid C., 2003. A performance evaluation of local descriptors. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Vol. 2, pp. 257-263.
- Nistér, D., 2004. An efficient solution to the five-point relative pose problem. *IEEE transaction on pattern analysis and machine intelligence* 26(6) pp.756-770.
- Nistér D., Naroditsky O., Bergen J., 2004. Visual Odometry, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Vol. 1, pp. 652-659.
- Philip J., 1996. A non-iterative algorithm for determining all essential matrices corresponding to five point pairs. *Photogrammetric record* 15(88), pp. 589-599.
- Pollefeys M., Verbiest F., Van Gool L., 2002a. Surviving dominant planes in uncalibrated structure and motion recovery. Proceedings of the seventh European Conference on Computer Vision (ECCV), Vol. 2, pp. 837-851.
- Pollefeys M., Van Gool L., Vergauwen M., Cornelis K., Verbiest F., Tops J., 2002b. Video-to-3D, Proceedings of Photogrammetric Computer Vision (PCV), *International Archive of Photogrammetry and Remote Sensing*, 34(3A), pp. 252-257.
- Pozzoli A., Mussio L., 2003. Quick solutions particularly in close range photogrammetry. *International Archives of the Photogrammetry, Remote Sensing*, 34(5/W12) pp. 273-278.
- Rother C., Carlsson S., 2001. Linear multi view reconstruction and camera recovery, Proceedings of the Eighth IEEE International Conference on Computer Vision (ICCV), Vol. 1, pp. 42-50.
- F. Schaffalitzky and A. Zisserman., 2002. Multi-view matching for unordered image sets, or "How do I organize my holiday snaps?". Proceedings of the 7th European Conference on Computer Vision, pp. 414-431.
- Torr P. H. S., Fitzgibbon A. W., Zisserman A., 1999. The Problem of Degeneracy in Structure and Motion Recovery from Uncalibrated Image Sequences, *International Journal of Computer Vision*, Vol. 32 (1), pp. 27-44.
- Tsai R. Y., Huang T. S., Zhu W. L. 1982. Estimating three-dimensional motion parameters of a rigid planar patch, ii: Singular value decomposition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 30, pp. 525-534.