ORIENTATION OF IMAGE SEQUENCES ACQUIRED FROM UAVS AND WITH GPS CAMERAS

Jan Bartelsen and Helmut Mayer

Institute for Applied Computer Science, Bundeswehr University Munich, Neubiberg, Germany
Jan.Bartelsen@unibw.de, Helmut.Mayer@unibw.de – www.unibw.de/ipk/photo

KEY WORDS: Automatic Image Orientation, Unmanned Aerial Vehicles, GPS cameras

ABSTRACT:

This paper presents an approach for orienting images taken with different cameras from unknown positions. Specifically, we show that we can relatively orient wide-baseline images taken by lightweight (micro) Unmanned Aerial Vehicles (UAVs) of about 1 kg overall weight and from the ground. For the latter, we employ a consumer GPS camera which allows us to upgrade relativ oriented 3D models to absolute orientation. For the remainder we assume that the camera calibration is at least weakly known.
Relative orientation based on matching point features as well as absolute orientation with consumer GPS cameras both suffer from possibly extreme outliers. We tackle this challenge by combining consensus-based filtering of meaningful observations with robust adjustment. While robust techniques based on weighting down observations alone are susceptible to even single extreme outliers (leverage points), techniques such as RANdom SAmple Consensus – RANSAC (Fischler and Bolles, 1981) make it possible to determine many or most of the meaningful observations (inliers) independently of leverage points. As solutions by consensus-based approaches will be correct, but might be poor, we employ a recursive procedure in the spirit of the expectation maximization (EM) algorithm, where the improved estimate of robust adjustment is employed to iteratively find extended sets of inliers. Results demonstrate the potential of the proposed approach.

1 INTRODUCTION

Automatic 3D reconstruction of urban areas from digital images is an active research area. (Pollefeys et al., 2004) presents earlier work. (Pollefeys et al., 2008) or the recent (Agarwal et al., 2009) show that it is feasible to automatically generate 3D models from thousands of images, even when they are taken only weakly constrained from the Internet. Yet this is restricted to pairs for which the viewing angle is not too large. While 3D data from laser scanners might have a higher reliability than from automatic image matching, scanners are costly and heavy. Compared to this, images from consumer cameras are an easily available and cheap data source.

We have also developed an approach in line of the above work. It particularly excels by being capable to deal with large viewing angles between pairs of images. This is achieved by combining Förstner points (Förstner and Gülch, 1987), correlation, least-squares matching, the 5-point algorithm (Nistér, 2004), and RANdom SAmple Consensus – RANSAC (Fischler and Bolles, 1981) with robust bundle adjustment. The key idea is to achieve reliability by means of a highly precise solution (computed by least-squares matching and robust bundle adjustment) for many points which is unlikely to arise by chance. By employing the direct 5-point algorithm (Nistér, 2004), no approximate values for position and attitude are needed and RANSAC allows us to deal with much less than 50% correct matches.

We investigated how far we can deal with images from very small Unmanned Aerial Vehicles (Micro-UAVs), because our approach has no need for approximations for position and attitude of the camera. In the last two years we made several experiments with Micro-UAVs with a diameter of about 1 meter and a weight under 1 kg (including the camera). They enable to take images of buildings and their facades from an off-ground perspective independently of ground conditions or obstacles on the ground. We could demonstrate that 3D reconstruction from images from such Micro-UAVs is feasible (Mayer and Bartelsen, 2008).

As we aim at a as complete as possible 3D modeling of urban areas, particularly buildings, we have started to combine terrestrial images with images from UAVs. This is especially promising, as Micro-UAVs are capable to fly through streets only a couple of meters above the ground, impossible with heavier UAVs. Therefore, the viewpoints are not so different than from the ground, both showing facades and partially also the roofs.

To estimate the absolute orientation, we have made experiments with cameras with an integrated GPS receiver. They determine a position in the world coordinate system whenever an image is taken. Because the GPS in the type of UAV we have used is not synchronized with the camera, we have only done experiments with a hand-held GPS camera. We particularly have used the Ricoh Caplio 500SE, which produces GPS positions, but there is no access to the raw data and thus no post-processing of the GPS signal to improve the precision is possible. While our approach for 3D reconstruction from images produces highly precise local models of the scene, the GPS camera positions allow by means of a 3D similarity transformation to estimate the absolute orientation for the model. A larger number of GPS camera positions allows an adjustment for the 3D similarity transformation.

Both, relative orientation of images based on matching point features as well as absolute orientation employing consumer GPS cameras suffer from possibly extreme outliers. For the images these are wrong matches, particularly if a wide-baseline is used, while the GPS camera might produce totally wrong values when satellites are in an inappropriate constellation or are occluded as in common in urban environments at which we aim at.

We tackle the above challenge by combining consensus-based filtering of meaningful observations with robust adjustment. While robust techniques based on weighting down observations are susceptible to single extreme outliers (leverage points), consensus-based techniques such as RANSAC render if possible to determine many or most of the meaningful observations (inliers) independently of leverage points. Once a correct approximation is generated via RANSAC, we refine it by robust adjustment using

weighting down as well as eventually throwing out weak points. As approximations by RANSAC are correct, but might be poor, we employ a recursive procedure in the spirit of the expectation maximization (EM) algorithm, where the improved estimate of the robust adjustment is employed to find an extended set of inliers, which is then used to find a further improved estimate, etc.

In the next Section, robust parameter estimation based on consensus-based filtering and robust adjustment is discussed as the key contribution of this paper. We introduce our approaches for relative and absolute orientation in Sections 3 and 4. Results are presented in Section 5. We end up with an outlook and conclusions.

2 ROBUST PARAMETER ESTIMATION

The combination of different robust techniques is a core of this paper. While robust adjustment allows to weight down observations which are (in a certain sense) close to the correct solution, but are still not part of it, it cannot deal with single extreme outliers (leverage points).

For the latter, consensus-based approaches such as RANSAC come to the rescue, at least if certain conditions hold:

- The solution can be computed from a limited number of observations to the latter avoiding combinatorial explosion.
- A criterion has to exist to tell good from bad observations (outliers), which can be evaluated based on the above solution.

RANSAC and similar consensus-based approaches work by randomly taking a minimum number of observations and computing a solution for them. (E.g., relative orientation can be computed from five point pairs, or a plane from three 3D points.) Then, all other observations not used for computing the solution are evaluated according to the above criterion and the number of the observations fulfilling the criterion is counted (**inlier**). E.g., for relative orientation one would check the distances from the corresponding epipolar lines or for planes simply the distance of the points from the plane. This is repeated a number of times always randomly choosing new observations to compute a solution. (Once one has a solution and thus the percentage of inliers, one can compute how many iterations are still necessary given a certain probability level (Fischler and Bolles, 1981).)

Counting inliers is in many cases not the best possible solution, though, as it does not account for the different qualities of inliers: They can be very close to the expected solution or at the very margin of the acceptance region. A way to take into account this information is to employ the Geometric Robust Information Criterion – GRIC (Torr, 1997). A constant penalty is given to all outliers, while the inliers are evaluated by a typically squared distance metric. The goal is thus to minimize the GRIC sum for all observations instead of maximizing the number of inliers.

While consensus-based approaches allow to come up with a correct solution for extreme outliers, even when the percentage of outliers is much larger than the 50% known as the breakdown-point for robust adjustment, the solution is often not extremely good as only a small part of the data is used. Thus, different people came up with the idea to improve the estimate inside the consensus-based approach by using more observations than just the minimum number leading, e.g., to locally optimized RANSAC (Chum et al., 2003).

We do not only combine a consensus-based approach with least squares optimization, but we robustify the latter and use it in an expectation maximization (EM) based fashion, iterating always between the E and the M step (start point is an approximate solution for a minimum number of observations):

- For the current solution all inliers are determined (E).
- Based on all inliers, an improved solution is obtained using robust least squares adjustment (M).

While in many instances only very few iterations are sufficient, there are examples where starting from observations in a relatively small area the number of inliers grows and grows by improving the solution by means of the inliers step by step.

We base robust adjustment on standardized residuals $\bar{v}_i = v_i/\sigma_{v_i}$ involving the standard deviations σ_{v_i} of the residuals, i.e., the differences between observed and predicted values. As a first means we employ reweighting with $w_i = \sqrt{2 + \bar{v}_i^2}$ (McGlone et al., 2004). Additionally, having obtained a stable solution concerning reweighting, outliers characterized by $\bar{v}_i$ exceeding a threshold, which we have set to 4 in accordance with theoretical derivations and empirical findings, are eliminated for the next iteration.

While many people use the average standard deviation of the residuals, we have found that computing the standard deviation for each and every residual although costly (it involves computing the covariance matrix of the observations and multiplying it with the design matrix from both sides) still leads to a much faster convergence and better results, often more than balancing the effort involved in the additional computations.

Usually it is sufficient to employ the above procedure only for the best hypothesis of a larger number, e.g., 100, of RANSAC iterations, as it is costly to use it for obviously bad hypotheses.

3 AUTOMATIC RELATIVE ORIENTATION OF IMAGE SEQUENCES

We have developed an approach for orienting images from consumer cameras (Mayer, 2008) which works without previous knowledge or approximations for position and attitude and without markers. While it is in principle possible to also work without knowledge about calibration (Mayer, 2005) using a direct self-calibration approach such as proposed in (Pollefeys et al., 2004), we resort in this paper to the assumption, that an approximate calibration in the range of a couple of percent is known, which is reasonable for many applications. Using calibration does not only allow to deal with totally planar scenes, but it also makes 3D reconstruction more efficient by constraining the solution space.

Relative orientation of images is based on patches around Förstner points (Förstner and Gülch, 1987). They are correlated taking into account the direction of the eigenvectors of the points making the matching in-plane-rotation invariant. If the correlation is above a low threshold, the patches are least-squares matched using the affine transformation as geometrical model and resulting in highly precise relative point positions. While we found that this procedure allows to match points from very different directions, i.e., is highly off-plane rotation invariant, it cannot cope with larger scale differences. Thus, we have started to experiment with Scale Invariant Feature Transform – SIFT features (Lowe, 2004), even though they are limited to similar view angles.

For image pairs, the relative sub-pixel coordinates of the centers of the patches are input to essential matrix calculation combining the direct 5-point algorithm (Nistér, 2004) needing no approximate values with RANSAC and robust bundle adjustment (cf. Section 2).

For image triplets, the 5-point algorithm is applied twice and the relative scale of the two pairs is computed as the median. Triplets are used as basic building block because for them points can be checked. Overlapping triplets are linked to form image sequences or blocks. All the above is done for images on higher levels of the image pyramids to reduce the computational complexity. The final result for the block is tracked down to the original resolution of the images. There, a final bundle adjustment is conducted. As we usually have larger scenes with sufficient 3D structure, the calibration parameters can be improved in the final bundle adjustment of the whole image block.

We employ a correction for radial distortion

$$ds = 1.0 + k_2 * (r^2 - r_0^2) + k_4 * (r^4 - r_0^4)$$

with r the distance to the principal point and r_0 the distance where ds is 0. From a larger number of experiments we have found, that it is important to employ radial distortion estimation only after robust adjustment.

To be able to work with images from different cameras, we link every image with a description of its camera parameters consisting of the calibration matrix as well as the radial distortion parameters. By using one description for each camera and not each image, we are able to reliably improve the camera parameters in the bundle adjustment, which would not be the case if the parameters would be optimized per image.

The approach procedure has been tested for hundreds of image sequences, also with images of beyond ten Megapixels and with up to several hundred images. Fig. 1 gives an example depicting the camera positions as colored pyramids with the tips of the pyramids symbolizing the positions of the projection centers of the cameras and the bases of the pyramids the image planes. The images for the green and red pyramids were taken at two different UAV flights while the blue pyramids are for images of the GPS camera acquired on the ground. 3D Points are given as colored spheres, with the color taken from the images.

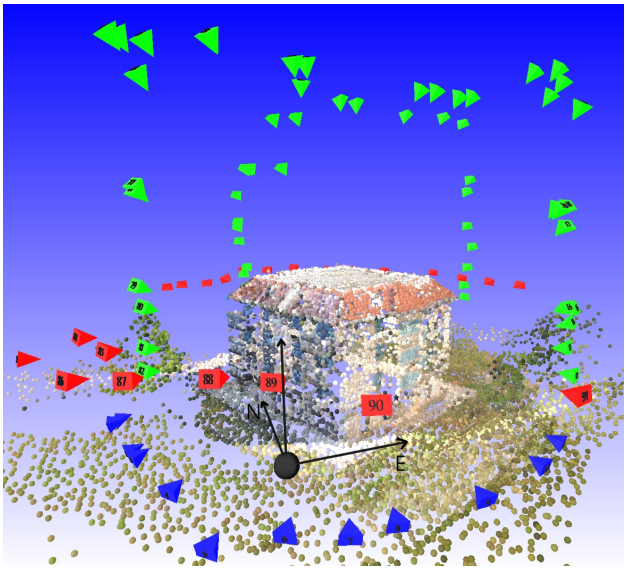


Figure 1: Building 51: Positions and attitudes of cameras for Micro UAV flights around and above the building (red and green pyramids) as well as of a consumer GPS camera (blue pyramids). The latter was used to absolutely orient the image block. 3D points are given as spheres with the colors taken from the images.

4 AUTOMATIC ABSOLUTE ORIENTATION USING CONSUMER GPS CAMERAS

4.1 3D Similarity Transformation Between Local and World Coordinate Systems

As detailed above, relative orientation produces a local 3D model of the scene in a Euclidean coordinate system. The mutual relation between this coordinate system and a Euclidean world coordinate system can be established by means of a 3D similarity transformation (Luhmann, 2000). Prerequisites for this are that both coordinate systems are Euclidean and three pairs of corresponding points exist whose respective vectors are linear independent.

The WGS84 world coordinate system is based on a reference ellipsoid and is thus not Euclidean. As the camera Ricoh Caplio 500SE that we have used produces WGS84 coordinates, a conversion into a Cartesian world coordinate system is necessary. The global standard Universal Transverse Mercator (UTM) system is particularly suited for this. Formulas for the conversion of WGS84 coordinates to UTM are given in (Schödlbauer, 1995). At the moment we do not consider, that UTM coordinates are restricted to a certain range of longitude, as we do only work in relatively small areas. We do use, though, a correction function included in the camera, transforming ellipsoidal heights to heights related to the Geoid, although with a limited precision.

The 3D similarity transformation from the local POS_{local} to the UTM position POS_{UTM} reads:

$$POS_{UTM} = Scale \times Rotation \times POS_{local} + Translation$$

$Translation$, $Scale$, and $Rotation$ are determined as follows based on the known camera positions in the local as well as in the UTM coordinate system: First, the center of gravity of the camera positions is computed for both coordinate system. The origin of the local coordinate system is shifted to its center of gravity. The center of gravity of the cameras in the world coordinate system defines the translation between model and world coordinate system. Additionally, the center of gravity is used as the center for the rotation. As GPS measurements tend to have a comparably bad precision for the height and we have worked in relatively planar areas until now, we assume that the camera positions in the world coordinate system have the same average height, mitigating errors in the height measurements. Yet, we note that for images taken on a slope, or from a UAV on different heights, this might be problematic and will need additional effort.

Because some GPS measurements might be grossly wrong, a consensus-based approach (cf. Section 2) is used to calculate an approximation for the transformation. For computing an individual hypothesis for the transformation, a minimum set of three different camera positions suffices, for each of which a position could be measured in the world coordinate system. The three points should neither be too close nor lie on a line.

Because we are dealing at the moment with a small number of camera positions with GPS measurements (usually less than 30), solutions for all possible triplets (full combinatorial search) are computed and the one with the highest number of inliers is cho-

sen. To deal with larger numbers, random sampling in the spirit of RANSAC will be used.

4.2 Adjustment of Local and GPS Camera Positions

The above Section presents means to compute the transformation between the local and the world coordinate system from three points. Usually, though, there are more points available and thus one wants to compute an improved estimate based on an adjustment of the inliers. For this, the following functional model has been developed:

- Observations: Camera positions in world coordinate system (E – easting, N – northing, and h – height) and related positions in the local 3D model (x, y, and z)
- 7 Unknowns: Rotation (3 parameters in rotation matrix R or elements of normed quaternions a, b, c, and d), scale (1 parameter m), and translation (3 parameters t_x, t_y, and t_z)

$$f :: \begin{pmatrix} E \\ N \\ h \end{pmatrix} = mR \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix}$$

with

$$R = \begin{pmatrix} 1-2(c^2+d^2) & 2(bc-ad) & 2(bd+ac) \\ 2(bc+ad) & 1-2(b^2+d^2) & 2(cd-ab) \\ 2(bd-ac) & 2(cd+ab) & 1-2(b^2+c^2) \end{pmatrix}$$

The elements of normed quaternions a, b, c, and d guarantee the uniqueness of the 3D rotation over all values. Yet, they are redundant. To reduce the number of unknowns for the adjustment, the largest of the four values is kept constant for an iteration, varying only the rest of the values while normalizing them.

The equations are linearized at the values of the approximation and then the system is iteratively solved.

5 RESULTS

In the following, we present results of our experiments demonstrating the potential of the proposed approach.

5.1 Estimation of Absolute Orientation

To verify our approach by comparing it to reference data, we conducted an experiment with reference point measurements using the SAtellite POsitioning Service (SAPOS) of the German State Survey allowing for position measurements with a very high precision of 1 to 3 centimeters including height. For the determination of the exact positions of the GPS camera in relation to the reference measurements, we used a special tripod which fixes the camera very close to the antenna of a Leica GPS system using the SAPOS service. For the reduction of the heights to the Geoid, data from the State Survey were employed. Thus, reference camera positions were available to estimate the precision of the measurements of the GPS camera as well as to verify the results of our adjustment. The GPS camera uses the NMEA-0183 protocol with an internal simplified Geoid reduction, which caused an error of about 3 meters for this image-sequence.

In the following, we compare the results of the adjustment with the measured GPS positions and the reference positions from SAPOS. We created an image sequence of building 20 of UniBw

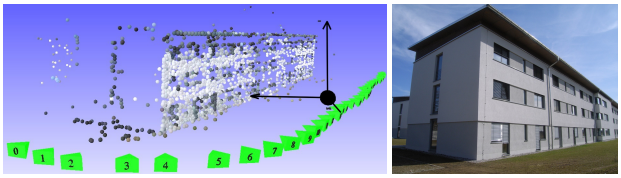


Figure 2: Left: 3D model of building 20 of Bundeswehr University Munich (UniBw). The big black sphere marks the center of the camera positions. Right: One image from the sequence.

with 33 images made by a Ricoh Caplio 500SE GPS camera in combination with our SAPOS system. The estimated precision of the bundle adjustment for the relatively oriented 3D model of building 20 is σ_0 = 0.33 pixels, which is of average quality. The impreciseness of the local model is cause of uncertainness in addition to the measured GPS positions. In spite of this Fig.3 shows a comparison which demonstrates the precision of the estimated absolute orientation.

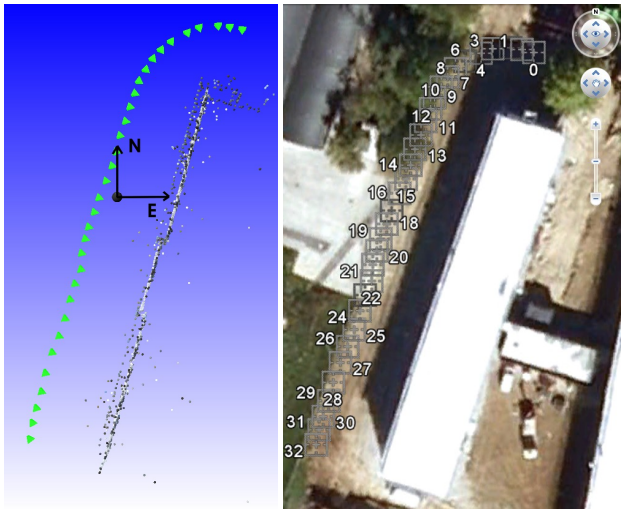


Figure 3: Left: Oriented 3D model of building 20 of UniBw from the top. Right: Oriented building 20 and camera positions in Google Earth.

We have selected three 3D similarity transformations. The first one is calculated from the SAPOS camera positions which have a precision of 1 to 2.5 centimeters. A precision like that is only possible under favorable conditions, even with SAPOS. We used this result as reference data without adjustment. As second, one approximation estimated from the GPS camerapositions and finally the result of the adjustment. In Table 1 and 2 the details of the 3D similarity transformations are illustrated.

The comparison has been done for building 20 of Bundeswehr University Munich (UniBw) (cf. Fig. 2).

	a	b	c	d	m
SAPOS	0.017	0.032	0,771	0.636	4.431
GPS Cam	0.028	-0.022	0.793	0.608	5.004
Adjusted	0.046	0.007	0.781	0.623	4.567

Table 1: Elements of normed quaternions and scalefactors of estimated 3D similarity transformations.

For the 33 camera positions we estimated the distances of the adjusted positions from the SAPOS reference positions and compared them to the distances of the initial measurements. The results in Table 3 and 4 show that the adjusted values are much better concerning dispersion and precision.

	Easting	Northing	Height
SAPOS	696794.88m	5328327.63m	547.98m
GPS Cam	696794.46m	5328328.76m	544.85m
Adjusted	696794.44m	5328328.81m	544.86m

Table 2: Translations of estimated 3D similarity transformations. The experiment was made in the UTM Zone 32U.

	Average	Variance	Worst	Best
GPS Cam	2.11m	$2.80m^2$	8.38m	0.41m
Adjusted	1.36m	$0.23m^2$	2.10m	0.70m

Table 3: Distances of initial and adjusted camera positions from SAPOS reference positions on the ground.

Fig. 4 presents another example. The information from the GPS camera has been used to absolutely orient the images. Additionally, the vertical direction has been determined from vanishing points of vertical structures on walls, vertical planes were fit to the 3D points, texture from the images was mapped on these planes, and, finally, all this has been integrated into Google Earth.

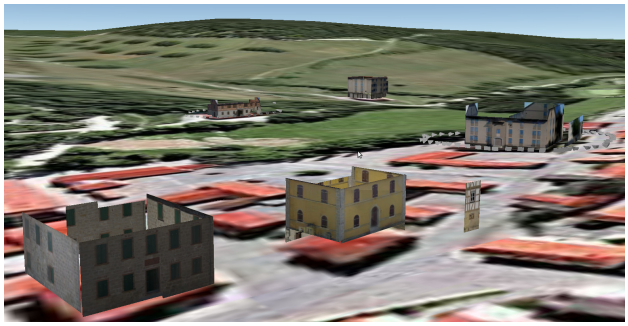


Figure 4: For several flights around buildings with a Micro UAV vertical planes have been determined from the 3D points generated during relative and absolute orientation, texture was mapped on the planes, and, finally, the resulting 3D models were included into Google Earth.

Fig. 5 shows building number 33 of UniBw, Germany. Again, the vertical direction has been estimated from the vanishing points and thus vertical walls were determined Fig. 5 (a) and (b) make clear that the measured GPS camera positions are sometimes several meters off the correct positions. Please particularly note the order of the positions in (a). After absolute orientation (c), the camera positions are much more plausible.

5.2 Combination of Several Cameras and Additional Estimation of Absolute Orientation

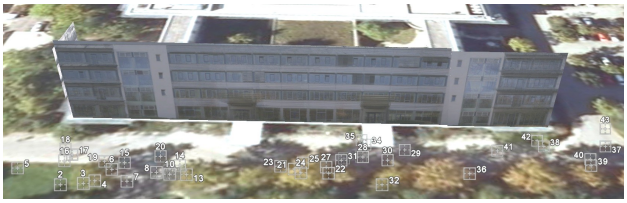
The following results for building 51 already introduced in Section 3 we firstly demonstrate, that we can determine the relative orientation for image sequences taken from different Micro UAVs with different cameras at different times (August 2008 and July 2009 – cf. Fig. 1). Additionally, the resulting 3D model could be upgraded to absolute orientation by acquiring suitable images by a GPS camera from the ground. Fig 6 illustrates the precision of the estimated orientation by comparing the absolutely oriented 3D model to the image from Google Earth at the same position.

6 OUTLOOK AND CONCLUSIONS

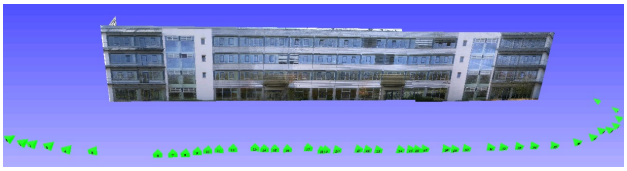
This paper presents an approach for robust estimation combining consensus-based techniques such as RANSAC which allow to obtain correct estimates also for extreme outliers (leverage points)

	Average	Variance	Worst	Best
GPS Cam	4.32m	$7.41m^2$	12.16m	0.42m
Adjusted	3.44m	$0.01m^2$	3.69m	3.26m

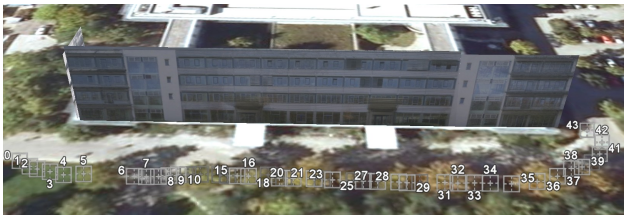
Table 4: Distances of initial and adjusted camera positions from SAPOS reference positions.



(a) Original GPS camera positions shown in Google Earth.



(b) 3D relative model including facade planes and camera positions.



(c) Adjusted camera positions shown in Google Earth

Figure 5: Building 33 of UniBw, Germany (a) Building in the form of the reconstructed facade plane in Google Earth and the camera positions measured via the GPS in the camera when taking images (b) 3D model with the facade plane and the camera positions after relative orientation. (c) Camera positions after absolute orientation.

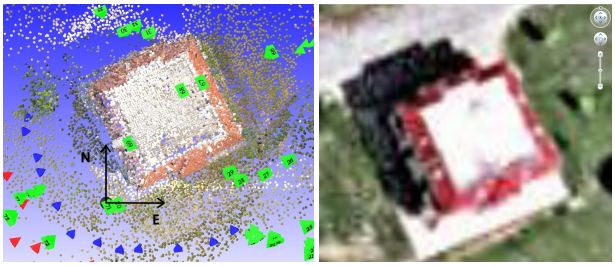
with robust adjustment. We include the latter in an iterative algorithm in the spirit of Expectation Maximization (EM), where inliers from the E-step are employed to compute a better estimate in the M-step which then lead to more inliers, etc.

We employ the robust approach to estimate relative and absolute orientation even when much more than 50% outliers exist in image matching or for occluded GPS measurements. Practical results show the potential of our approach for robust estimation.

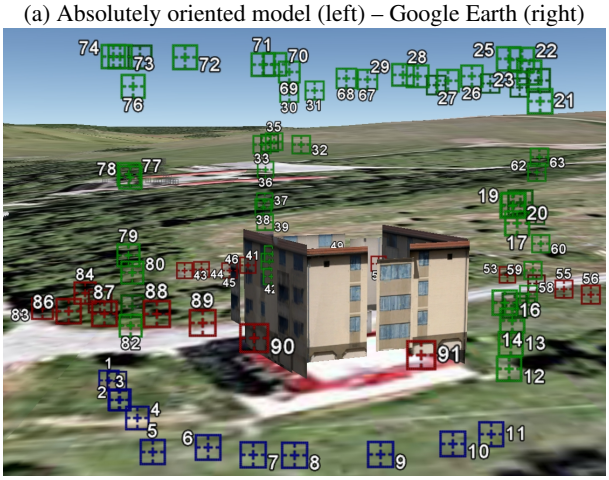
Until now, we only have made use of one type of GPS camera, namely the Ricoh Caplio 500SE, which offers no access to the raw data. Post processing would deliver more accurate camera positions and thus give initial values with lower uncertainty. We expect more consumer GPS cameras to be available in the near future.

A basic limit of our approach are the occlusions of the GPS in urban areas, particularly as we need to get close to the buildings to obtain highly detailed 3D models. Nevertheless, we have demonstrated for several urban areas that it is still feasible. In addition to the occlusions, also the satellite constellation is important. Under uncertain conditions which can be quantified the Dilution Of Precision (DOP), the initial GPS positions are too scattered. We have found that we need a DOP of better than 3 to obtain meaningful results.

As the vertical direction is not very precisely determined even



(a) Absolutely oriented model (left) – Google Earth (right)



(b) Camera positions in Google Earth

Figure 6: Absolutely oriented 3D model of building 51 (a) Left: Oriented Model – Right: Image from Google Earth with the same orientation. (b) Camera positions in Google Earth (blue: terrestrial images of GPS camera – red and green: images from Micro UAVs.

for many GPS measurements, we plan to make use of the fact that we aim at urban areas where many vertical lines exist. Thus, we are at the moment including our scheme described in (Mayer and Reznik, 2007) in our system. This allows to estimate the vanishing points of the vertical lines very precisely and reliably leading together with the calibration to the vertical direction.

Concerning future research, we note that no approach for image matching, which is one foundation of our approach, can yet handle at the same times large scale variances and wide viewpoint changes. Yet, this is typical for combined image sequences, e.g., from the UAV and from the ground, that we use. While scale variances in sequences with little viewpoint changes can be handled by SIFT Features (Lowe, 2004), which we have integrated into our approach as one option instead Förstner Points (Förstner and Gülch, 1987), no approach can handle larger viewpoint changes, i.e., above 50°, yet.

REFERENCES

Agarwal, S., Snavely, N., Simon, I., Seitz, S. and Szeliski, R., 2009. Building Rome in a Day. In: Twelfth International Conference on Computer Vision.

Chum, O., Matas, J. and Kittler, J., 2003. Locally Optimized RANSAC. In: Pattern Recognition – DAGM 2003, Springer-Verlag, Berlin, Germany, pp. 249–256.

Fischler, M. and Bolles, R., 1981. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. Communications of the ACM 24(6), pp. 381–395.

Förstner, W. and Gülch, E., 1987. A Fast Operator for Detection and Precise Location of Distinct Points, Corners and Centres of Circular Features. In: ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data, Interlaken, Switzerland, pp. 281–305.

Lowe, D., 2004. Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision 60(2), pp. 91–110.

Luhmann, T. (ed.), 2000. Nahbereichsphotogrammetrie. Herbert Wichmann Verlag, Heidelberg.

Mayer, H., 2005. Robust Least-Squares Adjustment Based Orientation and Auto-Calibration of Wide-Baseline Image Sequences. In: ISPRS Workshop in conjunction with ICCV 2005 "Towards Benchmarking Automated Calibration, Orientation and Surface Reconstruction from Images" (BenCos), Beijing, China, pp. 1–6.

Mayer, H., 2008. Issues for Image Matching in Structure from Motion. In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. (37) 3a, pp. 21–26.

Mayer, H. and Bartelsen, J., 2008. Automated 3D Reconstruction of Urban Areas from Networks of Wide-Baseline Image Sequences. In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. (37) 5, pp. 633–638.

Mayer, H. and Reznik, S., 2007. Building Facade Interpretation from Uncalibrated Wide-Baseline Image Sequences. ISPRS Journal of Photogrammetry and Remote Sensing 61(6), pp. 371–380.

McGlone, J., Bethel, J. and Mikhail, E. (eds), 2004. Manual of Photogrammetry. American Society of Photogrammetry and Remote Sensing, Bethesda, USA.

Nistér, D., 2004. An Efficient Solution to the Five-Point Relative Pose Problem. IEEE Transactions on Pattern Analysis and Machine Intelligence 26(6), pp. 756–770.

Pollefeys, M., Nistér, D., Frahm, J.-M., Akbarzadeh, A., Mordohai, P., Clipp, B., Engels, C., Gallup, D., Kim, S.-J., Merrell, P., Salmi, C., Sinha, S., Talton, B., Wang, L., Yang, Q., Stewénius, H., Yang, R., Welch, G. and Towles, H., 2008. Detailed Real-Time Urban 3D Reconstruction from Video. International Journal of Computer Vision 78(2–3), pp. 143–167.

Pollefeys, M., Van Gool, L., Vergauwen, M., Verbiest, F., Cornelis, K. and Tops, J., 2004. Visual Modeling with a Hand-Held Camera. International Journal of Computer Vision 59(3), pp. 207–232.

Schödlbauer, A., 1995. Rechenformeln und Rechenbeispiele zur Landesvermessung II. Wichmann Verlag, Karlsruhe, Germany.

Torr, P., 1997. An Assessment of Information Criteria for Motion Model Selection. In: Computer Vision and Pattern Recognition, pp. 47–53.