

# DETECTION OF CRITICAL CAMERA CONFIGURATIONS FOR STRUCTURE FROM MOTION

Mario Michelini and Helmut Mayer

Institute of Applied Computer Science  
Bundeswehr University Munich  
mario.michelini@unibw.de, helmut.mayer@unibw.de

**KEY WORDS:** Critical configurations, structure from motion, motion degeneracy, image triplets, image sets, 3D reconstruction

## ABSTRACT:

This paper deals with the detection of critical, i.e., poor or degenerate camera configurations, with a poor or undefined intersection geometry between views. This is the basis for a calibrated Structure from Motion (SfM) approach employing image triplets for complex, unordered image sets, e.g., obtained by combining terrestrial images and images from small Unmanned Aerial Systems (UAS). Poor intersection geometry results from a small ratio between the baseline length and the depth of the scene. If there is no baseline between views, the intersection geometry becomes undefined. Our approach can detect image pairs without or with a very weak baseline (motion degeneracy). For the detection we have developed various metrics and evaluated them by means of extensive experiments with about 1500 image pairs. The metrics are based on properties of the reconstructed 3D points, such as the roundness of the error ellipsoid. The detection of weak baselines is formulated as a classification problem using the metrics as features. Machine learning techniques are applied to improve the classification. By taking into account the critical camera configurations during the iterative composition of the image set, a complete, metric 3D reconstruction of the whole scene could be achieved also in this case. We sketch our approach for the orientation of unordered image sets and finally demonstrate that the approach is able to produce very accurate and reliable orientations.

## 1 INTRODUCTION

Our basic goal is to derive 3D structure and camera projection matrices from calibrated, but unordered image sets, where the motion is not known *a priori*. This is used as a basis for dense 3D reconstruction or image interpretation, e.g., of facades (Mayer and Reznik, 2007).

Most Structure from Motion approaches begin with the automatic determination of point correspondences between views of the image set. In the course of the reconstruction geometric constraints arising from scene rigidity and a general camera configuration are assumed to hold. However, problems arise if the assumed scene structure and/or camera configurations do not conform to these assumptions.

The first problem regards the scene geometry and occurs if the viewed scene structure is planar (*structure degeneracy*). The point correspondences are then related by a homography. Since for the fundamental matrix  $F$  holds  $F = [\mathbf{e}_2]_x H$  (where  $[\mathbf{e}_2]_x$  is the skew-symmetric matrix corresponding to the epipole  $\mathbf{e}_2$  in the second image and  $H$  the homography matrix), there exists a two parameter family of solutions for the epipolar geometry. Thus, the estimation of epipolar geometry would lead to a random solution based on the inclusion of outliers. For uncalibrated cameras there are approaches (Chum et al., 2005, Torr et al., 1999, Pollefeys et al., 2002b) which try to detect this case by comparing the models based on epipolar geometry and homography. Yet, the whole problem does not occur for calibrated cameras if the five point algorithm (Nistér, 2004) or (Li and Hartley, 2006) is used, what we do.

The second problem is more problematic and concerns camera configurations. Usually, a general camera configuration with translation and/or rotation of the camera between images is assumed. In absence of translation there remains only a pure rotational movement (*motion degeneracy*) and images are related by the infinite homography  $H_\infty$ . There exists no baseline between images and the epipolar geometry is undefined. For triangulation

based 3D reconstruction the accuracy of the reconstruction is proportional to the ratio between the baseline length and the depth of the scene. Thus, camera configurations without baseline may not be used and those with a very short baseline should be avoided in order to achieve a reliable and accurate reconstruction.

Critical camera configurations were analyzed in the context of *keyframe selection* approaches (Pollefeys et al., 2002a, Repko and Pollefeys, 2005, Thormählen et al., 2004, Beder and Steffen, 2006). There, image pairs which are most suitable for the estimation of the epipolar geometry and for which the triangulation is particularly well-conditioned are called *keyframes*. Hence, keyframes are those image pairs which comprise a sufficient baseline, so that the initial estimation of the 3D structure is reliable. (Pollefeys et al., 2002a) as well as (Repko and Pollefeys, 2005) estimate fundamental matrix and homography for each relevant image pair. Keyframes are selected from image pairs for which the fundamental matrix is found to be the more appropriate motion model based on the Geometric Robust Information Criterion – GRIC (Torr, 1998). Because they work with uncalibrated images, both approaches are not able to distinguish between structure degeneracy and the more critical motion degeneracy. Keyframe selection based on the result of the bundle-adjustment of the whole image set is proposed in (Thormählen et al., 2004). Unfortunately, the runtime of this method does not scale well. In (Beder and Steffen, 2006) the mean of the roundness of the error ellipsoids of the reconstructed 3D points as derived by bundle adjustment is used for keyframe selection. It is an efficient method which works on calibrated images and thus can be used to detect critical camera configurations.

From our point of view the problem with all these methods is that they are designed for keyframe selection and we found them to be unreliable for the detection of critical camera configurations. In keyframe selection the image pair with the highest score is used as keyframe, whereas for the detection of critical camera configurations one has to define a threshold, if the described approaches are to be used.

We think that the detection of critical camera configurations should be formulated as a binary classification problem. For such a problem various algorithms exist, e.g., (Breiman, 2001, Cortes and Vapnik, 1995). One often used approach is AdaBoost (Freund and Schapire, 1995) which belongs to the ensemble classifiers. It is based on the idea of creating a highly accurate prediction rule by combining many relatively weak and inaccurate rules.

In this paper we present an analysis of various metrics to determine no or a very weak baseline between views. We employ Machine Learning techniques and a classification algorithm based on AdaBoost which can detect critical camera configurations very reliably. By taking into account the critical camera configurations during the iterative composition of the image set from triplets, a complete, metric 3D reconstruction of the whole scene could be achieved. We shortly present our approach for the orientation of unordered image sets, in which the detection of critical camera configuration will be integrated, and demonstrate that it is able to produce accurate and reliable orientations for complex image sets.

The paper is organized as follows: In Section 2 we define several metrics which are to be analyzed concerning their suitability for the classification of critical camera configurations. An extensive evaluation of the metrics is presented in Section 3. In Section 4 our orientation framework for unordered image sets is described and results are presented. Finally, in Section 5 conclusions are given and future work is discussed.

## 2 ERROR METRICS

In this section we define several metrics, which will be analyzed concerning their suitability as features for classification in the remainder of the paper.

(Beder and Steffen, 2006) proposed an algorithm to determine the best initial image pair for a calibrated multi-view reconstruction based on the error ellipsoids of the reconstructed 3D points. The quality of a reconstructed 3D point is estimated by the roundness  $R$  of the error ellipsoid which is defined as

$$R = \sqrt{\frac{\lambda_3}{\lambda_1}} \quad (1)$$

where  $C$  is the covariance matrix and  $\lambda_1 \geq \lambda_2 \geq \lambda_3$  are eigenvalues of  $C$ .  $R$  lies between 0 and 1 and only depends on the relative geometry of the two cameras and the feature positions. If the two camera centers are identical and the feature positions were correct, the roundness would be equal to zero. For keyframe selection (Beder and Steffen, 2006) compute the mean roundness  $R_{mean}$  for all reconstructed points for an image pair. From a statistical viewpoint the mean is more sensitive to noise and thus less robust than the median. Hence, we compute also the median roundness  $R_{med}$  over all reconstructed points.

Motivated by the roundness  $R$  we have developed several other metrics based on the form of the error ellipsoid which take not only two, but all axes of the ellipsoid into account. We will show in Section 3 that this is more discriminative for the detection of critical camera configurations.

An error ellipsoid is defined by

$$(\mathbf{x} - \mathbf{p})^T C^{-1} (\mathbf{x} - \mathbf{p}) = 1$$

where  $C$  is the symmetric covariance matrix,  $\mathbf{p}$  the reconstructed point and  $\mathbf{x}$  a point on the ellipsoid. The eigenvectors of  $C$  define

the directions of the semi-axes and the eigenvalues  $\lambda_1 \geq \lambda_2 \geq \lambda_3$  the squares of the lengths of semi-axes  $a, b, c$ , i.e.:

$$a = \sqrt{\lambda_1} \quad b = \sqrt{\lambda_2} \quad c = \sqrt{\lambda_3} \quad (2)$$

The volume  $V$  of the error ellipsoid is given by the formula

$$V = \frac{4}{3} \pi abc = \frac{4}{3} \pi \sqrt{\lambda_1 \lambda_2 \lambda_3} = \frac{4}{3} \pi \sqrt{\det(C)} \quad (3)$$

and can be computed from the semi-axes or directly from the covariance matrix.

The computation of the surface area  $O$  is more complicated and comprises incomplete elliptic integrals. Instead, we employ an approximation (Michon, 2004) for the surface area

$$O \approx 4\pi \left( \frac{a^p b^p + a^p c^p + b^p c^p}{3} \right)^{1/p} \quad (4)$$

where  $a, b, c$  are semi-axes as in (2) and  $p$  is a constant. The choice of  $p = 8/5 = 1.6$  is optimal for nearly spherical ellipsoids which leads to a maximum relative error of 1.178% (Michon, 2004).

A further metric is the sphericity  $S$  (Wadell, 1935) of the ellipsoid which measures how spherical the ellipsoid is. It is defined as the ratio of the surface area of a sphere with the same volume as the ellipsoid to the surface area of the ellipsoid:

$$S = \frac{\pi^{1/3} (6V)^{2/3}}{O} = \frac{4\pi \sqrt[3]{\lambda_1 \lambda_2 \lambda_3}}{O} \quad (5)$$

The last metric based on the error ellipsoid is an alternative roundness measure  $K$  similar to  $R$  and  $S$ . We define it as the quotient of the ellipsoid volume  $V$  and the volume  $V_K$  of the minimum circumscribed sphere:

$$K = \frac{V}{V_K} = \frac{\frac{4}{3} \pi abc}{\frac{4}{3} \pi r^3} = \frac{abc}{r^3} = \frac{abc}{\max(a, b, c)^3} = \frac{\sqrt{\lambda_2 \lambda_3}}{\lambda_1} \quad (6)$$

The radius  $r$  of the minimum circumscribed sphere is given by the largest semi-axis  $\max(a, b, c)$  of the error ellipsoid.

Additionally, we have defined the depth of the reconstructed 3D points  $D$  as a metric which is independent of the error ellipsoid. The depth is proportional to the baseline and our assumption is that for motion degeneracy it should differ for general scenes significantly from the depth with no degeneracy.

In summary we have defined the following metrics:

- Metrics based on the shape of error ellipsoid:
  - Roundness  $R$
  - Volume  $V$
  - Sphericity  $S$
  - alternative Roundness  $K$
- Depth  $D$  of the reconstructed 3D point

All metrics are computed for each reconstructed 3D point and the median is used as global metric.  $R_{mean}$  as proposed in (Beder and Steffen, 2006) is employed for comparison.

## 3 DETECTION OF CRITICAL CAMERA CONFIGURATIONS

In our experiments concerning the classification of critical camera configurations we use about 1500 image pairs as ground truth

data. The images were taken with handheld cameras and cameras mounted on small Unmanned Aerial System (UAS). The ground truth data consists of about 30% known degenerate pairs which were taken using handheld cameras.

For evaluating binary classifiers common measures are the receiver operating characteristic (ROC) and the corresponding area under the ROC curve as well as precision and recall. For imbalanced datasets, as in our case, precision and recall give a more informative picture of an algorithm's performance (Davis and Goadrich, 2006). Hence, we use them primarily for our experiments instead of ROC. Precision and recall are defined as

$$\text{precision} = \frac{TP}{TP + FP}$$

$$\text{recall} = \frac{TP}{TP + FN}$$

where  $TP$ ,  $FP$  and  $FN$  are the number of true positives, false positives and false negatives. In order to obtain only one evaluation measure, we use F-score which is the  $\beta$ -harmonic mean of precision and recall:

$$F_{\beta} = (1 + \beta^2) \cdot \frac{\text{precision} \cdot \text{recall}}{\beta^2 \cdot \text{precision} + \text{recall}}$$

The most common choice for  $\beta$  is 1, which leads to:

$$F = F_1 = 2 \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (7)$$

To obtain more statistically reliable results, all our evaluations\* were performed using stratified cross validation with 10 folds. I.e., the data was randomly partitioned into ten subsets of equal size, the classifier trained on nine subsets and validated on the remaining subset. The whole evaluation process was repeated 10 times and the mean is used as the final evaluation score.

In Section 3.1 we evaluate the metrics defined in Section 2 concerning their suitability as classification features. Then, in Section 3.2, we determine the best feature subset which leads to the optimal classification based on AdaBoost.

### 3.1 Feature Comparison

In Section 2 we have defined several metrics which are to be used as classification features. We compare their suitability as features using information gain, a measure which is often used in decision trees to find best splits. Information gain  $IG$  is given by

$$IG(\text{class}, \text{metric}) = H(\text{class}) - H(\text{class}|\text{metric})$$

where  $H$  denotes the (information) entropy. Features with higher information gain tend to be more suitable for class separation than features with lower information gain. The average information gain per feature is shown in Fig. 1. It can be seen, that the features  $V_{med}$  and  $D_{med}$  perform clearly best and also comprise relatively small variations between folds. The features  $R_{med}$ ,  $S_{med}$ ,  $R_{mean}$  and  $K_{med}$  behave more or less similarly. Yet,  $S_{med}$  seems to be slightly more stable between folds than the other and  $R_{mean}$  turns out to be inferior in comparison with  $R_{med}$ .

Based on the information gain we trained a simple decision stump, i.e., selected an appropriate threshold, and evaluated its performance and thus the suitability of a single feature for classification. The results are given in Tab. 1. Again, the best features turn out to be  $V_{med}$  and  $D_{med}$ , whereas the worst is still  $R_{mean}$ .

\*Evaluations were performed using the Weka software package ([www.cs.waikato.ac.nz/ml/weka](http://www.cs.waikato.ac.nz/ml/weka)).

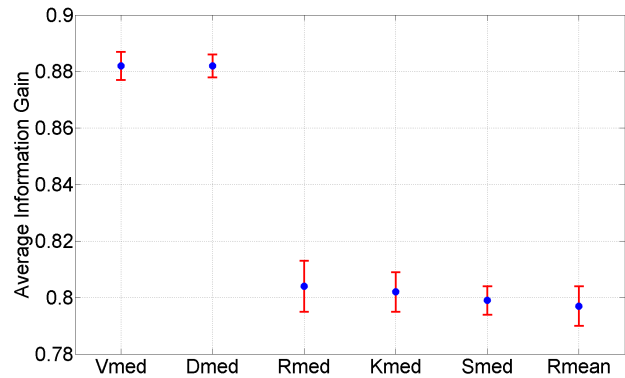


Figure 1: Average information gain per metric using stratified cross validation with 10 folds. The standard deviations between folds are represented as vertical red bars.

The features  $R_{med}$ ,  $S_{med}$  and  $K_{med}$  show no big difference and behave very similarly.

Feature	F-score	Area under ROC curve
$V_{med}$	0.9940	0.991
$D_{med}$	0.9915	0.988
$K_{med}$	0.9777	0.967
$R_{med}$	0.9792	0.969
$R_{mean}$	0.9738	0.957
$S_{med}$	0.9787	0.963

Table 1: Average feature performance based on classification using a single threshold and stratified cross validation with 10 folds

Next, we compared features  $R_{med}$ ,  $S_{med}$  and  $K_{med}$  which are roundness measures and have shown a similar behavior above. As presented in Fig. 2, the features are correlated pairwise, not linearly but rather quadratically, and there exists also a non-linear correlation between all three features. This proposition is confirmed by the correlation coefficients between feature pairs given in Tab. 2. The second column of Tab. 2 contains the values of the Pearson correlation coefficient which is used to detect a linear relationship. The Spearman's rank correlation coefficient in the third column can be used to detect a monotonic relationship, i.e., can be employed to detect also a non-linear relationship. Due to the correlation one can deduce, that it should be sufficient to use only one of the features for the classification. Because of its simplicity and modeling of the whole ellipsoid shape,  $K$  seems to be the appropriate choice. One could use  $S$  instead of  $K$ , but  $S$  is much more costly to compute and is also only an approximation of the ellipsoid surface area whereas  $K$  is an exact measure.

Features	Pearson	Spearman
$R_{med}, S_{med}$	0.95	0.99
$R_{med}, K_{med}$	0.84	0.96
$S_{med}, K_{med}$	0.77	0.95

Table 2: Correlation between roundness-based features

### 3.2 Feature Selection and Classification

From the results in Section 3.1 one can see that some metrics are more suitable for use as classification features than others. However, the results in Section 3.1 only hold if a single feature is used for classification. To find the best feature subset for a specific classifier, we performed an exhaustive search using F-score from equation (7) as evaluation measure and AdaBoost (Freund and Schapire, 1995) as the classification algorithm.

We use AdaBoost with 10, 30, 50, 80, 100, 150 and 200 trees (decision stumps, thresholds) and all non empty sets of the feature

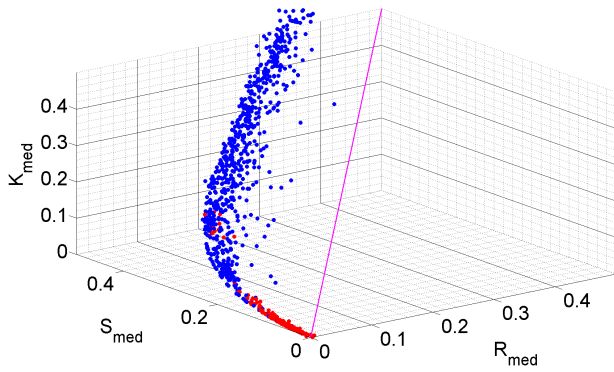


Figure 2: Correlation between roundness-based features. The feature values were normalized to  $[0; 1]$ . Blue points come from non-degenerate and red points from degenerate pairs. The magenta line is used to show the location of the linear relationship.

set's power set. As we found, that the performance of the feature subsets is not very sensitive to the number of trees, we use the mean over the tree scores for the evaluation. The average scores for all subsets are shown in Fig. 3 and a summary for the relevant subsets is given in Tab. 3. As can be seen in Fig. 3, the best feature subsets are located at the beginning of the right half (marked by a box). For these subsets F-Score as well as the area under ROC curve have the highest values. The combination of  $V_{med}$  and  $D_{med}$  clearly gives the best result. This result can be further improved by involving one of the roundness-based features, for which  $K_{med}$  seems to be slightly better than  $R_{med}$  and  $S_{med}$ . Note, that the good performance comes primarily from  $V_{med}$  and is only slightly improved by a combination with other features.

Finally, the classifier is trained on the whole data set and can then be employed for the detection of critical camera configurations for new image pairs.

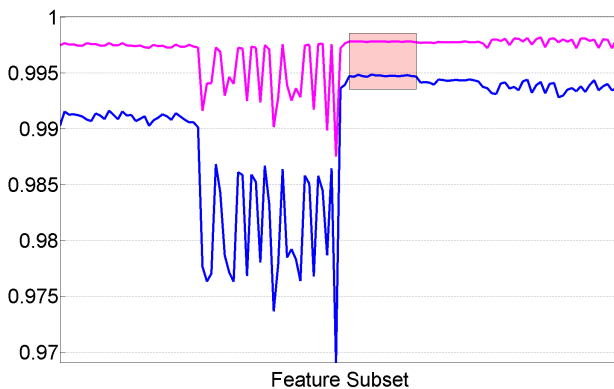


Figure 3: Average F-Scores (blue) and areas under ROC curve (magenta) for all feature subsets. The best range is highlighted by a box.

Feature Subset	F-score	Area under ROC curve
$V_{med}D_{med}K_{med}$	0.9947	0.9978
$V_{med}D_{med}R_{med}$	0.9944	0.9977
$V_{med}D_{med}S_{med}$	0.9943	0.9977
$V_{med}D_{med}$	0.9939	0.9976
$V_{med}$	0.9936	0.9972
$D_{med}$	0.9903	0.9974

Table 3: Evaluation results for feature subsets

#### 4 ORIENTATION OF UNORDERED IMAGE SETS

The detection of critical camera configurations is a prerequisite for a full automation of the orientation of unordered image sets. Only if critical camera configurations can be robustly detected, a reliable orientation avoiding scaling errors due to wrong lengths for corresponding baselines is possible. Thus, we plan to integrate the classifier described in the previous section in our approach for orientation described below in the near future.

Our approach for the orientation of possibly very large baseline image sets builds on (Bartelsen et al., 2012, Mayer et al., 2012). After detecting scale invariant feature transform – SIFT (Lowe, 2004) points, cross-correlation and affine least squares matching are used to obtain highly precise relative point positions as well as covariance information for them. This information is input to the random sample consensus – RANSAC (Fischler and Bolles, 1981), five point algorithm (Nistér, 2004) and robust bundle adjustment based determination of the relative orientation of image pairs and triplets.

In our previous work (Bartelsen et al., 2012, Mayer et al., 2012), the triplets are combined by means of tracking based on least squares matching with robust bundle adjustment necessary after the addition of few or even only one triplet. As this leads to a very high computational complexity, we have recently introduced a hierarchical procedure which employs unique identifiers for every point in every image. This is the basis for the determination of all points in all images, a 3D point can be seen in. The merging of image sets to ever larger sets can thus be computed in parallel and, therefore, efficiently. As in our preliminary work we use least squares matching and bundle adjustment to obtain a highly precise orientation, but the new procedure is considerably faster.

To deal with unordered image sets, we use the approach presented in (Bartelsen et al., 2012). The GPU implementation (Wu, 2007) of SIFT is employed to detect points and determine correspondences by pairwise matching. As result we obtain the matching graph which consists of images as nodes and edges connecting similar images. The weight of an edge is given by the number of correspondences between the connected images. Promising image pairs are obtained by construction of the maximum spanning tree of the matching graph. These image pairs are then used to derive and link image triplets. The latter are input for the orientation approach described above.

Our state concerning unordered image sets is still preliminary: Many image pairs, which can be oriented by our robust matching approach, are not found due to the limited capability of the employed fast matching method (Wu, 2007).

The result of our orientation framework for a set of 340 images is shown in Fig. 4. The images were acquired by handheld cameras from the ground and a camera mounted on a micro Unmanned Aerial System (UAS). A high quality dense 3D reconstruction of a part of the scene using the approach of (Kuhn et al., 2013) is given in Fig. 5.

#### 5 CONCLUSIONS AND FUTURE WORK

In this paper we have presented various error metrics and have analyzed their suitability using AdaBoost and cross validation as features for the classification of image pairs concerning critical camera configurations especially concerning motion degeneracy (no or very weak baseline). A combination of the volume, the distance and an alternative roundness measure for the ellipsoids

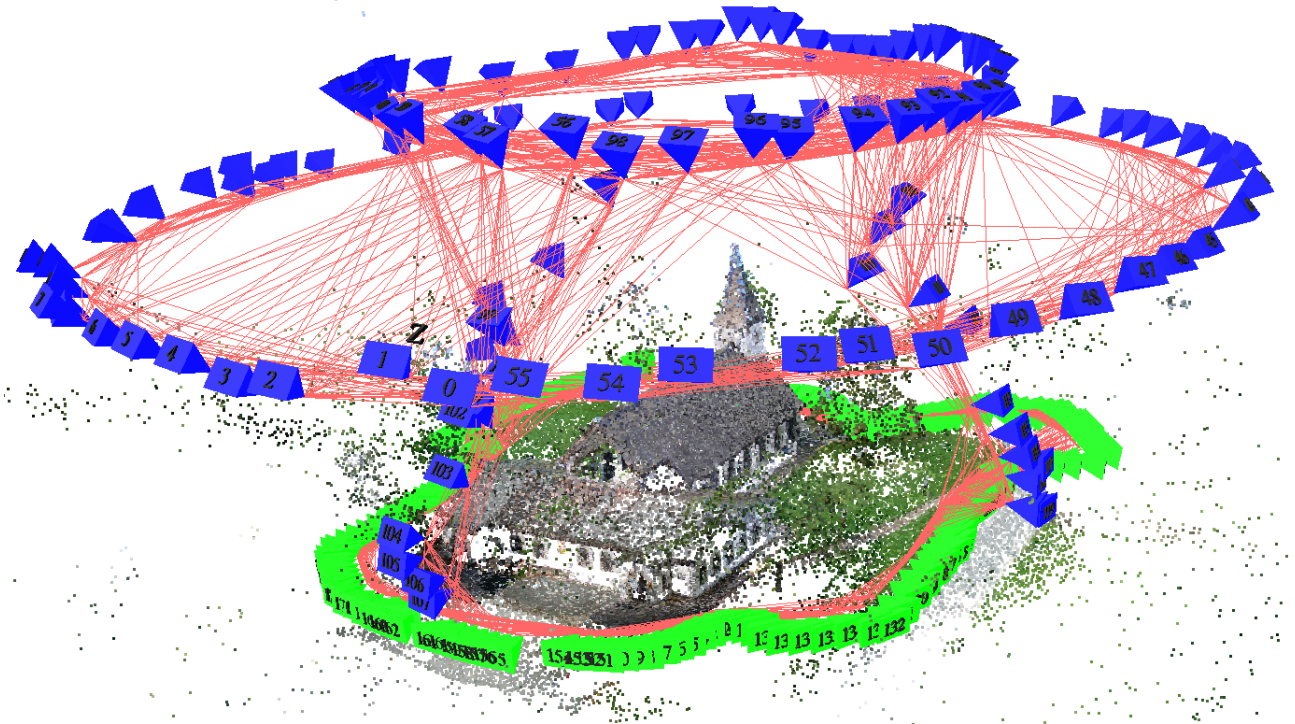


Figure 4: Orientation of 340 images taken from the ground and an Unmanned Aerial System (UAS) – pyramids represent cameras and links between cameras symbolize the existence of at least ten common points



Figure 5: Dense 3D reconstruction of a small part of the scene shown in Fig. 4

corresponding to the covariance matrices of the 3D points obtained by means of bundle adjustment were found to be particularly suitable, leading to a classification error of less than 1%.

In addition, we have sketched our approach for the orientation of unordered image sets which is able to produce very accurate and reliable orientations. This approach assumes that the image set does not contain image pairs arising from critical camera configurations. Thus, it will fail or yield a poor orientation if the image set contains such pairs. Therefore, we intend to integrate the above detection of pairs with a critical camera configuration into our orientation approach and use only non-degenerate pairs for the derivation of image triplets and image sets. By this means our framework should be able to produce very accurate and reli-

able orientations also for image sets comprising pairs with critical camera configurations.

In the future we intend to evaluate other classification algorithms and compare their performance with AdaBoost. Especially probability estimates instead of binary class membership provided, e.g., by Random Forests (Breiman, 2001), could be helpful. Also a classification into three classes, i.e., degenerate, non-degenerate and uncertain, could be useful. Uncertain image pairs could then be analyzed using more time-consuming techniques if they are found necessary for the connectivity of the image set or discarded otherwise.

## REFERENCES

- Bartelsen, J., Mayer, H., Hirschmüller, H., Kuhn, A. and Micheli, M., 2012. Orientation and Dense Reconstruction from Unordered Wide Baseline Image Sets. *Photogrammetrie – Fernerkundung – Geoinformation* 4/12, pp. 421–432.
- Beder, C. and Steffen, R., 2006. Determining an Initial Image Pair for Fixing the Scale of a 3D Reconstruction from an Image Sequence. In: *Pattern Recognition – DAGM 2006*, Springer-Verlag, Berlin, Germany, pp. 657–666.
- Breiman, L., 2001. Random Forests. *Machine Learning* 45(1), pp. 5–32.
- Chum, O., Werner, T. and Matas, J., 2005. Two-View Geometry Estimation Unaffected by a Dominant Plane. In: *Computer Vision and Pattern Recognition, 2005*, Vol. 1, IEEE Computer Society, Washington, DC, USA, pp. 772–779.
- Cortes, C. and Vapnik, V., 1995. Support-vector Networks. *Machine Learning* 20(3), pp. 273–297.
- Davis, J. and Goadrich, M., 2006. The Relationship Between Precision-Recall and ROC Curves. In: *23rd International Conference on Machine Learning*, ACM, New York, NY, USA, pp. 233–240.

- 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.
- Freund, Y. and Schapire, R., 1995. A Decision-theoretic Generalization of on-line Learning and an Application to Boosting. In: *Computational Learning Theory: Eurocolt '95*, Springer-Verlag, Berlin, Germany, pp. 23–37.
- Kuhn, A., Hirschmüller, H. and Mayer, H., 2013. Multi-Resolution Range Data Fusion for Multi-View Stereo Reconstruction. In: *Pattern Recognition, Lecture Notes in Computer Science*, Vol. 8142, Springer Berlin Heidelberg, pp. 41–50.
- Li, H. and Hartley, R., 2006. Five-Point Motion Estimation made Easy. In: *18th International Conference on Pattern Recognition*, Vol. 1, pp. 630–633.
- Lowe, D., 2004. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision* 60(2), pp. 91–110.
- 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.
- Mayer, H., Bartelsen, J., Hirschmüller, H. and Kuhn, A., 2012. Dense 3D Reconstruction from Wide Baseline Image Sets. In: *Real-World Scene Analysis 2011, Lecture Notes in Computer Science*, Springer-Verlag, Berlin, Germany, pp. 285–304.
- Michon, G. P., 2004. Final Answers. [www.numericana.com/answer/ellipsoid.htm#thomsen](http://www.numericana.com/answer/ellipsoid.htm#thomsen).
- 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., Gool, L. V., Vergauwen, M., Cornelis, K., Verbiest, F. and Tops, J., 2002a. Video-to-3D. In: *Photogrammetric Computer Vision 2002*, pp. 252–258.
- Pollefeys, M., Verbiest, F. and Gool, L. J. V., 2002b. Surviving Dominant Planes in Uncalibrated Structure and Motion Recovery. In: *7th European Conference on Computer Vision – Part II*, Springer-Verlag, pp. 837–851.
- Repko, J. and Pollefeys, M., 2005. 3D Models from Extended Uncalibrated Video Sequences: Addressing Key-Frame Selection and Projective Drift. In: *5th International Conference on 3-D Digital Imaging and Modeling*, IEEE Computer Society, pp. 150–157.
- Thormählen, T., Broszio, H. and Weissenfeld, A., 2004. Keyframe Selection for Camera Motion and Structure Estimation from Multiple Views. In: *8th European Conference on Computer Vision, Lecture Notes in Computer Science*, Vol. 3021, Springer, pp. 523–535.
- Torr, P. H. S., 1998. Geometric Motion Segmentation and Model Selection. *Philosophical Transactions of the Royal Society of London A* 356, pp. 1321–1340.
- Torr, P. H. S., Fitzgibbon, A. W. and Zisserman, A., 1999. The Problem of Degeneracy in Structure and Motion Recovery from Uncalibrated Image Sequences. *International Journal of Computer Vision* 32(1), pp. 27–44.
- Wadell, H., 1935. Volume, Shape and Roundness of Quartz Particles. In: *Journal of Geology*, pp. 250–280.
- Wu, C., 2007. SiftGPU: A GPU Implementation of Scale Invariant Feature Transform (SIFT). [www.cs.unc.edu/~ccwu/siftgpu](http://www.cs.unc.edu/~ccwu/siftgpu).