

# NON-INCREMENTAL DERIVATION OF SCALE AND POSE FROM A NETWORK OF RELATIVE ORIENTATIONS

Alessandro Cefalu, Dieter Fritsch

Institute for Photogrammetry, University of Stuttgart, Germany  
 alessandro.cefalu@ifp.uni-stuttgart.de  
 dieter.fritsch@ifp.uni-stuttgart.de

## Commission III

**KEY WORDS:** non-incremental Structure from Motion, Image Orientation, Pose Estimation, Ant Colony Optimization

### ABSTRACT:

The majority of approaches to Structure from Motion apply an incremental triangulate-and-resect strategy in order to reconstruct camera motion and scene structure in a common reference frame. The sequential addition of images may cause a drifting behaviour during the reconstruction, in some cases causing the process to fail. Over the last decade, more attention has come to non-incremental approaches, which exploit the network characteristics arising from the 2- or 3-view relations, given for a set of images through relative orientations. Most approaches employ rotation registration, followed by translation registration. The latter being carried out with or without simultaneous scene reconstruction. We suggest an approach which starts by estimation of relative scales, followed by simultaneous registration of rotation and translation. The latter is achieved by employing a path-finding algorithm based on Ant-Colony-Optimization. For scale estimation we propose a window-search adaption of Levenberg-Marquardt, which avoids unnecessary matrix inversions. We also suggest a simple method for detection of outlier orientations.

### 1. INTRODUCTION

Structure from Motion (SfM) is a strongly researched field and also well-established among users. Its application ranges from airborne survey to architecture and archaeology, just to name a few. Roughly described, the procedure consists of (i) establishment of geometric relations between image pairs or triplets by matching of visual features, (ii) fusion of camera poses and scene structure in a common coordinate frame (although i and ii not necessarily need to be executed in this order) and (iii) refinement of the result by bundle adjustment. The majority of approaches to SfM employ an incremental triangulate-and-resect strategy for the second task. Starting with a minimal solution from two or three images, further images are added successively, using the currently available 3D structure. This leads to four main drawbacks. (i) The choice of images for the initial solution and the order in which the remaining images are added, have strong impact on the quality of the results. (ii) Short baselines may generate erroneous point triangulations which cause incorrect results in subsequent steps. (iii) Errors are accumulated and lead to a drifting behaviour which may result in the well-known loop closure problem. (iv) Ambiguous structures, e.g. repetitive facades or vegetation, may cause strong false-positive 2-view-connections between images, often resulting in a failing reconstruction. Of course, various approaches within the incremental framework address these problems, but are not listed here.

Non-incremental (or non-sequential) approaches to SfM have come to more attention within the last decade. Generally speaking, these methods address the task of fusing camera poses and structure in a common coordinate frame by exploiting the network characteristics arising from the relative orientations between the images. Cycles within the network (or *camera graph*) correspond to a cyclic

concatenation of transformations and thus need to yield identity (Figure 1, Equation 1). In the following, we will refer to this circumstance as *identity constraint*. Since the scale of the baselines is unknown, the constraint only holds for the rotational part of the transformations. Within this context, a common coordinate frame can be interpreted as an additional node in the network, which is connected to all other nodes. Initially, the corresponding edges, i.e. the absolute orientations, are unknown. Yet, the identity constraint allows for expressing the absolute rotations as functions of the relative ones and can be solved with comparably little effort. This step is often entitled as *rotation averaging* or *rotation registration* and (although it can be solved in different ways) is the common starting point of all non-incremental SfM approaches we are aware of.

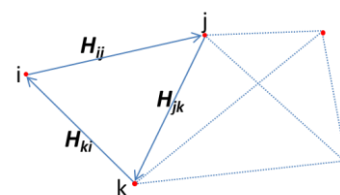


Figure 1. Sketch of a camera graph. Chaining the relative transformations  $H$  needs to yield identity.

The process is followed by a reconstruction of the global translations (camera positions). Here we encounter a larger diversity in the general strategies. In contrast to other approaches we begin our process by estimation of the unknown scales of the baselines and proceed by simultaneously solving for global rotation and translation, using the identity constraint. For this step we employ a robust, nature-inspired sampling of paths from the network of scaled relative orientations.

In the following chapter we try to classify existing approaches according to the underlying constraints and concepts for achieving robustness. Depending on the choice of constraint, scene reconstruction may need to be computed as an additional step.

Note: In this paper, when speaking of absolute or global orientation, we neither mean a georeferenced solution, nor metric scaling. Instead we refer to an arbitrary reference frame with arbitrary scale. That is, compared to a georeferenced solution, a datum defect of rank seven remains.

## 1.1 Related work

**Global rotation estimation:** A linear least squares solution based on quaternion formulation of rotations is given in (Govindu, 2001). (Martinec & Pajdla, 2007) suggest a solution based on approximate rotations and subsequent enforcement of orthonormality for the rotation matrices. A gradient descent method based on matrix completion is presented in (Arrigoni et al., 2014). (Hartley et al., 2013) discusses rotation averaging in different application scenarios and investigates different distance measures on  $SO(3)$ , the space of three-dimensional rotations.

**Global translation and structure estimation based on collinearity:** A reformulation of the collinearity constraint is used in (Kahl, 2005) to simultaneously estimate camera poses and scene structure. The concept is picked up by e.g. (Olsson & Enqvist, 2011) and (Martinec & Pajdla, 2007). The latter suggest a reduction of the number of reconstructed points by selecting four representative points per image pair.

**Global translation estimation based on coplanarity:** In (Arie-Nachimson et al., 2012) the relative orientation between cameras within the epipolar constraint is formulated as concatenation of the corresponding absolute orientations. The resulting system contains the observations of the scene points, but no object space coordinates; hence, the number of unknowns is reduced drastically. (Arrigoni et al., 2014) adopt this method for their reconstruction pipeline.

**Global translation estimation based on identity:** These approaches solely rely on the transformations between camera frames, thus creating very small systems. (Govindu, 2001) eliminates the unknown scale of the heading vectors by applying the cross-product with the correspondingly chained absolute camera positions and solves the system using a weighted least squares method. (Sim & Hartley, 2006) achieve scale elimination by setting the heading vector equal to the normalized version of the chained global camera positions and solve for unknowns employing  $L_\infty$  minimization. (Moulon et al., 2013) keeps the heading vector scales as unknowns in order to ensure chirality (right-handedness, points must be triangulated in front of a camera) while creating a linear system.

**Robustness to relative translation based on trifocal tensor:** As mentioned above, short baselines between images may cause erroneous triangulation results in incremental SfM frameworks. This is caused by a strong sensitivity of the relative translation estimation to short base lines (which is not the case for relative rotations). It is frequently suggested (also for non-incremental approaches) to compensate for this by employing image triplets and trifocal tensor estimation, e.g. (Sim & Hartley, 2006) or (Moulon et al., 2013).

The identity constraint to cycles of transformations not only gives rise to methods of point-free estimation of camera poses. It also allows detection of outliers among the relative orientations by evaluating the deviation from identity. The concept is often referred to as *cycle consistency*.

**Robustness to global rotation based on spanning trees:** A spanning tree of relative orientations, derived from the camera graph, can be used to compute putative absolute orientations, by chaining the relative orientations accordingly. When edges, which have not been used in the tree, are added, they create cycles which can be evaluated in terms of consistency. Since the number of orientations used in the tree is constant, also the number of unused edges is independent of the considered tree. (Govindu, 2006) suggests a RANSAC (Fischler & Bolles, 1981) approach, considering unused edges, whose consistency falls within a threshold, as inliers. The spanning tree yielding the highest number of inliers is kept for further computations. (Olsson & Enqvist, 2011) extend the approach by guided sampling of the trees, i.e. the selection probability of the edges is weighted according to the number of matches between the corresponding images. In (Enqvist et al., 2011) repeatedly building maximum spanning trees over the number of correspondences and rejection of inconsistent edges is suggested. The consistency measure is normalised by the square root of the cycle length, to take error accumulation into account.

**Robustness to relative rotations based on inference:** (Zach et al., 2010) utilizes Bayesian inference to the complete graph of relative orientations, over cycles of maximum length six, to detect erroneous relative rotations in the network and remove corresponding edges.

## 1.2 Outline of our approach

We pick up the idea of pure motion estimation from the identity constraint. Our goal is to implement a non-incremental SfM pipeline which avoids reconstruction of scene points as far as possible, to keep the number of unknowns small, for as long as possible. This includes global epipolar estimation as a replacement for bundle adjustment. In other words, we aim at fully decoupling motion estimation from scene reconstruction. As outcome of such an approach, we expect superior scalability to large data sets, but also see possibilities for application in SLAM or pure visual navigation, where computation power may be restricted or reconstruction is not, or only partially, needed. For instance, scene reconstruction could be run in a separate thread and turned off, when an area is traversed, which has already been visited. Presently we focus on the non-real-time scenario.

Starting with a set of relative orientations between images the outline of our suggested approach is:

1. Remove outlier orientations by rotational cycle consistency
2. Estimate the scale of the baselines
3. Remove outlier orientations by translational cycle consistency
4. Compute approximate absolute orientations using a robust, adaptively guided tree sampling procedure

5. A: Refine absolute orientations by global epipolar adjustment, then triangulate scene structure  
 B: triangulate scene structure, then refine using bundle adjustment

Since this is work in progress, this paper covers only steps 1 to 4, which represent the process of finding starting values for the bundle adjustment, or, as we plan for our future work, global epipolar adjustment. We suggest a simple accumulative procedure for detection of outlier orientations based on rotational cycle consistency. It is employed before scaling the heading vectors using least squares in a window search approach, based on the Levenberg-Marquardt framework. The previous outlier-detection method is then re-applied to the scaled translations. We conclude the process by retrieving approximate absolute orientations using a nature inspired tree sampling procedure, which adopts sampling probabilities to the encountered errors.

## 2. OUTLIER DETECTION

Trilaterally connected images form the shortest cycles in a camera graph. The identity constraint for a cycle formed by images  $i, j, k$  is given by:

$$\begin{aligned} I_{4 \times 4} &= H_{ki} H_{jk} H_{ij} \\ &= \begin{bmatrix} R_{ki} & s_{ki} t_{ki} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R_{jk} & s_{jk} t_{jk} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R_{ij} & s_{ij} t_{ij} \\ 0 & 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} R_{ki} R_{jk} R_{ij} & R_{ki}(R_{jk}(s_{ij} t_{ij}) + s_{jk} t_{jk}) + s_{ki} t_{ki} \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{aligned} \quad (1)$$

Or equivalently:

$$0_{4 \times 4} = H_{jk} H_{ij} - H_{ki}^{-1}$$

The indices are given as ‘from-to’ transformation direction.  $s$  is the unknown scale of a baseline. As mentioned, the rotational part is independent of  $s$  and can therefore be evaluated without scale information.

For each triplet of images we compute the geodesic angular distance  $\theta_{ijk}$  between identity and the chained rotations, derived from the Frobenius norm (*chordal distance*, Hartley et al., 2014) and weighted by the square root of the cycle length as suggested in (Enqvist et al., 2011).

$$d_{ijk}^{chord} = \|I_{3 \times 3} - R_{ki} R_{jk} R_{ij}\|_F \quad (2)$$

$$\theta_{ijk} = \frac{2}{\sqrt{3}} \arcsin\left(\frac{d_{ijk}^{chord}}{2\sqrt{2}}\right) \quad (3)$$

For each edge  $ij$  we compute the mean error of all  $n$  triangles sharing this edge.

$$\theta_{ij} = \frac{1}{n} \sum_k \theta_{ijk} \quad (4)$$

We consider edges with  $\theta_{ij} > 2^\circ$  as outliers. If there are outliers, we iteratively remove the edge with the largest error and re-compute the errors. Multiple computations are easily avoided by using a triangle/edge incidence matrix containing the errors  $\theta_{ijk}$  and setting corresponding entries to 0, when an edge is removed.

If scale (Chapter 3) is available for the relative camera translations, the same procedure can be used to detect translational outliers. Since no direct metric is available for the scales, we use an angular displacement  $\alpha$  between a ‘measured’ scaled heading vector  $t_{ij}^m$  and the heading vector  $t_{ij}^c$  resulting from ‘chaining’ the two remaining (upper index  $m$  or  $c$ , respectively).

$$\begin{aligned} t_{ij}^m &= s_{ij} t_{ij} ; t_{ij}^c = s_{jk} R_{ki} t_{jk} + s_{ki} t_{ki} \\ \alpha_{ij} &= \angle(t_{ij}^m, t_{ij}^c) \end{aligned} \quad (5)$$

As in contrast to the rotational case, permutation of the identity constraint leads to different results. We use the mean of the three possible results (again, weighted by cycle length) as error measure  $\gamma_{ijk}$  for a triangle.

$$\gamma_{ijk} = \frac{\alpha_{ij} + \alpha_{jk} + \alpha_{ki}}{3\sqrt{3}} \quad (6)$$

The remaining steps equal the rotational case, except, we allow for larger deviation; edges with  $\gamma_{ij} > 4^\circ$  are considered being outliers. When no more outliers are found, we re-estimate the scale of the remaining baselines.

## 3. SCALE ESTIMATION

The unknown scales  $s$  in the translational part of (Equation 1) can easily be solved, for example using singular value decomposition. Yet, in order to ensure chirality, it should be carried out including inequality constraints, such that  $s > 0$ . We implemented such an approach to compare it to our suggested approach, which is a window search adaption of the Levenberg-Marquardt (LM) framework. LM (Equation 7, often referred to as *damped least squares*) is a simple modification of iterative least squares and is known to be less sensitive to the quality of starting points  $x_0$  and to outliers. It introduces a modification to the normal equation matrix  $N$ , by superimposing a matrix  $N_d$ , which contains only the diagonal elements of  $N$ . Additionally,  $N_d$  is scaled by a factor  $\lambda$ , i.e. the diagonal elements of the normal equation matrix are emphasized for large  $\lambda$ .

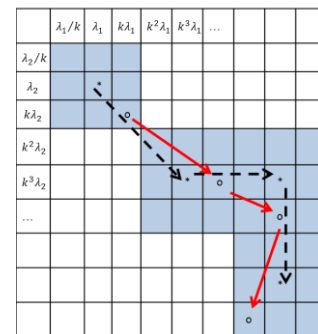


Figure 2. Sketch of our Levenberg-Marquardt-like window search approach. Stars depict the centres of the search windows, circles show the best solution found within the window.

For a small  $\lambda$  the solution  $dx$  is very close to the least squares solution. By increasing  $\lambda$ , the solution direction changes from least squares to steepest gradient and is simultaneously damped. The iterations are usually started with a damped

solution. In each iteration, the new solution is checked, whether it improves the residuals. If this is the case the result is accepted, the starting point  $x_0$  is updated and  $\lambda$  is decreased by dividing by a factor  $k$  (step size). Otherwise, the result is dismissed,  $\lambda$  is increased by multiplying by  $k$  and a new result  $dx$  is computed from the same starting point.

$$\begin{aligned} x &= x_0 + dx \\ dx &= (N + \lambda N_d)^{-1} A' l \end{aligned} \quad (7)$$

A drawback of LM is the necessity to repeatedly invert the modified normal equation matrix. We pick up the general idea of damping and direction modification, but try to avoid unnecessary inversions through a window search approach. In each iteration we compute two separate solutions  $dx_{ls}$  and  $dx_{sg}$  and combine them to a solution  $dx$  by:

$$\begin{aligned} dx_{ls} &= N^{-1} A' l \\ dx_{sg} &= N_d^{-1} A' l \\ dx &= \left( dx_{ls} + \lambda_1 \frac{dx_{sg} \|dx_{ls}\|_2}{\|dx_{sg}\|_2} \right) \frac{1}{\lambda_2} \end{aligned} \quad (8)$$

Here,  $\lambda_1$  induces the direction change, whereas  $\lambda_2$  damps the solution. Thus, we are able to modify damping and direction independently. Within an inner loop, we evaluate all combinations for  $\lambda/k$ ,  $\lambda$  and  $\lambda k$ , a total of nine solutions. Depending on which combination yields the best improvement of the residuals, we individually multiply or divide the damping factors by  $k^2$  (squared step size to avoid overlapping windows) or leave them untouched and proceed with a subsequent inner iteration (Figure 2). We repeat the inner loop until no improvement is achieved. Only after termination of the inner loop, we update  $x_0$  and compute new solution vectors by (Equation 8).

Using the described procedure, we solve for scales using a non-linear formulation of the translational identity.

$$t_{ij} = \frac{1}{s_{ij}} (s_{jk} R_{ki} t_{jk} + s_{ki} t_{ki}) \quad (9)$$

Again, permutation is taken into account, yielding nine equations per triangle. We also tested a linear formulation in our framework, but observe less sensitivity to outlier orientations, when using (Equation 9). This also holds in comparison to our implementation of an inequality-constrained svd solution (Figure 3). Although the current implementation does not take chirality into account, we have not encountered any problems so far. Yet, we are planning to add inequality constraints to our approach, to be sure in this point.

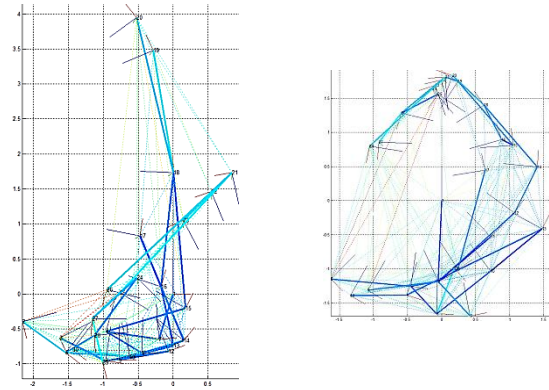


Figure 3. Absolute orientation results of our overall approach with outlier detection turned off (dataset CastleP30). Left: scales estimated with svd and inequalities. Right: scales estimated using our suggested approach. Although still erroneous due to no outlier removal being performed, it is much closer to the real solution. Compare this with Figure 7, right column.

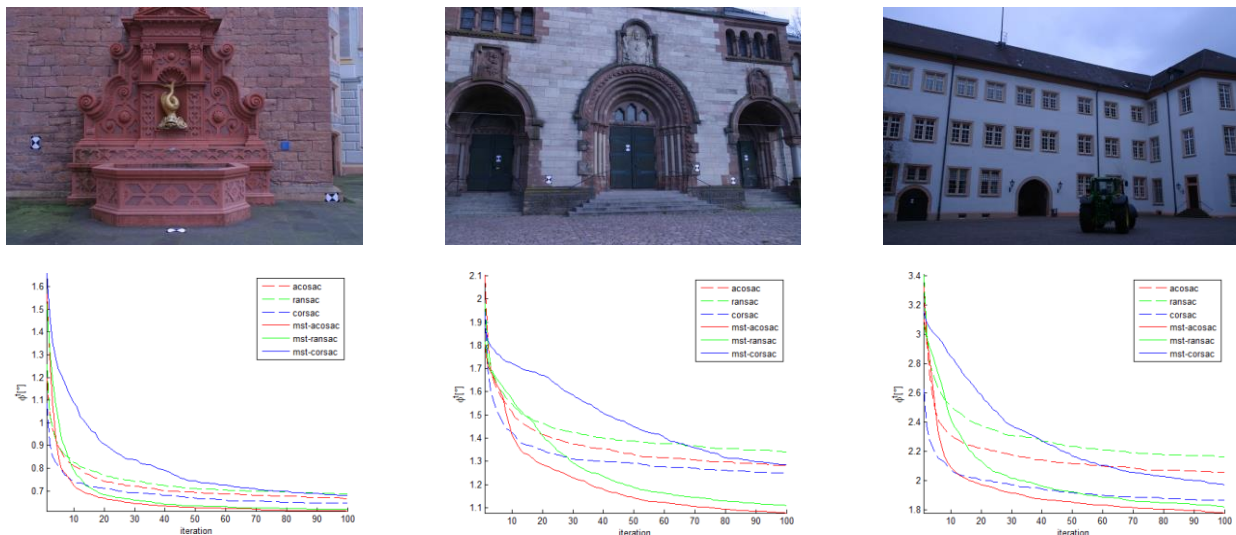


Figure 4. Mean convergence behaviour (median error in ° over iterations, 100 runs over 100 iterations) of different tree sampling methods. The dashed lines show the best directly picked result. The solid lines show the best result for minimal spanning trees derived from dynamic weighting. After a few iterations our suggested method (solid red) finds superior solutions. Datasets from left to right: FountainP11, HerzJesuP25 and CastleP30, all being part of the benchmark dataset of (Strecha et al., 2008). The images are undistorted. The numbers in the names of the dataset indicate the number of images.



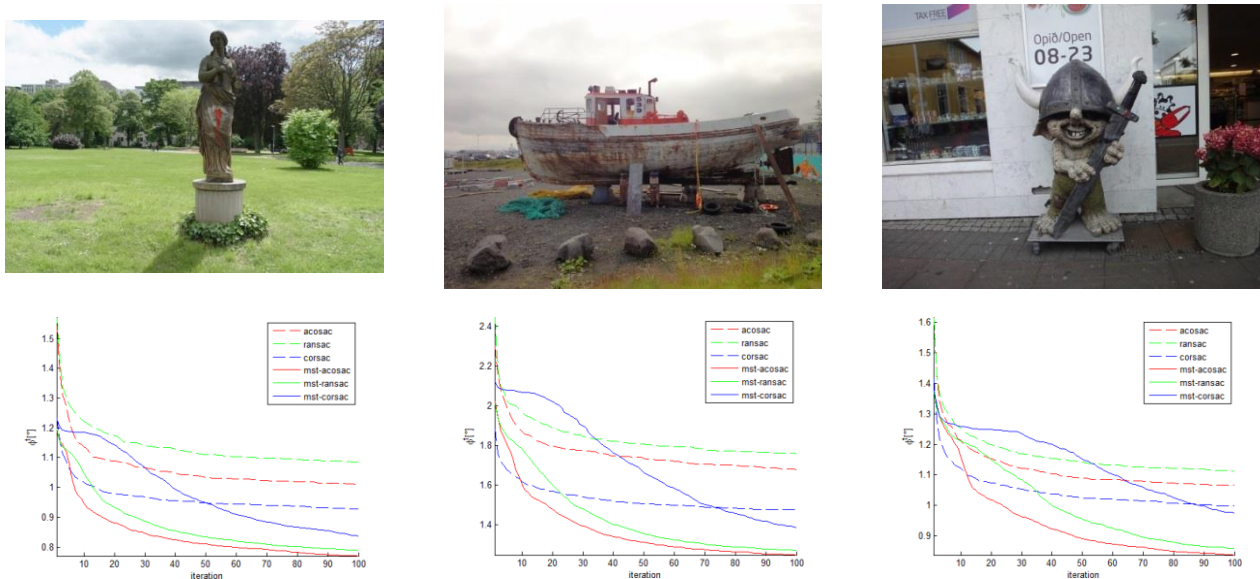


Figure 5. Mean convergence behaviour (100 runs over 100 iterations) on datasets with uncalibrated cameras. Left to Right: Statue (two circles, 55 images), Boat (one circle, 56 images), VikingTroll (three arches, 49 images)

#### 4. APPROXIMATE ABSOLUTE ORIENTATIONS

Having removed outliers among the relative orientations and estimated the scales of the baselines, we find approximate absolute orientations by tree sampling. We employ a path-finding algorithm which might be described as a dynamically reweighted RANSAC approach and is closely related to Ant-Colony-Optimization (ACO, Dorigo et al. 1999).

Initially, ants search for food in a random manner, leaving a trail of pheromones, which they can follow on their way back to the formicary. If the search was successful, they lay another layer of pheromones. The pheromone trail influences other ants in their choice of path. The more ants that follow a trail, the stronger the attraction of the path to a source of food. When the source is consumed, the pheromones evaporate and the search behaviour falls back to random. ACO simulates this procedure by picking a number of random paths from a graph and evaluating them with respect to some quality measure. The results are induced to the graph as weighting of the edges of the path. Basic variants evaluate picked trees as a whole and reweight corresponding edges of the graph by the same value; other variants evaluate each tree segment individually. Evaporation is often simulated by decreasing ‘old’ weights, in order to avoid premature convergence. In subsequent iterations, the same number of paths is picked, while adjusting the probability of choosing an edge to the current weightings. Over time, the minimal (or maximal) spanning tree of the graph converges.

##### 4.1 Outline of our approach

In our approach we only use one ‘ant’, which means, we update the weights of the graph after each sampled tree. Furthermore, we do not evaporate weights explicitly, but build mean weights over the iterations. Edges of a sampled tree are evaluated individually, as described in the following.

We start by initializing the weights of the camera graph with a positive value close to zero, to avoid division by zero in the following steps. We pick a random tree, using the reciprocal

values of the weights as probability measures. Each edge, not being part of the tree, forms a cycle, when added to the tree. A cycle is evaluated according to (4) and (6), the only difference being, that we allow arbitrary cycle length and reweight accordingly. We take the mean of both values as an error measure  $\phi$ . The errors for all cycles are computed and accumulated over the corresponding segments of the tree. Simultaneously, we count ‘traverses’ of the tree edges, i.e. the number of cycles a tree edge is part of (Figure 6). The accumulated errors at the edges are then normalized by the according number of traverses.

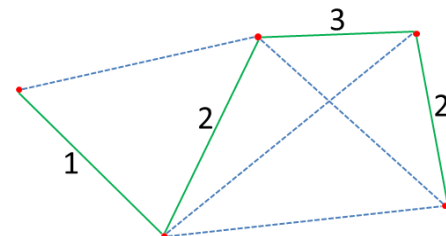


Figure 6. Sketch of a simple graph. The nodes are connected over a spanning tree (green). Unused edges (dashed blue) form cycles. The numbers indicate how many cycles a tree edge is part of.

Being  $\phi_i^e$  the error of a cycle  $i$  containing the tree edge  $e$ ; and  $n$  being the number of all cycles containing  $e$ , the total error measure  $\phi^e$  for the tree edge  $e$  is:

$$\phi^e = \frac{1}{n} \sum_{i=1}^n \phi_i^e \quad (10)$$

We evaluate all edges of the tree in this manner and add the results to the weighting of the overall graph. We repeat the tree sampling process iteratively, while counting how often an edge of the graph is chosen as a tree segment. The accumulated weights in the graph are normalized

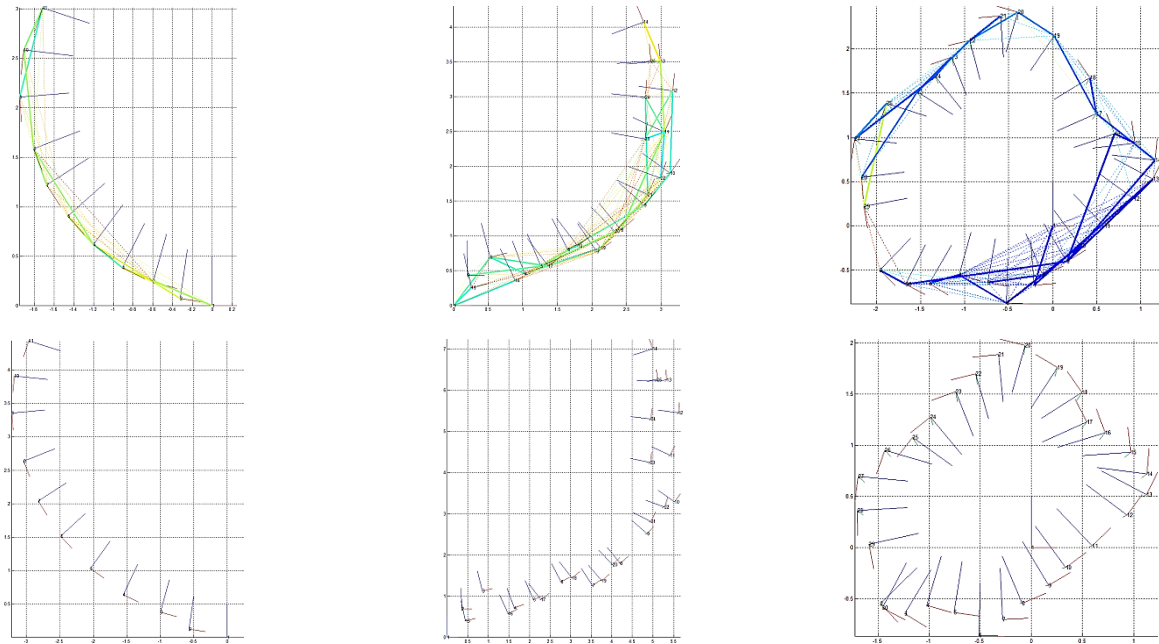


Figure 7. Top-view plots of the camera positions for the datasets of Figure 4 (same order), First image used as reference frame. Upper Row: Results derived after 20 iterations of our approach. The cameras are connected by edges corresponding to the established 2-view-connections. The colours indicate the weights created over the iterations. Thick lines represent the minimal spanning tree. Lower row: Reference solution computed using a commercial SfM software.

accordingly. The procedure results in the probability of choosing edges causing large errors being lower than that of choosing edges causing small errors.

To evaluate the quality  $\phi^t$  of a sampled tree  $t$ , we choose the median of all  $\phi^e$  of the tree. At this point, it might be considered to choose the best sampled path as solution, which would resemble a pure RANSAC approach. Yet, the goal of our procedure is to find a non-random path with superior quality, i.e. the minimal spanning tree over the weighted graph should converge to a solution better than the best randomly picked path, or at least find the solution earlier.

## 4.2 Experiments

In our experiments we let the process run for 100 iterations, and repeat 100 times for a dataset. The plots (Figure 4 and 5) show the mean convergence behaviour. We evaluate three different sampling variants:

- pure random sampling without influences on choice probabilities (unweighted, ‘ransac’), close to (Govindu, 2006)
- random sampling with probabilities according to the number of correspondences (fixed weights, ‘corrsac’), close to (Olsson & Enqvist, 2011)
- random sampling using our suggested approach (dynamic weights, ‘acosac’)

Additionally, we use all three variants to create weights for the edges as described and evaluate the convergence of the corresponding minimal spanning tree. The experiments show, fixed weighting according to the number of correspondences usually performs best among the tested random sampling approaches, but results in the worst minimal spanning tree result. Unweighted random sampling performs worst, but generates considerably good minimal spanning tree solutions.

This may be explained by a better coverage over the space of possible solutions, resulting in a better estimation of errors per edge. Although random sampling based on the ACO-weights only produces medium sampling results, it produces the best minimal spanning tree solution. The latter being the best performing of the tested variants, after a few iterations. Illustrations of absolute camera orientations for the test data sets, derived after only 20 iterations, can be found in Figure 7 and 8.

## 5. CONCLUSIONS

We presented a procedure to derive absolute orientations from a graph of relative orientations in a non-incremental way. A simple method, to detect outlier orientations based on cycle consistency, was suggested, which can be used for rotations and translations (in presence of scale information). To estimate scales we suggested a hybrid least squares / steepest gradient window search approach, which imitates the behaviour of Levenberg-Marquardt while decreasing the number of matrix inversions and being considerably insensitive to outliers. In order to derive approximate absolute orientations based on tree sampling, we propose a path finding algorithm based on Ant-Colony-Optimization and show in our experiments, that it outperforms other sampling methods after a few iterations.

We achieve good results for most datasets we have tested. However, our approach, which avoids rotation registration, may be over-optimistic in cases of weakly connected graphs with strong drifting behaviour, as can be seen in Figure 8 (middle column). Thus, we consider performing rotation registration before scale estimation, to make our process more reliable. In combination with the optimization schemes presented, we expect being able to present a fully working, scalable and robust pipeline for non-incremental Structure from Motion in near the future.

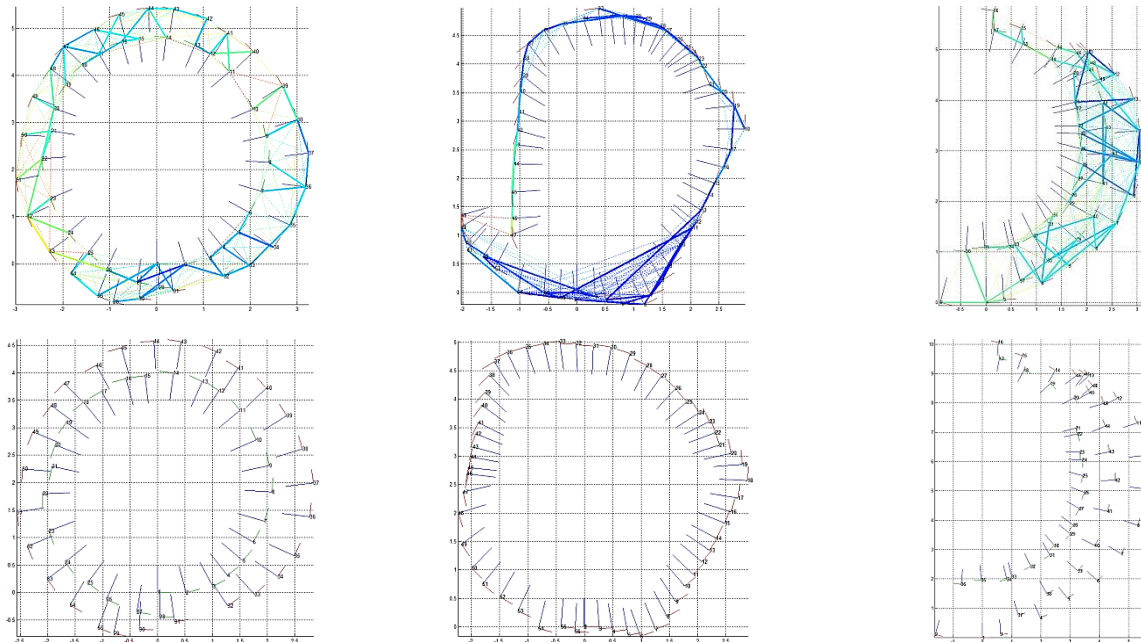


Figure 8. Top-view plots of the camera positions for the datasets of Figure 4.

## 6. REFERENCES

- Arie-Nachimson, M., Kovalsky, S. Z., Kemelmacher-Shlizerman, I., Singer, A., & Basri, R., 2012. Global motion estimation from point matches. In *3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT)*, Second International Conference on (pp. 81-88). IEEE.
- Arrigonis, F., Rossi, B., Malapelle, F., Fragneto, P., & Fusiello, A., 2014. Robust global motion estimation with matrix completion. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 1, 63-70.
- Dorigo, M., Caro, G., & Gambardella, L., 1999. Ant algorithms for discrete optimization. *Artificial life*, 5(2), 137-172.
- Enqvist, O., Kahl, F., & Olsson, C., 2011. Non-sequential structure from motion. In *Computer Vision Workshops (ICCV Workshops)*, IEEE International Conference on (pp. 264-271). IEEE.
- Fischler, M. A., & Bolles, R. C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6), 381-395.
- Govindu, V. M., 2001. Combining two-view constraints for motion estimation. In *Computer Vision and Pattern Recognition*, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on (Vol. 2, pp. II-218). IEEE.
- Govindu, V. M., 2006. Robustness in motion averaging. In *Computer Vision-ACCV 2006* (pp. 457-466). Springer Berlin Heidelberg.
- Hartley, R., Trampf, J., Dai, Y., & Li, H., 2013. Rotation averaging. *International journal of computer vision*, 103(3), p267-305.
- Kahl, F., 2005. Multiple view geometry and the  $L_\infty$ -norm. In *Computer Vision*, 2005. ICCV 2005. Tenth IEEE International Conference on (Vol. 2, pp. 1002-1009). IEEE.
- Martinec, D., & Pajdla, T., 2007. Robust rotation and translation estimation in multiview reconstruction. In *Computer Vision and Pattern Recognition (CVPR)*. IEEE Conference on (pp. 1-8). IEEE.
- Moulon, P., Monasse, P., & Marlet, R., 2013. Global Fusion of Relative Motions for Robust, Accurate and Scalable Structure from Motion. In *Computer Vision (ICCV)*, IEEE International Conference on (pp. 3248-3255). IEEE.
- Olsson, C., Enqvist, O., 2011. Stable structure from motion for unordered image collections. In *Image Analysis* (pp. 524-535). Springer Berlin Heidelberg.
- Sim, K., & Hartley, R., 2006.. Recovering camera motion using linfty minimization. In *Computer Vision and Pattern Recognition*, IEEE Computer Society Conference on (Vol. 1, pp. 1230-1237). IEEE.
- Strecha, C., von Hansen, W., Van Gool, L., Fua, P., & Thoennessen, U., 2008. On benchmarking camera calibration and multi-view stereo for high resolution imagery. In *Computer Vision and Pattern Recognition*, 2008. CVPR 2008. IEEE Conference on (pp. 1-8). IEEE.
- Zach, C., Klopschitz, M., & Pollefeys, M., 2010. Disambiguating visual relations using loop constraints. In *Computer Vision and Pattern Recognition (CVPR)*, IEEE Conference on (pp. 1426-1433). IEEE.