A New Minimal Solution to the Relative Pose of a Calibrated Stereo Camera with Small Field of View Overlap

Brian Clipp[1], Christopher Zach[1], Jan-Michael Frahm[1] and Marc Pollefeys[2]

[1]Department of Computer Science
The University of North Carolina at Chapel Hill, NC, USA
{bclipp,cmzach,jmf}@cs.unc.edu

[2]Department of Computer Science
ETH Zurich, Switzerland
marc.pollefeys@inf.ethz.ch

Abstract

In this paper we present a new minimal solver for the relative pose of a calibrated stereo camera. It is based on the observation that a feature visible in all cameras constrains the relative pose of the second stereo camera to be on a sphere around the feature which has a known position relative to the first stereo camera pose due to its triangulation. The constraint leaves three degrees of freedom, two for the location of the second camera on the sphere and the third for rotation in the plane tangent to the sphere. We use three temporal 2D correspondences, two correspondences from the left (or right) camera and one correspondence from the other camera to solve for these three remaining degrees of freedom. This approach is amenable to stereo pairs having a small overlap in their views. We present an efficient solution of this novel relative pose problem, theoretically derive how to use our new solver with two classes of measurements in RANSAC, evaluate its performance given noise and outliers and demonstrate its use in a real-time structure from motion system.

1. Introduction

In this paper we present a new minimal solver for use in stereo camera based structure from motion (SfM) or visual simultaneous localization and mapping (VSLAM). One popular application would be VSLAM for a humanoid robot. Our approach is analogous to human vision where both eyes overlap in only part of the total viewing frustum. Excluding prior models humans posses, such as relative sizes of objects, expected relative positions, expected ego-motion etc, depth could be perceived from the region of overlap between our eyes while rotation is derived from both overlapping and non-overlapping regions. This configuration of eyes (or cameras) provides a large total field of view of the two cameras while at the same time allowing

the scale to be fixed based on triangulation in the camera's region of overlap. This gives the best of what a two camera system can deliver, a wide field of view for accurate rotation estimation with an absolutely-scaled translation measurement.

Our solution method is based on the observation that a feature visible in all four cameras of the two poses of a stereo camera constrains the relative pose of the second stereo camera to be on a sphere around the feature which has a known position relative to the first stereo camera pose due to its triangulation, as shown in Figure 1. Features seen in only the left or right camera at both poses (two-view features) are labeled $S_{1..3}$ and the feature seen in all four camera views is labeled Q(four-view feature). Please note that within this paper we refer to 2D image matches as correspondences and 3D features or their projections as features. The constraint leaves three degrees of freedom, two for the location of the second camera on the sphere and the third for rotation in the plane tangent to the sphere. We have found algebraically that there is a degeneracy when using all three two-view correspondences from the same camera (left or right) and the camera is the same distance from the four-view feature at both poses. To avoid this degeneracy we use two two-view correspondences in the left (or right) camera and one two-view correspondence in the other camera to solve for the remaining three degrees of freedom.

The major advantage of our approach is that it uses the total field of view of the two camera system to determine the rotation, which maximizes accuracy. Given a stereo pair with a small overlap between its views, one could triangulate 3D points in the region of overlap in the first stereo camera and use the three point perspective pose solution [7] or its generalized incarnation [12] to find the relative pose of the second stereo camera with respect to the first camera. Due to the small overlap, the relative rotation and translation from these methods could be inaccurate because most of the 3D points would be in approximately the same viewing direction relative to the first camera. This problem is
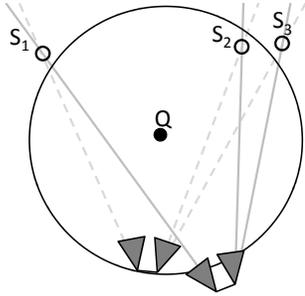
Figure 1. Geometry of stereo pair relative pose problem. Features seen in only the left or right camera are labeled S_i (two-view features). The feature seen in both left and right cameras at both times is labeled Q (four-view feature).
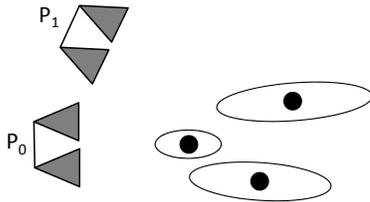
Figure 2. The effect of a depth uncertainty on the accuracy of the three point pose estimation method. At the initial camera pose P_0 the three features are triangulated. Their maximum likelihood position and error covariances are shown. This large error in depth can result in a highly inaccurate estimate of the second camera pose P_1.

illustrated in Figure 2. Our method fixes the rotation with features which do not have to be in the region of overlap. Accordingly, we get a more accurate rotation with small overlap while suffering from less inaccuracy in translation because we require only one feature to be triangulated accurately at two different poses rather than three features to be triangulated accurately in one pose.

Our solver necessitates some modification to standard RANSAC [6] to account for the differing amount of information in a two-view correspondence and a twice-triangulated four-view correspondence. We develop a method to weigh the relative information in each type of inlier to determine the best solution in the RANSAC process. We also show how to modify the RANSAC stopping criteria to account for two classes of measurement.

2. Background

Estimating SfM using multi-camera systems has been a topic of recent interest in the literature. In some instances one can take advantage of a reduced degree of freedom (DOF) motion model to simplify SfM. This can be used to estimate the planar motion of a wheeled robot for example. This approach was taken by Stewenius and Astrom in [16] where they theoretically derive solutions for planar SfM for many different combinations of number of rigidly mounted

cameras, features and multi-camera system poses. In contrast, our approach is to be used with two-camera systems which move with a full 6DOF.

A few approaches to solve for 6DOF motion of a multi-camera system exist which can solve for the system's motion without any overlapping views. Kim *et al.* described one such method in [8]. They first calculate essential matrices for all the system's cameras and decompose these essential matrices into rotation and translation direction. They show that the scale of the motion can then be solved as a triangulation problem.

Another approach was presented by Clipp *et al.* [3]. They solved the same problem using the minimal six point feature correspondences, five in one camera to determine the rotation and translation direction of the multiple camera system and the sixth in another camera of the multi-camera system which determines the scale. Both of these methods [3, 8] are based on decomposing the translation of the systems's cameras into translation of the total system and the rotation induced translation. This can only be done in special circumstances when these two vectors are not parallel or close to parallel. Accordingly, these methods are only applicable to certain key frames in any general video sequence.

The generalized epipolar constraint of multi-camera systems was developed by Pless in [13] to allow a network of rigidly-coupled cameras to be treated as a single imaging device for SfM. Pless suggests a linear 17-point solution for relative motion using the generalized epipolar constraint but left numerical results for this approach to future work. Li *et al.* show in [10] that this standard linear 17-point approach to solve for relative motion using the generalized epipolar constraint based on singular value decomposition does not work in many common scenarios and propose a non-minimal, linear solution that does.

Using Groebner bases Stewénius and Nistér [17] showed that there are two minimal cases for the relative pose of a generalized camera and developed the solution for the relative pose of two views with six ray correspondences. They showed that up to 64 solutions may exist for the proposed constraints. Their solution is degenerate if all cameras in the multi-camera system are on a line, which is the case with a stereo camera (our target here).

At the other end of the spectrum are methods which require all features used in pose estimation to be in the multi-camera system's region of overlap in the first pose. These include the before mentioned three-point perspective pose problem [7] and the generalized three point perspective pose problem [12]. While they don't suffer from the pure translation degeneracy, they do suffer when the views' overlap is small, necessitating a large tradeoff between overlap and total field of view.

Our approach, using one four-view feature to fix the

camera translation and three two-view correspondences to find the rotation, occupies a middle ground in the space of problems considered until now.

3. Solution Method

In this section we describe our approach for relative pose estimation between two poses of a stereo rig. Let P_0 and P_1 denote the projection matrices of the left and the right camera for the first time instance, and P'_0 and P'_1 are those for the second time instance. Without loss of generality, we assume a rectified stereo system, i.e. $P_0 = (I|0)$ and $P_1 = (I|b)$, where $-b$ is the baseline between the cameras on the stereo rig. General configurations can be reduced to this case by appropriate rotation of the 2D feature positions (corresponding to 3D camera rays emerging from the projection center).

Let $(R|t)$ denote the Euclidean transformation between the time instances, i.e.

$$P'_0 = (R|t) \quad \text{and} \quad P'_1 = (R|t+b).$$

The 3D point visible in both cameras has coordinates X for the first time instance and Y in the second instance (always with respect to the left camera in the rig). Hence,

$$Y = RX + t,$$

and $t = Y - RX$. 2D feature matches $p_0 \leftrightarrow q_0$ visible only in the left camera must satisfy the epipolar constraint,

$$q_0^T E_0 p_0 = q_0^T [t]_\times R p_0 = 0. \tag{1}$$

The epipolar constraint for feature correspondences $p_1 \leftrightarrow q_1$ only visible in the right camera can be easily derived as

$$
\begin{aligned}
q_1^T E_1 p_1 &= q_1^T [b+t-Rb]_\times R p_1 \\
&= q_1^T [b+Y-RX-Rb]_\times R p_1 \text{ with } [t=Y-RX] \\
&= q_1^T \big([b+Y]_\times R - R[X+b]_\times\big) p_1, \tag{2}
\end{aligned}
$$

where we used the fact that $[Rx]_\times Ry = (Rx) \times (Ry) = R(x \times y) = R[x]_\times y$ for rotation matrices R.

Overall, Eq. 1 and 2 allows to express both essential matrices in terms of the unknown rotation R, i.e.

$$E_0 = [t]_\times R = [Y]_\times R - R[X]_\times, \text{ and} \tag{3}$$
$$E_1 = [b+Y]_\times R - R[X+b]_\times. \tag{4}$$

In general, with two correspondences in the left camera and one correspondence in the right camera, there are three equations for the three degrees of freedom of the rotation. Using e.g. unit quaternions to represent the rotation matrix R, a polynomial system of four equations can be formulated, which contains the three epipolar constraints and

the unit quaternion constraint. By computing the elimination ideal (e.g. [4]) a 16th degree univariate polynomial (with only even powers) is obtained. The two solutions differing only in sign correspond to the quaternions (q_0, q_1, q_2, q_3) and $(-q_0, -q_1, -q_2, -q_3)$, which actually represent the same rotation.

In our initial experiments we compute a Groebner basis trace for the polynomial system of equations described above (using exact arithmetic in finite prime fields, as in [15]), and generate efficient code automatically to solve real instances of the pose estimation problem. In order to allow real-time processing, we utilize a root finding procedure based on Sturm bracketing instead of using one of the numerically more stable, but substantially slower approaches (e.g. [1, 2]). The observed numerical accuracy of this method degrades with decreasing baselines in terms of the 3D scene depths, which reduces its utility in real-world situations. Nevertheless, the assumption of small motion (in particular small rotations) of the camera system over time (also commonly employed in differential feature tracking methods) allows us to simplify the polynomial system as follows.

First, we represent the rotation R by modified Rodrigues parameters σ [14],

$$R(\sigma) = I + \frac{8[\sigma]_\times^2 - 4(1 - \|\sigma\|^2)[\sigma]_\times}{(1+\|\sigma\|^2)^2}, \tag{5}$$

where σ is a 3-vector. Since the modified Rodrigues parameters can be expressed in terms of the Euler axis $\bar{a}$ and angle θ as $\sigma = \bar{a}\tan(\theta/4)$, the linearization error increases with $\theta/4$ instead of e.g. $\theta/2$ for the classical Rodriguez parametrization. Hence, this particular representation relaxes the assumption of small rotation angles in comparison with other representations.

Under the assumption of small rotations, we approximate $R(\sigma)$ by its second order Taylor approximation and insert the resulting expression into Eqs. 3 and 4. The resulting polynomial system has three equations of degree two. The corresponding Groebner basis trace leads to an 8th degree polynomial and consists of only a few steps, hence the induced solution procedure is considered to be numerically stable. Root finding and backsubstitution give the modified Rodrigues parameters σ and the corresponding rotation through Eq. 5, which is only an approximate solution to the original problem. The reported possible rotations are non-linearly refined to satisfy Eqs. 3 and 4 precisely.

4. RANSAC Considerations

The minimal solution method described in the previous section uses two modes of data points—point feature matches in 3D space, and feature correspondences in 2D. Hence, we have to deviate from the uniform treatment of samples employed in traditional RANSAC settings.

The first modification addresses the refined stopping criterion to account for the potentially different inlier ratios for the two different types of correspondences. The algorithm maintains the inlier ratio ϵ_l for features correspondences only visible in the left camera, the inlier ratio ϵ_r for features visible only in the right camera, and the inlier ratio ϵ_d for features visible in both cameras (four-view features). Without loss of generality we assume that two temporal correspondences from the left camera and one correspondence from the right camera are utilized in the minimal solver. Then the modified stopping criterion is given by

$$n = \frac{log(1 - c)}{log(1 - \epsilon_l^2 \epsilon_r \epsilon_d)}, \quad (6)$$

where n is the number of samples required to achieve confidence c in the solution.

The inlier scoring function for a pose hypothesis also needs adjustedment. In standard RANSAC the total count of inliers is used to score the current hypothesis, which assumes that all data points contain the same amount of information. In the given setup we face a mixed set of data points consisting of 3D points visible in 4 images in total, and 2D feature correspondences. In both cases the latent variables have 3 degrees of freedom per sample—the coordinates of the underlying 3D point. However, the dimension of the observed measurement is either 4-dimensional (two 2D feature positions in either the left or the right camera) or 8-dimensional (2D positions in both cameras at both points in time). Consequently, the residual error lies in the space orthogonal to the fitted manifold [18]. Accordingly, the error space for 3D points visible in all cameras is 5-dimensional (8-3), and the residuals for points visible either in the left or right camera is 1-dimensional (4-3). Therefore, inliers visible in both cameras carry five times more information than single camera correspondences. Thus, the utilized weighting to combine the respective inlier counts is given by

$$score = 5 \times \#Inliers(3D)$$
$$+ \#Inliers(left) + \#Inliers(right), \quad (7)$$

where $\#Inliers(3D)$ denotes the number of inliers among fully visible 3D points (four-view correspondences), and $\#Inliers(left)$ and $\#Inliers(right)$ designate the inlier counts of the respective two-view correspondences.

5. Minimal Problem for Stereo Cameras

In Table 1 we show the excess constraints for the case where we have two stereo camera views for differing numbers of two-view and four-view correspondences. The case with three two-view correspondences and one four-view correspondence has zero excess constraints as it is a minimal solution. A second, as yet unstudied, minimal case

		3D			
		0	1	2	3
	0	-6	-3	-1	3
	1	-5	-2	**0**	4
	2	-4	-1	2	5
2D	3	-3	**0**	3	6
	4	-2	1	4	7
	5	-1	2	5	8
	6	**0**	3	6	9

Table 1. Excess constraints for two stereo camera views

uses two four-view correspondences and one two-view correspondence. That solution is left to future work.

Minimal cases are when the number of constraints (independent equations) equals the number of degrees of freedom (unknowns). Our proposed constraint allows for a minimal solution. Our situation is: M poses of a stereo camera, N two-view correspondences, each observed in either the left or right camera only, and Q four-view correspondences seen in both the left and right camera at both points in time. The degrees of freedom are six for each camera, three for each two-view correspondence and three for each four-view correspondence less the six for the coordinate frame of the first camera leaving $6M + 3(N + Q) - 6$ degrees of freedom. The number of constraints is $2MN + 3MQ$, two for each two-view correspondence in each camera and three for each four-view correspondence in each camera. Each four-view feature yields three constraints because it is triangulated using the known stereo camera geometry. Table 1 gives an overview of the number of constraints for all solvable combinations of 3D feature matches and 2D correspondences for two stereo camera poses.

When a solution is under-constrained not all constraints are independent and the geometry must be considered in calculating the degree of under-determinacy. When over-constrained each additional feature or correspondence gives independent information to the system. This is why there appears to be an inconsistency in the pattern of Table 1 between under-constrained and over-constrained geometry.

6. Degenerate Cases

In this section we describe certain configurations of features which lead to degeneracies in our solution. The major reason we use two two-view correspondences in the left (or right) camera and one two-view correspondence in the other camera to solve for the rotation is a degeneracy that can occur when using correspondences only from one of the cameras. If thee correspondences are selected from the left camera to solve for the rotation and the four-view feature is equidistant from the left camera at both poses then we have determined algebraically that the solution is degenerate. The same goes for the right camera. However,
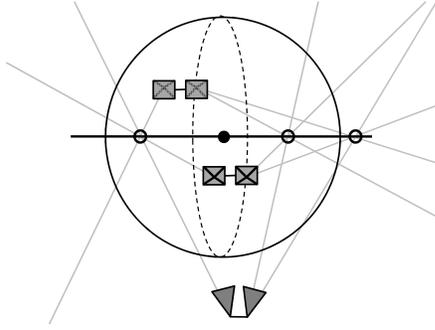
Figure 3. The first degenerate case. When all the features (two-view and four-view) lie on a 3D line through the four-view feature the second camera can be anywhere on a circle which is the intersection of the sphere and a plane orthogonal to the line containing the four-view feature.

Figure 4. Rotation error and translation direction error after non-linear refinement under varying noise with and without outliers.

this is not a degeneracy of our method because we use two-view correspondences from both the left and right cameras to solve for the rotation.

One truly degenerate case that may arise with our method in practice is when all three features, which give rise to the 2D correspondences, lie of a line through the 3D point used in the minimal solution at the center of the sphere. In this case the camera can be anywhere on a circle described by the intersection of the sphere and a plane through the center of the sphere orthogonal to the line. This configuration is depicted in Figure 3. This configuration might occur in man made environments where straight lines are present. However, this configuration is a common degenerate for all relative pose solution methods.

7. Synthetic Experiments

In this section we evaluate the performance of our minimal solver using two synthetic experiments. First we evaluate the performance after nonlinear refinement of the solver under varying levels of image noise with and without outliers to test the solver's ability to deal with corrupted data. Second we compare our solver to the three point perspective pose solution without refinement while decreasing the overlap of the stereo pair by rotating the cameras on the rigid system around their vertical axes.

The first experimental setup tests random motions of a stereo camera. For ease of explanation we assume all units of length are in meters. The two cameras have a baseline of 0.5 m, have parallel optical axes and are placed in a standard stereo camera configuration where both camera centers are on a line from the left optical axis to the right optical axis orthogonal to the axes. The first camera is placed with identity rotation and the left camera of the stereo head at the origin. Three dimensional feature points are distributed in a 20x20x20m volume in front of the camera. The second camera pose is generated by first randomly translating the
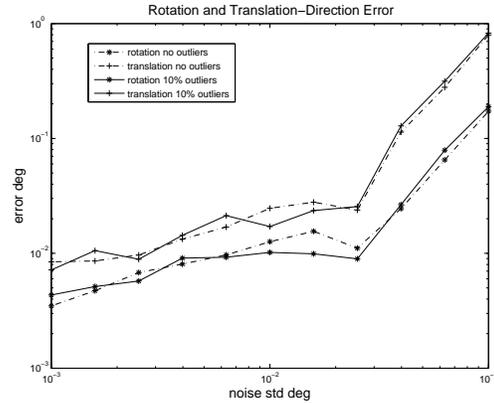
camera between 0.2 and 3.5 meters in a random direction. The minimum translation reduces the effect of being close to the degenerate case when the two cameras are the same distance from the 3D feature, which is the center of rotation. The second stereo pair is then rotated randomly up to five degrees in each of the three rotation axes. Based on visibility we divide the 3D features into three classes: those that can be seen in both cameras of the stereo system at both times (four-view features), those that can be seen only in the left camera at both times and those that can be seen only in the right. In this way we model the effect of a limited overlap in the field of view.

We tested the performance of the novel minimal solver under varying levels of noise and outliers. We use RANSAC [6] to find an initial solution and do a bundle adjustment on the inliers to refine the relative pose solution. Figure 4 shows the rotation and translation direction error with varying levels of noise added to the features in the images with and without outliers. Given that our method uses the 3D location of one feature, we triangulate the noisy image measurements of this feature in both stereo pairs independently and use the triangulated point locations as input to our solver. Image noise is reported in degrees. For context 0.1^o corresponds to two pixels of a camera with a sixty degree field of view and 1200 horizontal pixels. Figures 4 and 5 clearly show that our solver is able to separate inliers from outliers in the presence of noise.

The second experiment is designed to test our method's performance vs. the three point perspective pose method (p3p) in a typical indoor scenario. The camera is placed in a corridor which has 3D features randomly distributed in the 0.1m thick walls. The corridor is 20m long, 2m high and 2m wide. The first stereo camera is placed in the middle of the corridor pointing down the corridor at the far wall, which is 10m away from the camera. The second stereo camera is randomly translated and rotated in a way that it is moving
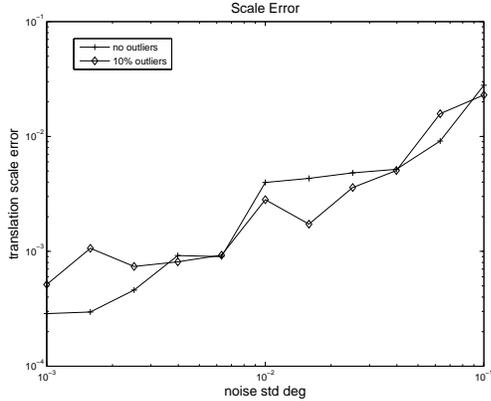
Figure 5. Error in the scale of camera translation under varying noise with and without outliers. Error is given as
$$abs((norm(T_{est}) - norm(T_{true})/norm(T_{true}))).$$

down the hall and points on the far wall are visible.

We progressively reduce the overlap between the cameras by rotating the left and right cameras' optical axes away from each other. We then compare the accuracy of the relative pose calculated using our method and (p3p) after RANSAC but without non-linear refinement. This provides a measure of how close the RANSAC solution is to the true solution. The closer to the true solution the more likely it is that a non-linear refinement will find the global minimal error solution. We test both methods on exactly the same random motions, 3D feature locations and same noisy feature measurements over 1000 trials and report the average results.

For the p3p method we first triangulate 3D points between the left and right cameras of the stereo system and the initial pose, P_0. We then use the projections of these triangulated features in the left image of the stereo head at the second pose, P_0', to calculate the relative pose. We calculate inliers and outliers and score both the p3p solution and our solution method in the same manner. We use an adaptive stopping criterion so that we can compare the required number of RANSAC samples to reach 99% confidence. We also compare the rotation and translation direction error and scale error of the two methods. In Figures 6 through 9 the percentage on the legend shows the percent overlap at infinity between the cameras. The cameras have a 60° field of view horizontally and vertically.

Comparing Figures 6 and 7 one can clearly see that the performance of the p3p method decreases with decreased overlap in the cameras while our method has virtually constant performance regardless of overlap. With 100% overlap p3p out-performs our method. However, with 25% overlap the two methods perform comparably and with 5% overlap our minimal solution out-performs p3p for typical noise values. Figure 8 shows that our method performs with
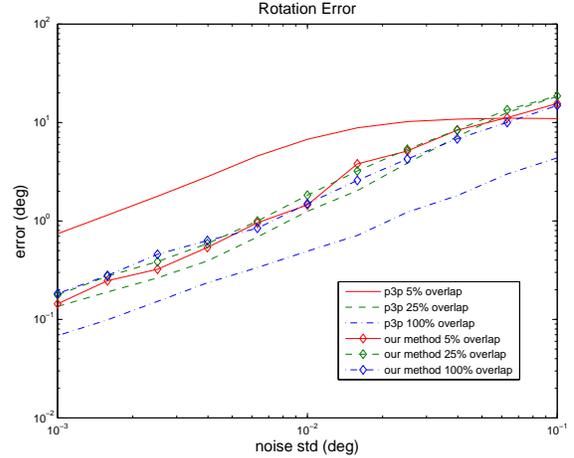
Figure 6. Comparison of rotation error after RANSAC without outliers. This is the error for the best sample from RANSAC without non-linear refinement.
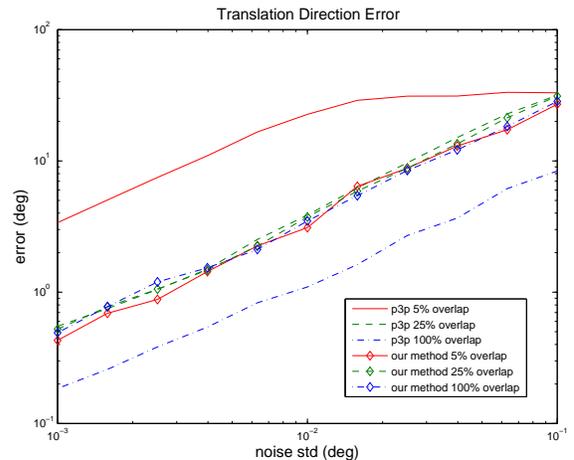
Figure 7. Comparison of translation direction error after RANSAC without outliers. This is the error for the best sample from RANSAC without non-linear refinement.

roughly the same scale error regardless of overlap while the p3p method degrades. Figure 9 shows that p3p requires more RANSAC samples to find a 99% confidence solution than our method when the region of overlap is reduced.

8. Experimental Evaluation in Structure from Motion

To demonstrate our minimal solver we have incorporated it into a real-time (12fps processed), stereo camera based structure from motion system. The system does 2D feature tracking using a graphics processing unit (GPU) implementation of multi-camera scene flow [5]. This work is an extension of Kanade-Lucas-Tomassi (KLT) tracking [11] into three dimensions. Features are matched between the two cameras in the stereo head and triangulated. Image motion
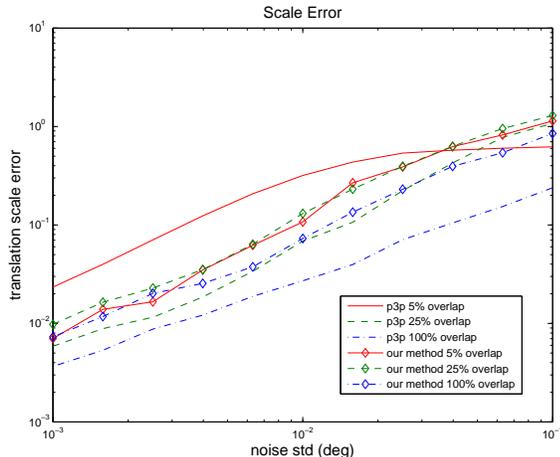
Figure 8. Error in the length of translation direction after RANSAC without outliers for out solution method and p3p with varying stereo camera overlap. No refinement is performed on the best RANSAC sample.
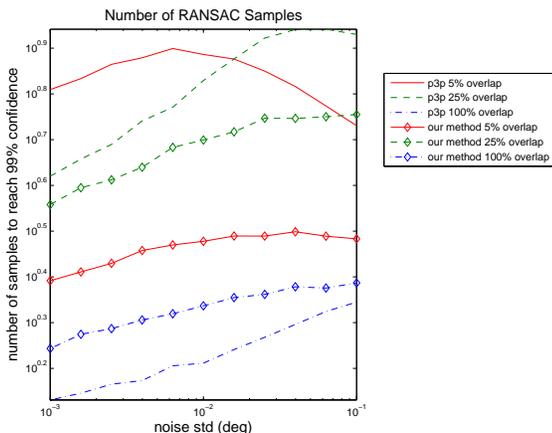
Figure 9. Number of samples required to achieve 99% confidence in the best solution for out solution method and p3p with varying stereo camera overlap.

can then be parameterized as the motion of the 3D feature in front of the camera. This gives accurate correspondences between the four cameras of a stereo pair at two different times. In addition, we also track features that cannot be matched between the left and right image using a standard image space KLT formulation. The image feature tracking runs at approximately seventy frames per second including stereo feature matching and feature re-detection.

Estimating the scene structure and camera motion is done using a RANSAC framework to find an initial pose estimate, followed by a local bundle adjustment to refine the camera poses and 3D feature estimates. Structure from motion is performed only on key frames, which are determined based on the average feature motion in the images. This considerably speeds up processing of a video sequence

without significant loss in accuracy. The RANSAC framework uses the minimal solver described in this paper and makes the modifications to RANSAC mentioned in section 4. The local bundle adjustment is performed on the previous seven key frames and all of the features visible in at least two of those views. Additionally, the oldest two camera poses are held fixed to ensure the continuity of the SfM results as the sliding window of bundle adjusted views is moved forward in time.

Figure 10 shows a top down view of the left camera path calculated using a video sequence shot in an office environment. Example images of the video sequences are show in Figure 11. The office loop is approximately 18 by 10 meters. The camera rounded the loop twice. The path was not exactly the same in both trips around the loop, which accounts for most of the variation of the paths. Note the upper left of Figure 10, where the camera path crosses over itself three times just before the clockwise turn. This was a point of constriction in the environment, which forced the camera to take the same position on each trip around the loop and is shown in the top right image of Figure 11.

9. Conclusion

In this paper we have introduced a novel minimal solution for the relative pose of a stereo camera using one feature observed in all four views, two two-view correspondences from the left (or right) camera and one two-view correspondence from the other camera. Our approach allows the scaled translation to be estimated between poses while at the same time enables a wide total field of view to increase the relative motion estimation accuracy. We have evaluated our solver on synthetic data with noise and outliers. An accuracy evaluation of an automatically generated solver using our constraints as proposed in [9] will be the subject of future work. Additionally, we demonstrated our solver's application in a real-time structure from motion system.

References

[1] M. Bujnak, Z. Kukelova, and T. Pajdla. A general solution to the P4P problem for camera with unknown focal length. In *CVPR*, 2008. 3

[2] M. Byröd, K. Josephson, and K. Aström. Improving numerical accuracy of Gröbner basis polynomial equation solver. In *ICCV*, 2007. 3

[3] B. Clipp, J.-H. Kim, J.-M. Frahm, M. Pollefeys, and R. Hartley. Robust 6dof motion estimation for non-overlapping, multi-camera systems. *IEEE WACV*, 2008. 2

[4] D. Cox, J. Little, and D. O'Shea. *Ideals, Varieties, and Algorithms*. Springer, 2nd. edition, 1997. 3

[5] F. Devernay, D. Mateus, and M. Guilbert. Multi-camera scene flow by tracking 3-d points and surfels. *CVPR*, 2006. 6
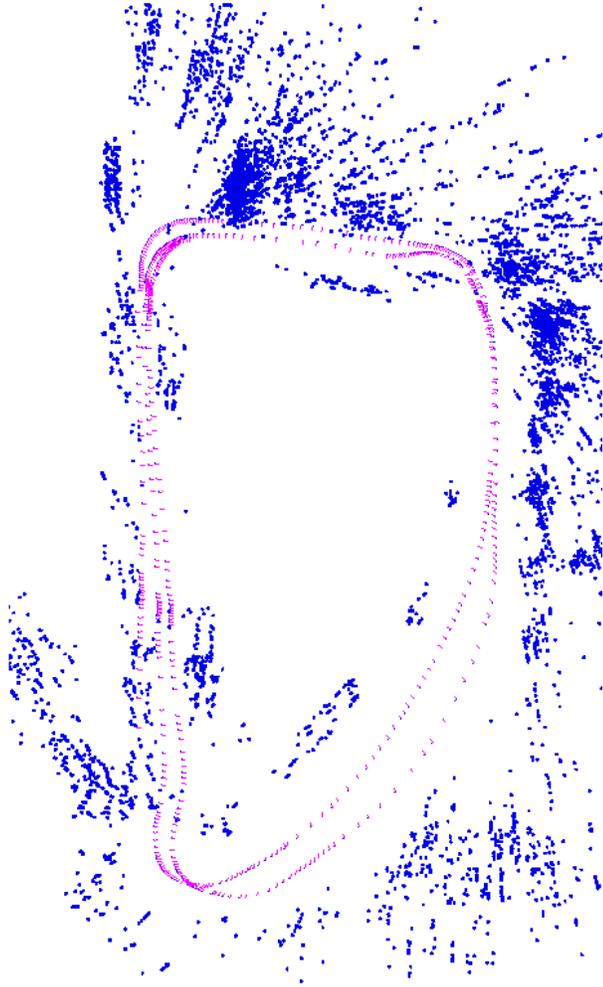
Figure 10. View from above of the reconstructed camera path showing the completed loops. The camera made just over two roughly 18x10m laps around an office environment. No global bundle adjustment was performed. We have attempted to remove features on the ceiling and floor so that the layout of the environment is visible. Left camera axes are drawn as well as a purple line for the baseline.

Figure 11. Sample frames from the left camera of the stereo pair for the office sequence. The images are ordered left to right, top to bottom according to their time in the sequence.

[6] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 1981. 2, 5

[7] R. M. Haralick, C.-N. Lee, K. Ottenberg, and M. Nölle. Review and analysis of solutions of the three point perspective pose estimation problem. *IJCV*, 1994. 1, 2

[8] J. Kim, R. Hartley, J. Frahm, and M. Pollefeys. Visual odometry for non-overlapping views using second-order cone programming. In *ACCV*, 2007. 2

[9] Z. Kukelova, M. Bujnak, and T. Pajdla. Automatic generator of minimal problem solvers. In *ECCV*, 2008. 7

[10] H. Li, R. Hartley, and J. hak Kim. A linear approach to motion estimation using generalized camera models. *CCVPR*, 2008. 2

[11] B. Lucas and T. Kanade. An Iterative Image Registration Technique with an Application to Stereo Vision. In *Int. Joint Conf. on Artificial Intelligence*, 1981. 6

[12] D. Nistér and H. Stewénius. A minimal solution to the generalised 3-point pose problem. *J. Math. Imaging Vis.* 1, 2

[13] R. Pless. Using many cameras as one. In *CVPR*, 2003. 2

[14] H. Schaub and J. L. Junkins. *Analytical Mechanics of Space Systems*. AIAA, 2003. 3

[15] H. Stewénius. *Gröbner Basis Methods for Minimal Problems in Computer Vision*. PhD thesis, Lund University, Apr. 2005. 3

[16] H. Stewenius and K. Astrom. Structure and motion problems for multiple rigidly moving cameras. In *ECCV*, 2004. 2

[17] H. Stewénius and D. Nistér. Solutions to minimal generalized relative pose problems. In *In Workshop on Omnidirectional Vision*, 2005. 2

[18] P. H. S. Torr. Bayesian model estimation and selection for epipolar geometry and generic manifold fitting. *IJCV*, 2002. 4