RELATIVE POSE ESTIMATION USING IMAGE FEATURE TRIPLETS

T. Y. Chuang*, F. Rottensteiner, C. Heipke

Institute of Photogrammetry and GeoInformation, Leibniz Universität Hannover, Germany -
(chuang, rottensteiner, heipke)@ipi.uni-hannover.de

Commission VI, WG VI/4

KEY WORDS: Three-view Geometry, Line Triplet Matching, Relative Orientation, Feature Integration

ABSTRACT:

A fully automated reconstruction of the trajectory of image sequences using point correspondences is turning into a routine practice. However, there are cases in which point features are hardly detectable, cannot be localized in a stable distribution, and consequently lead to an insufficient pose estimation. This paper presents a triplet-wise scheme for calibrated relative pose estimation from image point and line triplets, and investigates the effectiveness of the feature integration upon the relative pose estimation. To this end, we employ an existing point matching technique and propose a method for line triplet matching in which the relative poses are resolved during the matching procedure. The line matching method aims at establishing hypotheses about potential minimal line matches that can be used for determining the parameters of relative orientation (pose estimation) of two images with respect to the reference one; then, quantifying the agreement using the estimated orientation parameters. Rather than randomly choosing the line candidates in the matching process, we generate an associated lookup table to guide the selection of potential line matches. In addition, we integrate the homologous point and line triplets into a common adjustment procedure. In order to be able to also work with image sequences the adjustment is formulated in an incremental manner. The proposed scheme is evaluated with both synthetic and real datasets, demonstrating its satisfactory performance and revealing the effectiveness of image feature integration.

1. INTRODUCTION

Relative pose estimation is the problem of recovering the relative orientation of images and is an indispensable ingredient for any 3D exploitation of imagery such as structure from motion (Scaramuzza and Fraundorfer, 2011). Numerous solutions can be found in the literature. Most feature-based techniques for pose estimation of image sequences were designed in conformity with the following brief description. Features in each image are extracted independently and then tracked/matched into pairs (Jazayeri, 2010; Barazzetti et al., 2011) or triplets (Nistér, 2000; Bartelsen et al., 2012) to estimate fundamental matrices (Torr and Murray, 1997; Nistér, 2004;) or trifocal tensors (Spetsakis and Aloimonos, 1991; Hartley, 1997; Nistér et al., 2004; Reich et al., 2013), often using RANSAC (Fischler and Bolles, 1981) to deal with blunders. Subsequently, the image pairs or triplets are transformed into a common projective frame and image orientation is refined using bundle adjustment (Beder and Steffen, 2008; Sibley et al., 2010; Fraundorfer and Pollefeys, 2010; Schneider et al., 2013). The majority of existing methods was established under the assumption that sufficient image point features can be accurately detected, tracked and matched. However, scenes such as indoor environments, consisting mainly of planar surfaces with little texture, are frequently encountered, and in these environments point features may be hardly detectable, so that a stable point distribution for pose estimation may not be available. In contrast, line features are usually abundant in such conditions and can be more reliably detected and matched despite partial occlusion. Since point and line features supply complementary information of scene geometry, using the combination of these two primitives should render a more robust estimation than those only using one type of feature.

This paper proposes a scheme that solves relative pose estimation using calibrated cameras from point and line triplets, and investigates the effectiveness of feature integration. Compared to

image pairs, a triplet-based scheme poses strong and reliable constraints on pose estimation, and is thus preferred, also because a pair-wise approach for lines is not possible. In this work, acquisition of homologous point triplets is conducted using existing techniques. However, line matching across images raises challenges due to the deficiencies in line extraction and the absence of strong geometric constraints. These often lead to problems such as unreliable endpoints, incomplete topological connections, and asymmetrical radiometric information among putative line matches. Geometric parameters (Roux and McKeown, 1994; Heuel and Förstner, 2001), trifocal tensors and epipolar geometry (Hartley, 1995; Schmid and Zisserman, 1997), radiometric information (Baillard et al., 1999; Scholze et al., 2000; Herbert et al., 2005), and image gradients (Baillard and Dissard, 2000; Wang et al., 2009) are common foundations used to overcome ambiguities in line matching. However, the trifocal tensor as well as epipolar methods generally need known relative orientation parameters, in particular, when using point-to-point correspondences along line segments (Schmid and Zisserman, 2000). Besides, matching groups of line features has the advantage that more geometric information is available for disambiguation (Beder, 2004; Deng and Lin, 2006; Ok et al., 2012), yet it usually is computationally expensive and susceptible to incorrect topological connections or inaccuracy of endpoints. Other methods exploit assumptions such as the projective invariant (Lourakis et al., 2000) or specific configurations of lines in space (Elqursh and Elgammal, 2011), and thus these methods are limited to specific conditions.

In summary, the necessity of accurate prior knowledge, computational complexity, and limitations in specific assumptions are the problems in line matching which need to be addressed according to current literature. We propose a line triplet matching method considering the multiple view geometric relations as well as image gradient similarity, based on available point matches for generating initial values for the pose, to

---

* Corresponding author

alleviate the problems mentioned above. Rather than randomly choosing potential line matches, we generate an associated lookup table that stores information of line candidates to guide the process of selecting putative lines. Thus, the number of potential matches to be checked is reduced. Subsequently, an incremental adjustment (Beder and Steffen, 2008; Reich et al., 2013) for simultaneous pose estimation based on point and line triplets is carried out.

To address this work, the proposed methodology is elaborated in section 2. Afterwards, an experiment using synthetic data is performed to assess the effectiveness of the line matching scheme and to quantify the performance of the proposed approach. These results and the results achieved for a real image sequence are presented in section 3, where we additionally compare the proposed method with a point-based non-commercial software tool. Finally, section 4 concludes this work and gives an outlook into future prospects.

2. METHODOLOGY

Our workflow can be split into two threads for points and lines, respectively.

2.1 Extracting and Matching Points

For point matching, we detect SURF features (Bay et al., 2008) for pair-wise point correspondences, and then perform a cross-search between consecutive pairs for point triplet acquisition. As previously mentioned, the number of point measurements might be inadequate in unfavourable scenes, such as indoor environments. We thus use SURF with loose thresholds to collect more potential point matches. This, however, results in a potentially high number of outliers.

Rather than using RANSAC for blunder elimination, we developed a more specific scheme, which makes use of the fact that the 3D structure of indoor scenes is normally simpler than that of outdoor scenes and mainly consists of planar surfaces. We thus eliminate blunders based on the consistency of local projective transforms. In light of our experiment results, this method rendered good repeatability of correct matches and was found to be robust in dealing with data containing a high ratio of blunders. We do acknowledge, however, that the developed method is not as generally applicable as RANSAC.

The approach starts with a 2D Delaunay triangulation for all matched points in one image of the pair. Then, pairs of triangles that share the longest edge are merged to form a quadrilateral for calculating local projective transformations. OPTICS (Ankerst et al., 1999) and k-means are used to identify the clustering structure of the local transformation parameters. OPTICS was proposed to cluster data based on the notion of density reachability. Neighbouring points in parameter space are merged into the same cluster as long as they are density-reachable from at least one point of the cluster. Then, k-means clusters the filtered transformation parameters into k clusters. Consequently, points belonging to transformations that are not part of any cluster are considered as outliers and are removed. Also, clusters containing less than a minimum of three transformations are deemed to be outliers. Subsequently, a rigorous least-squares adjustment with outlier detection is performed for each accepted cluster. The 2D points that contributed to every transformation are used to calculate the projective parameters for the whole group. A rejection threshold is set for the process in order to remove any points that have a high residual and are potentially outliers. This threshold is calculated dynamically during every iteration of the

adjustment while the value of one pixel is set as the maximum. Finally, all remaining points are considered as correct point matches. More details can be found in (Stamatopoulos et al., 2012).

2.2 Extracting and Matching Lines

Line features are extracted using the LSD detector (Grompone von Gioi et al., 2010) from each image, and introduced into the proposed matching procedure. Based on local projective transformations estimated from the point correspondences (see section 2.1) we start the matching process by mapping lines from the two search images into the reference image. Compatible lines are found as projected lines that lie within a distance tolerance τ_s and an angular tolerance τ_a of the lines in the reference image. Then, we construct a lookup table of all triplet combinations based on the list of the compatible lines and perform similarity measures in terms of three-view geometry and image gradients on each potential line triplets as follows.

The geometric relations among the lines are modelled following Petsa and Patias (1994): As shown in Figure 1, assume that lines l_1, l_2, l_3 are a potential match in an image triplet (I_1, I_2, I_3), where each line is described by two parameters a_i and b_i serving as observations in the following estimation process.
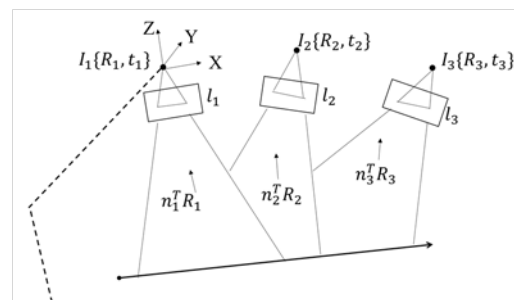

Figure 1. Triplet-line geometry for lines.

For each line l_i, the projection plane is constructed using the line and the projection centre. This plane is described by its normal vector n_i which is given by the cross product of the imaging rays through the start and the end point of the line. In model space these planes can be expressed as:

$$n_i^T R_i (X + t_i) = 0 \qquad (1)$$

where R_i is the rotation matrix, t_i describes the projection centre (R_i and t_i thus containing the parameters of relative orientation) and X is a point on the projection plane. We define the datum by fixing the orientation parameters of the first image and the base between the first and the second images.

Setting the normal vector n_i expressed in model space, to $n_i^T R_i = [u_i \quad v_i \quad w_i]$ and intersecting the three lines a condition equation can be derived (see Petsa and Patias 1994 for details) reading:

$$u_1(v_2 w_3 - w_2 v_3) + v_1(w_2 u_3 - u_2 w_3) + w_1(u_2 v_3 - v_2 u_3) = 0 \qquad (2)$$

Note that while eq. 2 contains the elements of the rotation it does not contain the translations t_i. These can be determined when considering the fact that a point in model space must result in homologous points in image space, yielding eq. (3); for details see again Petsa and Patias (1994):

$$d_2(q_1 v_2 - q_2 v_1)(q_3 w_3 - q_2 w_1)(q_3 v_3 - q_2 v_1) -$$
$$d_3(q_3 v_3 - q_2 v_1)(q_1 w_2 - q_2 w_1)(q_1 v_2 - q_2 v_1) = 0 \quad (3)$$

where $q_1 = u_1 u_3$; $q_2 = u_2 u_3$; $q_3 = u_1 u_2$, and $d_i = n_i^T R_i t_i$.

Consequently, a discrepancy value ε_1 for the line triplet can be computed from the mean of the two formulae when the initial values of the orientation parameters are provided. Besides, the overlapping neighbourhoods between the two lines in the search images and the line in the reference image are identified after the initial mapping. Then a discrepancy value ε_2 is determined using the mean differences of normalized image gradients, compared to those of the line in the reference image:

$$\varepsilon_2 = \frac{|\bar{g}_2 - \bar{g}_1| + |\bar{g}_3 - \bar{g}_1|}{2} \quad (4)$$

where $\bar{g}_1$, $\bar{g}_2$, and $\bar{g}_3$ indicate the mean image gradients across the overlapping neighbourhoods of l_1, l_2, and l_3 in each image, respectively. If the mapped line does not overlap the reference line, the gradient difference will be assigned a pre-defined value so as to decrease the similarity of the potential match. Finally, the similarity measure of the line triple is computed:

$$Sim(l_1, l_2, l_3) = exp\left(-\left(\frac{|\varepsilon_1|}{\alpha} + \frac{|\varepsilon_2|}{\beta}\right)\right) \quad (5)$$

where ε_1 and ε_2 indicate the discrepancy values of the geometric relations and the image gradients, respectively. The parameters α and β control the relative impact of the two indices, and should be adjusted experimentally. $Sim(l_1, l_2, l_3)$ ranges from 0 to 1, a larger value reveals stronger similarity among the line matches.

Since we use the relative orientation parameters estimated from corresponding points as initial values, the similarity measures are correlated to the quality of the prior knowledge. We thus only use it to arrange the order of the combinations in a lookup table containing potential line matches, instead of regarding the score as a criterion for matching. Consequently, line candidates with a higher degree of similarity will be investigated first. Besides, to alleviate the dependency on the quality of the initial values, we gradually refine the orientation parameters and rearrange the lookup table during the matching process.

Once the preliminary lookup table is generated, an iterative procedure is carried out to verify the combinations of potential matches. In view of the geometric conditions, six line triplets suffice for determining the parameters of relative orientation of two images with respect to the reference one. Thus in each iteration of the procedure, six candidate triplets are selected from the lookup table. The selection starts with the six best combinations. From the selected six triplets the relative pose is estimated via a least-squares adjustment. We then check whether the adjustment has converged, and we also check whether the residuals lie below a given threshold τ_r. Candidate triplets that do not satisfy the checks are excluded from further computations, otherwise they are enrolled as accepted triplets, and in the next iteration the next best set of six lines is investigated. After obtaining n sets ($n = 5$ in this work) of six the accepted triplets, a cluster analysis is performed for the pose parameters. The idea is that the valid pose parameters should be similar and thus clustered together in parameter space. All the accepted triplets whose estimations are assembled in the main cluster are used to determine new pose parameters, refining the initial values, and the lookup table is updated accordingly. The matching process is repeated until all the potential line matches in the lookup table are investigated.

2.3 Unified Adjustment for Point and Line Features

Following the matching process described in Section 2.2, for each image triplet, we unify point and line triplets into a common estimation procedure, allowing for the recovery of the optimal relative orientation. For the line triplets we use eqs. (2) and (3), yielding two condition equations per line triplet. For each point triplet, three condition equations are formulated according to the well-known coplanarity condition: for each of the three pairs which can be formed from the three images, the two imaging rays of the corresponding points and the base vector must lie in a plane.

The estimation of the relative orientation of an image triplet with calibrated sensors has 11 degrees of freedom. Thus, the minimal required information is four point triplets or six line triplets or a combination of these, disregarding degenerate cases. Besides the distinctive geometric characters of the individual feature types, considering the combination of features in a minimum solution for an image triplet, which can be realized by one point + four lines, two points + three lines, three points + two lines, or three points + one line, is practically meaningful.

The problem of estimating relative orientation parameters of an image triplet can be formulated using the Gauss-Helmert model:

$$w = A\xi + B(y + e), \ e \sim (0, \Sigma = \sigma_0^2 P^{-1}) \quad (6)$$

where y, e, w, ξ, and P denote the observation vector, the error vector, the discrepancy vector, the vector of incremental unknowns, and the weight matrix, respectively; A and B are the partial derivative coefficient matrices with respect to unknowns and observations, respectively. The mathematical model in eq. 6 is utilized during the line matching process, see section 2.2. In order to be able to also work with image sequences the adjustment is formulated in an incremental manner. We reformulate the Gauss-Helmert model into a Gauss-Markov model and use the incremental least squares adjustment technique described in (Beder and Steffen, 2008; Reich et al., 2013) using sliding image triplets. We first rearrange eq. 6 as $w - By = A\xi + Be$. Then, let $\bar{y} = w - By$ be the new observation vector and $\bar{e} = Be$ be the new error vector. The linear model is yielded:

$$\bar{y} = A\xi + \bar{e}, \ \bar{e} \sim (0, \Sigma_{\bar{e}} = B\Sigma B^T = \sigma_0^2 \bar{P}^{-1}) \quad (7)$$

The unknowns are obtained via:

$$\hat{\xi} = [A^T (BP^{-1}B^T)^{-1}A][A^T (BP^{-1}B^T)^{-1}\bar{y}] \quad (8)$$

The incremental solution is then found as described in Reich et al., 2013).

3. EXPERIMENTS

In this work, we used synthetic and real datasets to validate our approach. The synthetic test is designed to quantify the performance with respect to line triplet matching and pose estimation. The results conducted under realistic conditions serve to better evaluate the potential and limitations of the proposed method. They are evaluated and compared to the result obtained from a point-based software tool (VISCODA, 2012).

3.1 Synthetic Data Test

As shown in Figure 2, image features comprising 10 points and 50 lines, were generated in three views by projecting simulated features, residing in a box of approximately $10 \times 10 \times 10 \ m^3$ in

model space, into image space. Outliers (30 lines) were also included in the dataset. The imaging distance was about 5 to 15 m with respect to the box. The focal length and pixel size were 9 mm and 7 μm, respectively, yielding an image scale of approximately 1:1,000. The lines were derived via fitting of sampled points. A noise of zero mean and σ pixels standard deviation, where $\sigma \in \{0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5\}$, was added to the image coordinates of these points. In addition, due to the lack of grey value information, the discrepancies of the image gradients were given a constant value for the similarity measures. Each test was repeated 1,000 times with feature reallocation. Thus, the feature primitives in each computation differed in distribution, as well as in length and image coverage. The unstructured distribution of the simulated features demonstrated in Figure 2 is notably more complex than the one in a common man-made environment, increasing potential problems in line matching.
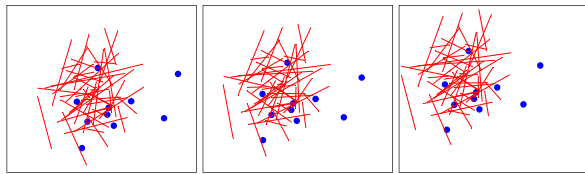

Figure 2. Demonstration of feature distribution.

Tables 1 and 2 provide an insight into the matching results of the synthetic data. Under each specific standard deviation of feature observations, the effectiveness of the proposed method in terms of the matching performance (Table 1) and the pose estimation (Table 2) is reported. Table 1 indicates that the proposed matching scheme achieved an average matching rate up to 98% in which the false cases were largely caused by unsatisfactory line candidates contaminated by the random errors. Besides, it also reveals that the matching effectiveness is stable and robust to outliers especially when the standard deviation of the coordinates is below three pixels.

Std. deviation (pixel)	±0.5	±1.0	±1.5	±2.0	±2.5
Average rates (%)	98	96	95	93	91
Std. deviation (pixel)	±3.0	±3.5	±4.0	±4.5	±5.0
Average rates (%)	88	83	79	74	73

Table 1. Matching performance with synthetic data.

Table 2 contains the estimated relative orientation parameters for the different cases. It can be seen that the differences to the true values ($\sigma = \pm 0$) are rather small. The findings validate the pose estimation approach acceding to the proposed method and suggest, not surprisingly, that more accurate orientation parameters are produced when more accurate feature primitives are available.

σ	Rotations (degree)			Translations (m)		
	ω_2	φ_2	κ_2	t_{x_2}	t_{y_2}	t_{z_2}
	ω_3	φ_3	κ_3	t_{x_3}	t_{y_3}	t_{z_3}
±0.0	0.3	4.2	-6.0	-	-0.3	-0.5
	-0.5	-6.0	-2.0	1.5	-0.4	-0.6
±1.0	0.302	4.194	-6.001	-	-0.301	-0.501
	-0.505	-5.998	-2.001	1.501	-0.401	-0.598
±3.0	0.299	4.205	-6.007	-	-0.301	-0.500
	-0.514	-5.977	-1.967	1.500	-0.402	-0.601
±5.0	0.317	4.193	-6.023	-	-0.300	-0.502
	-0.461	-5.986	-2.003	1.503	-0.403	-0.603

Table 2. Results of estimated orientation parameters. Note that t_{x_2} is not estimated but set to a constant, defining the scale.

Moreover, as shown in Table 3, we quantified the advantages of using both point and line features for pose estimation via the root mean square error (RMSE) derived from 40 check points. As anticipated, the result of the combined features outperformed those obtained from using only points or only lines due to the better distribution and redundancy.

	σ	Points	Lines	Points and lines
	±1.0	0·021	0·016	0·013
RMSE (m)	±3.0	0·028	0·019	0·015
	±5.0	0·037	0·026	0·021

Table 3. Assessment of estimation quality.

3.2 Real Data Test

This experiment involved an image sequence, it was undertaken to verify the results of the simulations under realistic conditions and to conduct a comparison with a purely point-based method. To highlight the advantages of involving line features for relative pose estimation, the images sequence used for the study shows an indoor scene with little texture. It was captured using a hand-held camera in a corridor, see Figure 3. Starting from one point, a person holding the camera moved forward to the end of the corridor and took images at approximately a 45 degree horizontal angle with respect to the direction of movement. The image sequence consists of 40 frames and can be split into two sections. The scene in the first 30 frames contained distinct grey value variations and is favourable for image point detection. On the other hand, the rest of the views consist mainly of planar surfaces with little texture.

(a) Frame 1   (b) Frame 10   (c) Frame 22
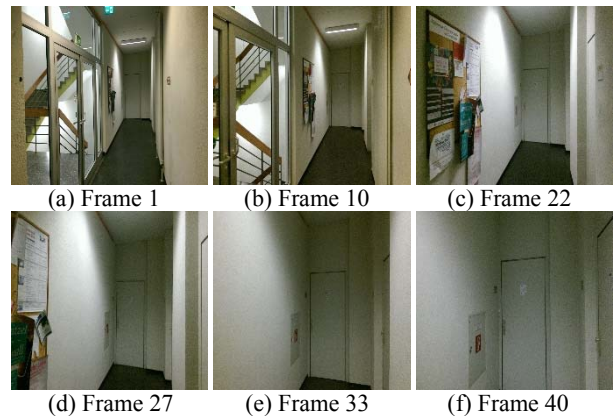(d) Frame 27   (e) Frame 33   (f) Frame 40
Figure 3. Six frames from the corridor sequence.

As described, points were automatically extracted and matched using the SURF descriptor coupled with the introduced outlier removal process. Lines were detected via the LSD detector and matched using the proposed procedure. Figure 4 shows the matching results of the homologous lines in one of the triplets.
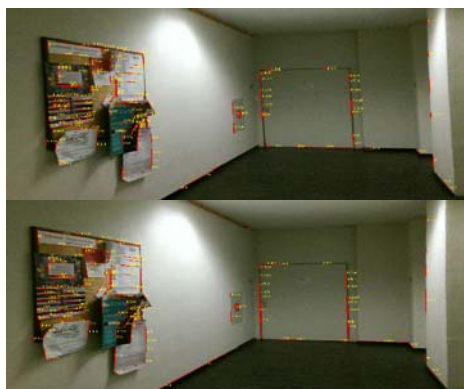
Figure 4. Illustration of the line triplets.

After that, we used the acquired point and line triplets to conduct the relative pose estimation, and compared the recovered trajectory with the one estimated by the point-based software. Figure 5 shows the reconstructed trajectories of the camera motion based on our method and the point-based method, respectively. The two reconstructed trajectories are quite similar until frame 29, from where the point-based method failed to deliver a correct pose estimation for the remaining images (see Figures 5(b) and 5(d)). It is largely due to the fact that the scene in frames 30 to 40 mainly consists of planar surfaces with poor texture. Thus, the point-based method does not have sufficient point features to maintain the computation. Even though we do not have reference data for quality assessment, we can say that our method is superior to the point-based method, because it is able to recover the whole trajectory of this indoor sequence. The result not only highlights the effectiveness of the proposed scheme but also underlines the advantages of feature integration in visual odometry applications.

(a) Top view of our method (b) Top view of the software

(c) Side view of our method
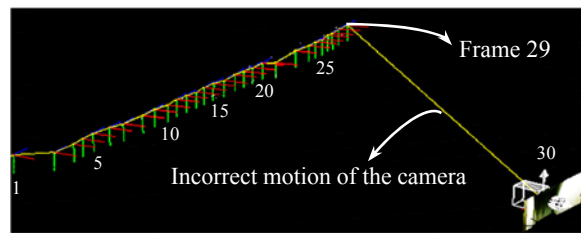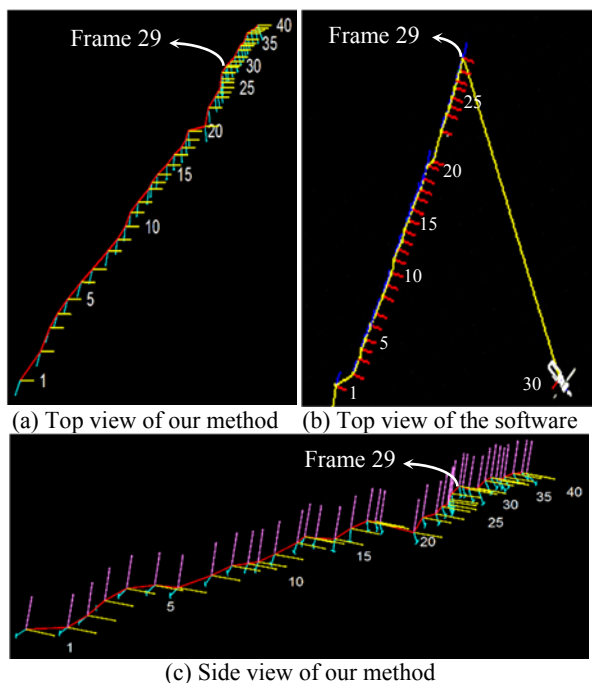
(d) Side view of the software

Figure 5. Top and side views of the reconstructed trajectories using our method (a)(c) and the point-based software (b)(d), respectively.

4. CONCLUSIONS

In this work, we have presented a triplet-wise approach for relative pose estimation from image point and line triplets. The line triplet matching and estimation scheme has been successful and has demonstrated its robustness and effectiveness using synthetic and real datasets. In light of the experiments, our approach outperformed a purely point-based method for an indoor image sequence. This highlights the advantages of integrating different image features and suggests a comprehensive way of dealing with different geometrical structures of scenes. As a result, our approach can be considered as an effective addition to point-based methods for computing the relative pose in unfavourable environments.

Future improvements of the proposed method will address computational optimization and the exploitation of further advantages of line geometry in 3D space. Considering that reconstructed model lines supply more degrees of freedom than model points, it seems promising to align model lines with provided 3D control entities, such as 3D digital vector maps, revealing a potential for automated absolute pose estimation.

REFERENCES

Ankerst, M., Breunig, M.M., Kriegel, H.P., and Sander J., 1999. OPTICS: ordering points to identify the clustering structure, In: *Proceedings of the 1999 ACM SIGMOD International Conference on Management of Data*, Philadelphia, Pennsylvania, USA, pp. 49-60.

Barazzetti, L., Mussio, L., Remondino, F., and Scaioni, M., 2011. Targetless camera calibration. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVIII-5/W16, pp. 335-342.

Baillard, C., Schmid, C., Zisserman, A., and Fitzgibbon, A., 1999. Automatic line matching and 3D reconstruction of buildings from multiple views. In: *International Archives of Photogrammetry and Remote Sensing,* Vol. 32, Part 3-2W5, pp. 69-80.

Baillard, C. and Dissard, O., 2000. A stereo matching algorithm for urban digital elevation models. *Photogrammetric Engineering & Remote Sensing,* 66(9), pp. 1119-1128.

Bartelsen, J., Mayer, H., Hirschmüller, H., Kuhn, A., and Michelini, M., 2012. Orientation and dense reconstruction of unordered terrestrial and aerial wide baseline image sets. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* Melbourne, Australia, Vol. I-3, pp. 25-30..

Bay H., Ess A., Tuytelaars T., and Gool L.V., 2008. SURF: Speeded Up Robust Features. In: *Computer Vision and Image Understanding*, Vol. 110, No. 3, pp. 346-359.

Beder, C., 2004. A unified framework for the automatic matching of points and lines in multiple oriented images. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 35, Part B3, pp. 1109-1113.

Beder, C. and Steffen, R., 2008. Incremental estimation without specifying a-priori covariance matrices for the novel parameters. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Anchorage, USA, pp. 1-6.

Deng, D. and Lin, X.Y., 2006. A fast line segment based dense stereo algorithm using tree dynamic programming. In: *European Conference on Computer Vision*. Lecture Notes in Computer Science, Vol. 3953, Springer, Heidelberg, pp 201-212.

Elqursh, A. and Elgammal, A., 2011. Line-based relative pose estimation. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3049-3056.

Fischler, M.A. and Bolles, R.C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. In: *Commun. ACM*, Vol.24, Part 6, pp. 381-395.

Fraundorfer, F. and Pollefeys, M., 2010. A constricted bundle adjustment parameterization for relative scale estimation in visual odometry. In: *Proc. IEEE Int. Conf. Robotics and Automation,* pp. 1899-1904.

Grompone Von Gioi, R., Jakubowicz, J., Morel, J.M., and Randall, G., 2010. LSD: a fast line segment detector with a false detection control. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 32(4), pp. 722-732.

Hartley, R.I., 1994. Lines and point in three views - an integrated approach. In: *Proc. ARPA Image Understanding Workshop. 6 p.*

Hartley, R.I., 1995. A linear method for reconstruction from lines arid points, In: *IEEE International Conference on Computer Vision.* pp. 882-887.

Hartley, R.I., 1997. Lines and points in three views and the trifocal tensor. *International Journal of Computer Vision*, 22(2), pp. 125-140.

Herbert, B., Vittorio, F., and Luc, V.G., 2005. Wide-base line stereo matching with line segments. In: *IEEE Conference on Computer Vision and Pattern Recognition. pp.* 329 - 336.

Heuel, S., and Förstner, W., 2001. Matching, reconstructing and grouping 3D lines from multiple views using uncertain projective geometry. In: *Proc. Computer Vision and Pattern Recognition,* pp. 517-524.

Jazayeri, I., 2010. *Image-based modelling for object reconstruction.* Ph.D. thesis, The University of Melbourne.

Kim, Z., and Nevatia, R., 2004. Automatic description of complex buildings from multiple images. *Computer Vision and Image Understanding,* 96(1), pp. 60-95.

Lourakis, M.I.A., Halkidis, S.T., and Orphanoudakis, S.C., 2000. Matching disparate views of planar surfaces using projective invariants, *Image and Vision Computing*, 18(9), pp. 673683.

Mayer, H., 2002. Estimation of and view synthesis with the trifocal tensor. In: *International Archives of Photogrammetry and Remote Sensing and Spatial Information Sciences*, Vol. 34, Part 3A, pp. 211-217

Nistér, D., 2000. Reconstruction from uncalibrated sequences with a hierarchy of trifocal tensors. In: *European Conference on Computer Vision*, LNCS, Vol. 1842, Springer Berlin Heidelberg, pp. 649-663.

Nistér, D., Naroditsky , O., and Bergen, J., 2004. Visual odometry. In: *IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 1, pp. 652-659.

Nistér, D., 2004. An efficient solution to the five-point relative pose problem. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 26(6), pp. 756-770.

Ok, A.O., Wegner, J.D., Heipke, C., Rottensteiner, F., Soergel, U., and Toprak, V., Matching of straight line segments from aerial stereo images of urban areas. *ISPRS Journal of Photogrammetry and Remote Sensing,* 74, pp. 133-152.

Petsa, E., and P. Patias, 1994. Relative orientation of image triples using straight linear features, In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (39) III*, pp. 663-669.

Reich, M., Unger, J., Rottensteiner, F., and Heipke, C., 2013. Online compatible orientation of a micro-UAV based on image triplets. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* II-3/W2, pp. 37-42.

Roux, M. and McKeown, D. M., 1994. Feature matching for building extraction from multiple views. In: *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, Seattle, Washington, pp. 46-53.

Wang, Z., Wu, F., and Hu, Z., 2009. MSLD: a robust descriptor for line matching. *Pattern Recognition,* 42(5), pp. 941-953.

Scaramuzza, D. and Fraundorfer, F., 2011. Visual odometry [tutorial]. *IEEE Robotics and Automation Magazine*, 18(4), pp. 80-92.

Schmid, C. and Zisserman, A., 1997. Automatic line matching across views, In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 666-671.

Schmid, C. and Zisserman, A., 2000. The geometry and matching of lines and curves over multiple views, *International Journal of Computer Vision*, 40(3), pp. 199-233.

Schneider, J., Läbe, T. and Förstner, W., 2013. Incremental realtime bundle adjustment for multi-camera systems with points at infinity. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XL-1/W2, pp. 335-360.

Scholze, S., Moons, T., Ade, F., and Van Gool, L., 2000. Exploiting colour for edge extraction and line segment stereo matching. In: *International Archives of the Photogrammetry,*

Remote Sensing and Spatial Information Sciences, Vol. 33, Part B3, pp. 815-822.

Sibley, G., Reid, I., Newman, P., 2010. Vast-scale Outdoor Navigation Using Adaptive Relative Bundle Adjustment. *Journal of Robotics Research.*

Spetsakis, M. and Aloimonos, J.Y., 1991. A multi-frame approach to visual motion perception. *International Journal of Computer Vision,* 6(3), pp. 245-255.

Stamatopoulos, C., Chuang, T.Y., Fraser, C.S., and Lu, Y.Y., 2012. Fully automated image orientation in the absence of targets. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* Melbourne, Australia, Vol. I-3, pp. 303-308.

Torr, P. and Murray, D., 1997. The development and comparison of robust methods for estimating the fundamental matrix. *International Journal of Computer Vision*, pp. 1-33.

VISCODA, 2002. "Voodoo Camera Tracker: A tool for the integration of virtual and real scenes", Hannover, Germany. http://www.viscoda.com/en/voodoo-manual (10 Jan. 2015).