

Full 6DOF Pose Estimation from Geo-Located Images

Clemens Arth, Gerhard Reitmayr, and Dieter Schmalstieg

Graz University of Technology

Abstract. Estimating the external calibration – the pose – of a camera with respect to its environment is a fundamental task in Computer Vision (CV). In this paper, we propose a novel method for estimating the unknown 6DOF pose of a camera with known intrinsic parameters from epipolar geometry only. For a set of geo-located reference images, we assume the camera position - but not the orientation - to be known. We estimate epipolar geometry between the image of the query camera and the individual reference images using image features. Epipolar geometry inherently contains information about the relative positioning of the query camera with respect to each of the reference cameras, giving rise to a set of relative pose estimates. Combining the set of pose estimates and the positions of the reference cameras in a robust manner allows us to estimate a full 6DOF pose for the query camera. We evaluate our algorithm on different datasets of real imagery in indoor and outdoor environments. Since our pose estimation method does not rely on an explicit reconstruction of the scene, our approach exposes several significant advantages over existing algorithms from the area of pose estimation.

1 Introduction

The topic of *location recognition* became an increasingly important area of research in Computer Vision (CV) recently. In the literature the meaning of *location* is ambiguous, however. Though the term is mainly used to refer to the task of position estimation in 2D or 3D, *location recognition* is also used to denote algorithms returning only a qualitative location description rather than a numeric geo-location. Such a qualitative description might be, for example, a *location tag* such as the name of a street or a building, or even the name of a city or country only.

In contrast to the rather weak definition of *location recognition*, the term *pose estimation* can be clearly defined. It describes the task of determining the external calibration of a camera with respect to a more-or-less known environment. Solutions to the *pose estimation* problem can be categorized by the number of degrees of freedom (DOF) considered. A higher number of DOF implies a more complete extrinsic camera calibration and, in turn, more accurate estimation results. At the same time, the complexity of the required algorithms increases significantly. Indeed, *location recognition* can be considered a subproblem of the

pose estimation problem, as the approaches proposed for location recognition consider only up to four DOF at maximum (i.e. a 3D location and a compass direction). Thus *location recognition* approaches are limited concerning their practical relevance for certain applications. The applicability of a 2, 3, or 4DOF pose is restricted to simple operations, like visualizing the position on a map. In contrast, for Simultaneous Localization and Mapping (SLAM) or Augmented Reality (AR), a full 6DOF pose estimate is required. This is more demanding and traditionally requires some reconstruction or model of the scene. Consequently, application requirements strictly dictate the suitability of algorithms and may render *location recognition* approaches completely useless.

The contribution of this work is the introduction of a novel algorithm for full 6-DOF pose estimation *without the need for 3D reconstruction*. The general idea is illustrated in Figure 1. Traditional pose estimation methods require an explicit model of the environment, such as a sparse point cloud reconstruction [1,2]. Building the reconstruction itself requires lots of images and a considerable number of evolved algorithms - besides vast amounts of computational power - to recover the initial poses of the images and the scene structure. In contrast, our algorithm builds upon a database of images for which only the center of projection of the camera is known, such as images annotated with position estimates from GPS.

For the pairs formed by the query image and each of these reference images, we estimate the epipolar geometry using image features such as SIFT [3]. From known epipolar geometry, the epipole is obtained and, together with the known position of the reference image, is treated as a single 2D-3D correspondence for the query image. Using at least 3 reference images, we finally arrive at a standard 3-point-pose estimation problem. By applying a RANSAC-based scheme in a robust solver, the algorithm is able to estimate the full 6DOF pose of our query camera in a global context. This approach stands in contrast to estimating the pose of a camera from real world points and their image observations. We treat the known camera positions as the world points and the placement of the cameras relative to the query camera as the observations,

Our method has several major benefits for practical applications:

(i) There is no need for the creation of a 3D reconstruction of the scene. We circumvent the tedious data acquisition and model building stage completely. The algorithm still delivers competitive results in terms of speed and accuracy.

(ii) Given a query image, we use standard image retrieval methods to identify a set of reference images depicting roughly the same scenery. We can therefore directly benefit from the improvements in organization and maintenance of image databases and the evolution of image retrieval algorithms.

(iii) The method is intended to work on community photo collections that are available without laborious preparation. The quality of our algorithm is assumed to automatically increase with an increasing number of geo-tagged images and improved geo-tag accuracy.

(iv) Our approach is inherently suitable for application in mobile scenarios. Although most current mobile GPS receivers are known to deliver erroneous

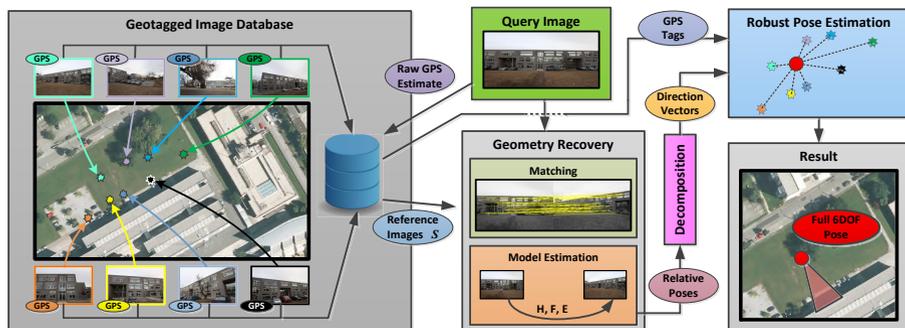


Fig. 1. A schematic illustration of the proposed algorithm. For a query image I a set S of reference images is retrieved. The estimated epipolar geometry between all pairs of I and members of S is expressed in terms of relative poses. This facilitates the reformulation of the pose estimation problem as a 3PP problem.

position estimates, our algorithm is estimated to automatically improve in quality with enhanced GPS receivers in the future.

(v) For mobile applications, any information has to go through a relatively narrow and high-latency communication channel. With the approach proposed, the amount of information to be transmitted can be scaled to shift the workload between a server and a mobile client, depending on the computational resources on the client and the capacity of the communication channel available.

2 Related Work

We divide related work in the area of pose estimation roughly into three groups. The first set of approaches relates to the *location recognition* problem and is mainly based on image recognition methods. The second set of algorithms solves the problem with the help of additional models of the environment, where textured planar models, wireframe models or point cloud reconstructions are used. The last set of approaches is related to *camera calibration*. Approaches from all three areas are discussed in the following.

2.1 Pose Estimation as an Image Recognition Problem

Solutions to the localization problem are usually based on retrieval from a database of images, image features extracted from the images, and some kind of matching algorithm. To speed up the entire procedure, approximate search structures, like vocabulary trees, are commonly used.

Related work based on this idea for localization outdoors using GPS and image databases includes [4,5,6,7,8,9,10], amongst others. The main advantage of these methods is the availability of vast amounts of data. At the same time, image databases can be maintained with relatively little effort and the problem

of image retrieval is well understood. However, known algorithms from this area are not capable of delivering an accurate 6DOF pose, but rather a 2, 3, or 4DOF pose at best.

2.2 Localization from Scene Reconstructions

The second large group of approaches deals with the use of special models of the environment and how to perform localization from them. Since registration is one part of the SLAM problem, a large number of approaches directly stems from this area [11,12,13], another large group is found in the area of live real-time reconstruction [14,15]. Representing the environment as sparse point clouds and using them for localization has been investigated recently, especially in the context of mobile devices [1,2,16,17].

Approaches based on scene reconstructions may be capable of delivering highly accurate 6DOF pose estimates. Especially the use of point cloud reconstructions allows for self-localization in real-time, even on current mobile hardware. However, the automatic creation of sparse point cloud models is still a topic of active research. Moreover, no workflow for maintaining these models over a longer period of time is known so far. SLAM based solutions are prone to drift errors and are known to perform well only in moderately sized scenarios. Real-time reconstruction of large environments requires vast amounts of computational power, making these approaches hardly usable on mobile devices.

2.3 Relation to Camera Calibration

Solving the pose estimation problem is equivalent to finding the external camera parameters in camera calibration, as covered by well-known literature [18]. Early works partially cover the theoretical foundation of our approach, i.e. the possibility to estimate the translation and rotation of cameras on a non-rigid camera rig [19,20]. Also related to our work is the investigation of the viewing graph [21], which studies the minimally required information to recover the full external calibration parameters for a set of cameras. More recently, epipolar geometry is decomposed into epipoles and mutual projections between cameras are investigated to reduce the required number of correspondences [22].

In our approach, we avoid calculating a full calibration for a set of images by intent. However, our approach can be considered a solution to a viewing graph with a star-like configuration (i.e. having an unknown query image in the center and epipolar constraints to a larger set of surrounding images).

3 6DOF Pose Estimation from Epipolar Geometry

The approach described in the following calculates a full 6DOF pose for a query camera with known intrinsic camera parameters by using a database of images tagged with their positions. For the rest of this paper, we denote the cameras, respectively the centers of projection, as C , the images taken from the cameras as

I. For brevity, we do not explicitly mention the use of homogeneous coordinates when it is clear from the context. Our algorithm consists of four separate parts.

1. Given a query image I with a position estimate, first we determine a set \mathcal{S} of nearby images from a database using standard image retrieval methods. The position estimate might be a rough GPS measurement outdoors, or information from Bluetooth or WiFi triangulation indoors. Since image retrieval is a well-known problem, we will not discuss this part further.
2. For all pairs formed by the query image I and the images from \mathcal{S} , the epipolar geometry is estimated using image features. Since the results of traditional feature matching approaches are not guaranteed to be free of errors, we use a robust scheme for outlier removal.
3. From the epipolar geometry extracted from I and each I' in \mathcal{S} , we can derive information about the placement of each reference camera C' relative to the query camera C . This is described in Section 3.1. Epipolar geometry does not enforce any requirements on the real structure present in the scene. However, regular structures are quite common in man-made environments, for example in the form of large planar surfaces. We propose an improvement to our algorithm in Section 3.2 that is based on the estimation of planar structure.
4. The 6DOF pose of the query camera C is estimated by solving a modified 3PP problem. As an analogy to the relationship between world points and their observations in an image, we consider the relationship between the rough positions of the reference cameras C' and their relative direction with respect to the query camera C . A detailed description is given in Section 3.3.

3.1 Relative Translation between Cameras

Given an estimate of the fundamental matrix \mathbf{F} between the query image I and some other image I' from \mathcal{S} , we can extract information about the relative translation of C' with respect to C as follows.

The right null vector of \mathbf{F} is the epipole \mathbf{e} , which is basically the observation of the camera center C' in image I . Effectively, \mathbf{e} describes the direction under which C' is seen from C . After normalizing the epipole by the internal calibration \mathbf{K} of camera C ,

$$\hat{\mathbf{e}} = \mathbf{K}^{-1}\mathbf{e} \quad (1)$$

describes the direction in the coordinate system of C on which C' must reside. Equivalently, if it is possible to estimate the essential matrix \mathbf{E} for images I and I' , the epipole $\hat{\mathbf{e}}$ can directly be calculated as the right null vector of \mathbf{E} .

3.2 Dominant Planar Structure

Planar structure is common in man-made environments. Assuming a scene with dominant planar structure and the internal calibration \mathbf{K} and \mathbf{K}' of both cameras C and C' are known, we can optionally employ a stricter geometric model than

epipolar geometry to estimate the relative placement of cameras with respect to each other.

Given a single physical plane Π , a minimum of four correspondences between I and I' is sufficient for the estimation of a 2D homography H . H encapsulates a 2D projective transformation between images I and I' that is valid for this single underlying plane Π only. For arbitrary scene points \mathbf{X}_i on Π , the 3×3 matrix H transforms the point's observation $\hat{\mathbf{x}}_i$ in I to its observation $\hat{\mathbf{x}}'_i$ in I' such that

$$H\hat{\mathbf{x}}_i = \hat{\mathbf{x}}'_i. \quad (2)$$

Similar to epipolar geometry, the homography H encodes the relative pose change P between C and C' :

$$P = [R|\mathbf{t}], \quad (3)$$

where R and \mathbf{t} denote the rotational and translational parts of the pose change respectively. Let H denote the homography between images I and I' , then

$$H = R + \frac{1}{d}\mathbf{t}\mathbf{n}^T \quad (4)$$

with d being the distance from C to the given plane Π , and \mathbf{n} being the normal vector of Π w.r.t. C . R and \mathbf{t} are the relative rotation and translation of camera C' with respect to camera C . The decomposition of equation 4 can be obtained using the method proposed in [23]. For numerical stability, H is normalized prior to decomposition such that $\det(H) = 1$. As a result, two physically possible solutions for R , \mathbf{t} , d and \mathbf{n} are computed.

Unfortunately, from H the real translation between the cameras C and C' cannot be inferred completely. However, considering the position of a particular camera C' , \mathbf{t} defines a direction along which the camera C must reside. We can transform \mathbf{t} into the coordinate system of camera C , obtaining

$$\mathbf{v} = -R^T\mathbf{t}. \quad (5)$$

\mathbf{v} denotes the direction of camera C' with respect to the coordinate system of camera C . Similarly to the direction estimates obtained in Section 3.1, these estimates can be treated as input to the pose estimation solver described in Section 3.3. Note, that estimating multiple homographies for single image pairs can help to select the correct direction hypothesis in advance (i.e. the one that is a consistent solution to multiple homography decompositions).

3.3 Pose Estimation

After calculating the translation estimates of the reference cameras C'_i in \mathcal{S} with respect to the query camera C , we arrive at a 3PP problem [24]. For pinhole camera models, a known camera calibration K allows conversion of the image measurements \mathbf{x}_i to rays \mathbf{v}_i and their pairwise angle $\angle(\mathbf{v}_i, \mathbf{v}_j)$ can be measured. In this case, three known 3D points \mathbf{X}_i and their corresponding image measurements \mathbf{x}_i give rise to three pairwise angular measurements. These are sufficient to

compute a finite number of solutions for the camera location and orientation. To determine the plausibility of a solution, a fourth observation of a 3D point has to be used.

In our case, we do not work in image space. Our rays \mathbf{v}_i are the direction estimates $\hat{\mathbf{e}}_i$ of the camera centers C'_i obtained either as epipoles (section 3.1) or translation directions (section 3.2). The camera centers C'_i are the known 3D points \mathbf{X}_i . Again we can calculate pairwise angle measurements $\angle(\hat{\mathbf{e}}_i, \hat{\mathbf{e}}_j)$, which leads to the same equation system as in the pinhole case. For three known camera positions C'_i , the pairwise 3D point distances $l_{i,j}$ can be computed. Furthermore, the angles $\theta_{i,j}$ are known from the corresponding direction vectors $\hat{\mathbf{e}}_i$ and $\hat{\mathbf{e}}_j$. The unknowns are the distances d_i between the center of the query camera C (the equivalent for the center of projection in the pinhole case) and the camera center C'_i :

$$\begin{aligned} l_{i,j} &= \|C'_i - C'_j\| \\ \theta_{i,j} &= \angle(\hat{\mathbf{e}}_i, \hat{\mathbf{e}}_j) \\ d_i &= \|C - C'_i\|. \end{aligned}$$

Using the law of cosines, each of the three point pairs gives one equation:

$$l_{i,j}^2 = C_i'^2 + C_j'^2 - 2C'_i C'_j \cos(\theta_{i,j})$$

This is the same polynomial system as in the case of the more commonly used pinhole camera model and can be solved with the same techniques [24]. The main difference is that in the pinhole case the camera calibration matrix \mathbf{K} is used to convert image measurements to vectors and therefore pairwise Euclidean angle measurements. However, in our case, the rays are defined by the translational part of the differential poses.

It is important to mention that in all cases there is a sign ambiguity in the direction under which the camera C' is seen by C . Implementations of the standard 3PP algorithm usually consider directed rays (i.e. basically ignoring the possibility of 3D points lying *behind* the camera center). One way to overcome this issue is to modify the implementation accordingly, considering the negative roots of the quartic as valid solutions. This would then possibly give (at max 8) poses that are partially point mirrored. Those are validated by checking a negative determinant of the rotation component (giving at max 4 solutions). Instead we employ the generalized 3PP estimation algorithm by Nister [25], since it implicitly works on undirected rays at negligible additional cost and returns the same 4 solutions at max. The maximum number of poses is therefore at max 4 for E and F, and at max 32 for H, if the aforementioned elimination by using multiple homographies is ignored.

4 Experimental Results

We evaluated the method on two data sets. The first data set is an indoor scene of an office corner. This data set consists of 111 images and is used to



(a) Images from the indoor corner scenario (111 images).



(b) Images from the outdoor campus data set (156 images).



(c) SfM corner reconstruction.



(d) Google Earth image with a set of camera locations marked.

Fig. 2. Sample images of both scenarios, the indoor SfM reconstruction and a Google Earth image with GPS positions of several images from the campus data set

demonstrate the functionality of our approach in an idealized case. We computed a SfM reconstruction from the images to obtain camera positions (see Figures 2(a) and 2(c)). This procedure allows us to (i) acquire ground truth poses for the 111 cameras with respect to the reconstruction, and to (ii) generate pseudo ground truth poses for arbitrary images using 3PP. An alternative experimental evaluation using a robot arm to acquire ground truth data is scheduled for a future experiment.

The second data set consists of 156 images that were taken outdoors on our campus site. The camera location was recorded with a highly accurate Realtime Kinematic GPS receiver to obtain ground truth camera positions (see Figures 2(b) and 2(d)). Note that we are aware of the fact that current mobile devices are not equipped with such high quality sensors. However due to issues discussed in Section 4.3 we decided to prove our algorithm viable under meliorated conditions, i.e. anticipating future improvements to sensor technology.

4.1 Indoor Scenario

For the indoor data set, each of the 111 images of the reconstruction was matched as a query image against all other images. As described in Section 3, three different relative motion estimates between the query image and all other images

Table 1. Results for the indoor corner data set. Translation error is relative to the ground truth pose from the SfM reconstruction. Rotation error is the rotation angle of the relative rotation between ground truth and estimated pose.

Data	Method	# of Successful Estimates (of at max. #)	# Inl.	Translation Error [mm]		Rotation Error [°]		
				Mean	Std. Dev.	Mean	Std. Dev.	
Rec. Cameras	Fundamental Matrix	110 / 111	66	<i>x</i>	-0.22	6.56	0.70	0.72
				<i>y</i>	-0.23	9.36		
				<i>z</i>	-0.78	11.72		
	Essential Matrix	110 / 111	67	<i>x</i>	-0.23	5.61	0.69	0.72
				<i>y</i>	-0.15	8.00		
				<i>z</i>	0.72	9.99		
	2D Homographies	110 / 111	92	<i>x</i>	0.11	5.87	0.62	0.35
				<i>y</i>	0.31	8.92		
				<i>z</i>	-0.16	9.50		

were calculated, the fundamental matrix F , the essential matrix E and a dominant homography H . All estimations used a RANSAC scheme to obtain an outlier free result. If not enough inliers were found for an image pair, the correspondence was rejected. Finally, we estimated the location of the query camera using the pose estimation approach of Section 3.3 and compared it to the ground truth estimate from the SfM reconstruction. We accepted a localization result, if the RANSAC processes obtained enough inliers and the estimated location was within 10cm of the ground truth location, and within 5° of ground truth orientation. Table 1 shows the localization success rate and the errors in the localizations. We obtain a very high localization rate and a pose accuracy that is comparable to 3D localization from point correspondences.

4.2 Outdoor Scenario

In the second experiment, again each of the 156 reference images was used as a query image against all other images. For this data set, we did not compute a 3D reconstruction but relied on the ground truth position from the RTK-GPS receiver with an accuracy of $\pm 5cm$. Therefore, the orientation of the reference cameras is not known. The ground truth positions were transformed into Earth-Centered, Earth-Fixed (ECEF) coordinates to obtain a map-independent, Cartesian coordinate system. As before query images are matched against reference images and the relative orientation was estimated. We restricted the estimation to the homography model only (see Section 3.2) because the real scene was dominated by planar building surfaces. The estimated camera location was then compared to the ground truth location from GPS. We accepted a localization result, if the error was below $50cm$. Table 2 shows the localization success rate and localization errors in each dimension, Figure 4 shows a histogram of the error in $x/y/z$ direction for the successful pose estimates.

Figure 3 shows the estimated camera locations with respect to the ground truth positions (shown as red cubes) as a qualitative result. The general offset is due to the visualization of the images planes in front of the camera centers. Figure 5 shows the error distributions for the three indoor experiments and the

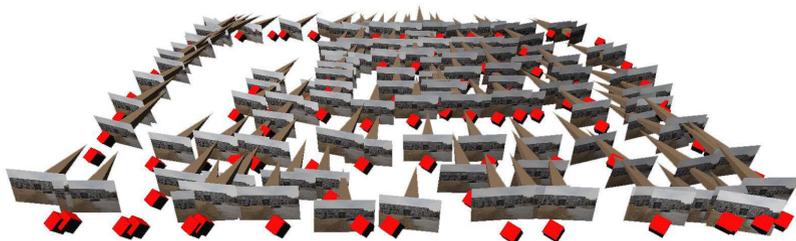


Fig. 3. Localized cameras with respect to the ground truth locations (shown as red cubes)

Table 2. Results for the outdoor campus data set. Translation error is relative to the GPS ground truth locations. We did not estimate rotation error, because no ground truth was available.

Camera Data	Method	# of Successful Estimates (of at max. #)	Axis		Translation Error [mm]	
			x	y	Mean	Std. Dev.
Campus	2D Homographies	142 / 156 (91.03%)	x	-2.63	107.48	
			y	5.94	121.18	
			z	0.61	97.29	

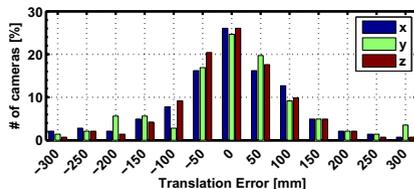


Fig. 4. Histogram of the translation error for all valid camera estimates

outdoor experiment as sorted distances for all cameras. For most cameras we obtain similarly accurate results with only a small percentage of large errors.

4.3 Discussion

The results of the previous sections demonstrate that our algorithm performs very well for the limited test cases. There is a number of significant findings and considerations to mention:

- The estimation of E and F is subject to large errors or can completely fail, if the structure in the scene is largely planar. If the internal camera parameters of the reference images are known, computing H is clearly preferable under these circumstances.
- An important factor causing E and F to fail is the inability of the feature detection algorithm to deliver correspondences reflecting the real 3D structure of the scene. Matching features across wide baselines and over large viewpoint changes is crucial for our approach, however, the employed SIFT algorithm is limited. From the results it appears that pose estimation from E and F is only of theoretical interest, but improved matching algorithms might very well make these methods relevant.
- Preliminary tests with mobile phone images geo-tagged by the device itself exposed several issues requiring further attention. For example, Apple iPhone devices deliver coarsely quantized GPS estimates and the geo-tags of images

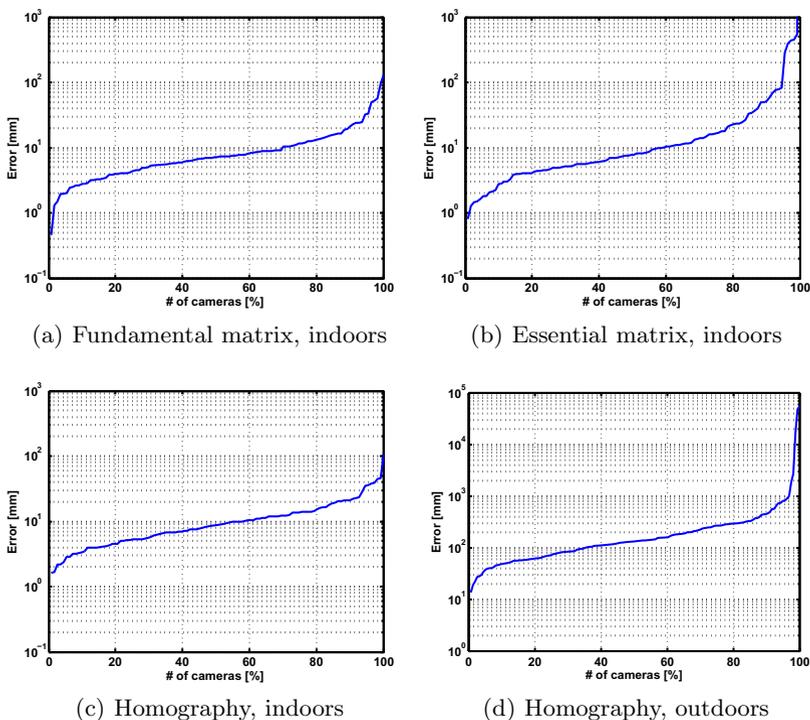


Fig. 5. Plots showing the sorted position errors for all cameras for each of the four experiments. Plot (a-c) show the results for the indoor data set, and (d) for the outdoor data set.

end up on a regular grid of about 13x20 meters (for middle Europe). IOS API gives the GPS info as double floating-point numbers, which are internally filtered to be discrete values with a precision of 4 decimal digits only. This indicates that restrictions for the application of our approach are induced by external factors for the time being. However, a simulation with gaussian noise added to the RTK-GPS data gives evidence that our approach is still applicable for estimates from customer-grade hardware.

- Comparing our approach to SfM based ones is difficult due to the amount of amortization involved. SfM requires some effort first, also storing features for images. This extraction and storage can be done in our approach incrementally with each added query image (becoming a reference image later on). Shared - and comparably expensive - components might be feature matching and geometric validation steps. H/E/F decomposition cost are negligible, and the final P3P might even be faster in our case due to the reduced amount of *correspondences*. We therefore argue that the entire localization approach is comparably efficient to a standard SfM based approach, however, omitting the initial SfM reconstruction effort. We want to stress the fact that our

algorithm tackles typical problems of SfM based solutions, like integration over time or lighting issues.

- Performing localization with a large set of images *avoiding* SfM seems artificial, however, avoiding SfM was one of the main goals in the development of this algorithm. It is of course possible to integrate parts of SfM into our approach. Doing a 'local' SfM reconstruction employing bundle adjustment for comparison reasons is an open point, for example.
- To further improve robustness and accuracy of the pose estimates, a modified algorithm could encounter *confidence* measurements, such as e.g. the number of image matches used for epipolar geometry estimation to reweight individual camera correspondences. Investigating this idea in detail is still an open issue, however.

5 Conclusion

In this work we proposed a novel method for estimating the unknown 6DOF pose of a camera with known intrinsic parameters from epipolar geometry only. Given a set of reference images with image overlap and known position, we estimated epipolar geometry between the image of the query camera and the individual reference images using image features. Extracting the information about the relative positioning of the query camera with respect to each of the reference cameras, we obtain a set of relative pose estimates. Combining this set of pose estimates and the approximate positions of the reference cameras in a robust manner allows us to estimate a full 6DOF pose for the query camera.

Overall, we demonstrated that localization from databases of only geo-referenced images alone is feasible. The effort to build and maintain a 3D reconstruction of the camera poses and environment is avoided as only relative pose information between the query and reference images is used.

There is an inherent tradeoff in that the described method requires more images to be matched to obtain robust pose estimates compared to matching against a recovered 3D structure. However, we believe that this approach is useful for large-scale, dynamic and evolving image databases which can be maintained with less effort than partial or full 3D reconstructions.

Acknowledgements. Thanks to Christian Pirchheim for valuable comments, Manfred Klopschitz for helping with data collection and preparation and to Albert Walzer for his efforts in video editing. This work was partially sponsored by the Christian Doppler Laboratory for Handheld Augmented Reality.

References

1. Arth, C., Klopschitz, M., Reitmayr, G., Schmalstieg, D.: Real-Time Self-Localization from Panoramic Images on Mobile Devices. In: Proc. ISMAR, pp. 37–46 (2011)
2. Irschara, A., Zach, C., Frahm, J.M., Bischof, H.: From Structure-from-Motion Point Clouds to Fast Location Recognition. In: Proc. CVPR, pp. 2599–2606 (2009)

3. Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints. *IJCV* 60, 91–110 (2004)
4. Chen, D., et al.: City-scale Landmark Identification on Mobile Devices. In: Proc. CVPR (2011)
5. Chen, C.Y., Grauman, K.: Clues From the Beaten Path: Location Estimation with Bursty Sequences of Tourist Photos. In: Proc. CVPR, pp. 1569–1576 (2011)
6. Hays, J., Efros, A.A.: Im2gps: Estimating geographic information from a single image. In: Proc. CVPR (2008)
7. Knopp, J., Sivic, J., Pajdla, T.: Avoiding Confusing Features in Place Recognition. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part I. LNCS, vol. 6311, pp. 748–761. Springer, Heidelberg (2010)
8. Schindler, G., Brown, M., Szeliski, R.: City-Scale Location Recognition. In: Proc. CVPR (2007)
9. Zamir, A.R., Shah, M.: Accurate Image Localization Based on Google Maps Street View. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part IV. LNCS, vol. 6314, pp. 255–268. Springer, Heidelberg (2010)
10. Zhang, W., Kosecka, J.: Image Based Localization in Urban Environments. In: Proc. TDPVT, pp. 33–40 (2006)
11. Davison, A.J., Mayol, W.W., Murray, D.W.: Real-Time Localisation and Mapping with Wearable Active Vision. In: Proc. ISMAR, pp. 18–27 (2003)
12. Karlekar, J., et al.: Positioning, Tracking and Mapping for Outdoor Augmentation. In: Proc. ISMAR, pp. 175–184 (2010)
13. Klein, G., Murray, D.: Parallel Tracking and Mapping for Small AR Workspaces. In: Proc. ISMAR, pp. 1–10 (2007)
14. Lothe, P., Bourgeois, S., Royer, E., Dhome, M., Collette, S.N.: Real-Time Vehicle Global Localisation with a Single Camera in Dense Urban Areas: Exploitation of Coarse 3D City Models. In: Proc. CVPR (2010)
15. Mouragnon, E., Lhuillier, M., Dhome, M., Dekeyser, F., Sayd, P.: Real-Time Localization and 3D Reconstruction. In: Proc. CVPR, pp. 363–370 (2006)
16. Hile, H., Grzeszczuk, R., Liu, A., Vedantham, R., Košecka, J., Borriello, G.: Landmark-Based Pedestrian Navigation with Enhanced Spatial Reasoning. In: Tokuda, H., Beigl, M., Friday, A., Brush, A.J.B., Tobe, Y. (eds.) Pervasive 2009. LNCS, vol. 5538, pp. 59–76. Springer, Heidelberg (2009)
17. Takacs, G., et al.: Outdoors Augmented Reality on Mobile Phone using Loxel-based Visual Feature Organization. In: MIR, pp. 427–434 (2008)
18. Hartley, R.I., Zisserman, A.: *Multiple View Geometry in Computer Vision*, 2nd edn. Cambridge University Press (2004) ISBN: 0521540518
19. Shashua, A.: Omni-Rig Sensors: What Can be Done with a Non-Rigid Vision Platform? In: Workshop on Applications of Computer Vision, WACV (1998)
20. Zomet, A., Wolf, L., Shashua, A.: Omni-Rig: Linear Self-recalibration of a Rig with Varying Internal and External Parameters. In: Proc. ICCV, pp. 135–141 (2001)
21. Levi, N., Werman, M.: The Viewing Graph. In: Proc. CVPR., vol. 1, p. 518 (2003)
22. Sato, J.: Recovering Multiple View Geometry from Mutual Projections of Multiple Cameras. *IJCV* 66(2), 123–140 (2006)
23. Ma, Y., Soatto, S., Kosecka, J., Sastry, S.: *An Invitation to 3D Vision*. Springer (2003)
24. Haralick, R.M., Lee, C.N., Ottenberg, K., Nölle, M.: Review and Analysis of Solutions of the Three Point Perspective Pose Estimation Problem. *IJCV* 13, 331–356 (1994)
25. Nistér, D.: A Minimal Solution to the Generalised 3-Point Pose Problem. In: Proc. CVPR, pp. 560–567 (2004)