Revisiting the PnP Problem with a GPS

Timo Pylvänäinen, Lixin Fan, and Vincent Lepetit*

Nokia Research Center

Abstract. This paper revisits the pose estimation from point correspondences problem to properly exploit data provided by a GPS. In practice, the location given by the GPS is only a noisy estimate, and some point correspondences may be erroneous. Our method therefore starts from the GPS location estimate to progressively refine the full pose estimate by hypothesizing correct correspondences. We show how the GPS location estimate and the choice of a first random correspondence dramatically reduce the possibility for a second correspondence, which in turn constrains even more the remaining possible correspondences. This results in an efficient sampling of the solution space. Experimental results on a large 3D scene show that our method outperforms standard approaches and a recent related method [1] in terms of accuracy and robustness.

1 Introduction

The recent development of mobile devices has made applications such as localization using a simple embedded camera realistic on such devices. Research in this direction, however, has mostly focused on image retrieval techniques to consider large-scale environments [2]. Such approach can only provide a coarse pose, and more accuracy will be needed, for example in Augmented Reality applications.

In this paper, we focus on the estimation of an accurate 3D pose from correspondences between 3D points and their projections in the image. This is most certainly one of the oldest problems in Computer Vision, however in this work, we explicitly target pose estimation of photographs taken with a hand held device. In particular, this implies that we can exploit the other sensors these devices are typically equipped with beside the camera.

These sensors include accelerometers, magnetometers (i.e. electronic compasses), and GPS. Accelerometers, or inertial sensors, have been used for several years in tracking applications [3]. They can measure the camera motion, which is of no use in our application[1], and the camera orientation. In practice, however, it is not uncommon to see errors of over 90 degrees in the obtained orientation estimate, and so we choose not to use orientation measurements. Instead, we only consider the GPS.

* Computer Vision Laboratory, École Polytechnique Fédérale de Lausanne (EPFL).

[1] In practice the accuracy of the accelerometers available for mobile devices combined with the erratic motion of such a device together make it virtually impossible to integrate acceleration signal to obtain location.

G. Bebis et al. (Eds.): ISVC 2009, Part I, LNCS 5875, pp. 819–830, 2009.

The standard GPS is typically said to have an accuracy of around few meters. This gives a strong prior on the camera pose, but the uncertainty must still be properly taken into account for an accurate pose estimation. Our method starts from the GPS location estimate to progressively refine the full pose estimate. This is done by sequentially hypothesizing correct correspondences. Because the choices for the previous correspondences dramatically constrain the possibilities of the next correspondence, this allows a particularly efficient sampling of the solution space.

The closest method in the literature is [1], which starts from a prior on the full pose and applies an Extended Kalman filter each time a correspondence is picked to shrink the search space for the next hypotheses. Our experimental results show that our method performs better in terms of robustness and accuracy. This is most likely due to the analytical solution to constrain the search where [1] has to linearize the Kalman filters equations.

In the remainder of the paper, previous work which has used GPS and other sensors is first briefly reviewed. A formalization of the problem is given in Section 3, and Section 4 describes and analyses the sequential sampling procedure. Finally Section 5 compares the proposed method to the standard solution using RANSAC and [1] on a large real 3D scene.

2 Related Work

Carceroni et al. [4] have studied a similar problem of estimating camera orientation from multiple views, given the locations of the viewpoints and a set of point correspondences between views. The uncertainty in the viewpoint locations, however, has not been taken into account. Their method essentially reduces the problem to three degrees of freedom, while we optimize over the full six degrees of freedom of the camera pose.

Some work has been done with inertial sensors and GPS where pose estimation is relevant [3,5,6,7]. Many of these systems do not truly use all the redundancy of the given measurements, but rather use sensors to initialize and help the visual system [3]. When low level sensor fusion is used, it is often done as the prediction part of a Kalman filter such as in [5] and is only useful for tracking applications.

Pollefeys et al. [7] use GPS and inertial tracking for camera pose estimation, and then correspondences for reconstruction. There is therefore no fusion of image and sensor information. GPS and orientation sensor data are used in [6] to compute an initial pose estimate by solving a linear system. This estimate is then refined using a non-linear optimization of an objective function that incorporates a GPS term with an arbitrary weighting factor. Both of these methods require a measurement of the full pose, while we rely only on the GPS and its uncertainty to initialize our method.

We want to be able to deal with outlier correspondences, and RANSAC-like methods have proved their robustness and efficiency. Many variants have been proposed over the years, and we can roughly classify them into four categories:

1. Methods to reduce cost of evaluation to increase the number of iterations that can be done in a given time [8,9].
2. Methods that use prior knowledge to guide sampling [10,11,12].
3. Methods that exploit the connectivity of nearly optimal models [13,14,15,16,17].
4. Methods that modify the evaluation function to get better results [18,19].

Most improvements of RANSAC fall into Categories 1 and 3. In most cases, these improvements can be combined with our method. For instance, our method has been successfully combined with the Hill Climbing strategy [13] to create a combined optimization strategy which performs better than Hill Climbing strategy alone.

The method we propose falls mostly in the second category, taking advantage of very specific prior knowledge. The closest method in this category is perhaps [11] in the sense they use a model specific sampling strategy. They, however, consider a different problem than ours since they focus on homography estimation. They use a different consistency constraint, enforced as a preprocessing step which introduces a fixed overhead.

Like our method, NAPSAC [15] uses the idea of generating sample sets sequentially, but its only assumption was that neighbouring points of an inlier are more likely to be inliers. This method does not use any prior knowledge and falls in Category 3.

Moreno-Noguer [1] also uses sequential sample generation and considers the camera pose estimation problem but requires a prior on the full pose, while we use a prior on the camera center location only. In [1], the correspondences between 3D points and their reprojection are not assumed to be known, and each consecutive point is treated as an observation of the unknown pose in a Kalman filter setting. In contrast, we take a fundamentally different approach as each consecutive point is instead used to reduce degrees of freedom of the pose estimate and the covariances are propagated to represent the uncertainty in the fixed degrees.

However, the algorithm presented in [1] is probably the closest one to ours in that it solves the same problem, uses similar prior knowledge and builds minimal sets sequentially. We therefore compare it against our method in Section 5.

3 Problem Statement

This section gives a formal mathematical definition of the problem we solve. The camera pose estimation is formulated as a minimization problem of a cost function that considers the log likelihood of the pose given the observed correspondence and of the location measurement from the GPS. Because these measurements have meaningful units, the relative weights of the two different kinds of measurements are properly described by the covariance matrices.

We assume we are given a set of world points $\mathbf{X}_i \in \mathbb{R}^3$ and their reprojections $\mathbf{x}_i \in \mathbb{R}^2$ in the image. In practice we use SURF [20] to find these correspondences. The reprojections $\mathbf{x}_i$ are corrupted by noise, and some can even be completely mismatched.

We denote by θ the camera pose and by $\text{proj}_\theta(\mathbf{x}_i)$ the projection of point $\mathbf{X}_i$ by the camera. In other words, for the inliers it is expected that

$$\mathbf{x}_i = \text{proj}_\theta(\mathbf{X}_i) + \epsilon\,, \tag{1}$$

where ϵ is a Gaussian noise term.

The location provided by the GPS is denoted by $\boldsymbol{g}$, and to simplify we assume its true value $\bar{\boldsymbol{g}}$ can be computed directly from the true value $\bar{\theta}$ of the camera parameters as the camera center $\bar{\boldsymbol{g}} = c(\bar{\theta})$.

The problem can finally be stated as recovering the pose θ which minimizes the cost function

$$cost(\theta) = E_c(\theta) + \sum_i \rho(e_i^2)\,, \tag{2}$$

where

$$\begin{aligned} E_c(\theta) &= (\mathbf{g} - c(\theta))^\top \Sigma_c^{-1} (\mathbf{g} - c(\theta)) \\ e_i^2 &= (\mathbf{x}_i - \text{proj}_\theta(\mathbf{X}_i))^\top \Sigma_x^{-1} (\mathbf{x}_i - \text{proj}_\theta(\mathbf{X}_i)) \end{aligned} \tag{3}$$

and $\rho(.)$ is a robust estimator:

$$\rho(e_i^2) = \begin{cases} e_i^2 & e_i^2 < T^2 \\ T^2 & e_i^2 \geq T^2 \end{cases}\,. \tag{4}$$

4 Sequential Sampling

For the pose estimation problem, three correspondences define a pose. In the proposed sequential sampling, this minimal set is generated by selecting each consecutive correspondence from a different distribution, starting from the uniform distribution. Each consecutive correspondence reduces the degrees of freedom of the unknown pose. The probability distribution of the location measurement is mapped to a probability distribution of the correspondences, with the assumption that the first correspondence was correct.

This effectively uses the location measurement to guide the generated minimal sets to be consistent with the location measurement. The proposed poses are evaluated against the robust cost function to find a good inlier set.

4.1 Sampling the First Correspondence

Assuming the rotation of the camera is unknown, then even with known location any single correspondence is a priori equally likely. It is always possible to align any point in 3D to any point in the camera image by rotating the camera. So the first correspondence $\mathbf{x}_1 \leftrightarrow \mathbf{X}_1$ is randomly selected from a uniform distribution.

4.2 Sampling the Second Correspondence

Lets first assume that the camera location and the image location $\mathbf{x}_1$ of the first correspondence are known exactly. The only remaining degree of freedom is the rotation about the axis formed by the camera center and the corresponding 3D point $\mathbf{X}_1$. Under this constraint, the projection $\mathbf{x}_2$ of a given world point $\mathbf{X}_2$ lies on an ellipse in the image plane.

Since the value of the angle defined by $\mathbf{X}_1$, the camera center $\mathbf{g}$, and $\mathbf{X}_2$ should remain the same when expressed in the camera coordinate system and in the world coordinate system, the following formula should hold:

$$\frac{\mathbf{x}_1^\top \mathbf{x}_2}{\|\mathbf{x}_1\|\|\mathbf{x}_2\|} - \frac{(\mathbf{X}_1 - \mathbf{g})^\top (\mathbf{X}_2 - \mathbf{g})}{\|\mathbf{X}_1 - \mathbf{g}\|\|\mathbf{X}_2 - \mathbf{g}\|} = 0\,, \tag{5}$$

where $\mathbf{x}_1$ and $\mathbf{x}_2$ are considered to be in homogeneous coordinates. This equation defines an implicit function constraint of the type:

$$f(\mathbf{x}_2, \mathbf{x}_1, \mathbf{g}) = 0\,. \tag{6}$$

In practice, however, the location $\mathbf{g}$ provided by the GPS and the projection $\mathbf{x}_1$ are corrupted by noise. For a given $\mathbf{x}_2$, there is then some probability that Equation (6) is satisfied. The exact computation of this probability involves integrating over the sets of $\mathbf{x}_1$ and $\mathbf{g}$ for which the constraint holds. This is not computationally feasible, so instead, we linearize f in the neighbourhood of the observations:

$$f(\mathbf{x}_2, \mathbf{x}_1, \mathbf{g}) \approx J_{\mathbf{x}_2}\mathbf{x}_2 + J_{\mathbf{x}_1}\mathbf{x}_1 + J_{\mathbf{g}}\mathbf{g}, \tag{7}$$

where $J_{\mathbf{x}_1}$, $J_{\mathbf{x}_2}$ and $J_{\mathbf{g}}$ are the Jacobians of (6) with respect to to $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{g}$ respectively, and evaluated at the measured points.

Since we assumed that $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{g}$ are normally distributed, the residual $f(\mathbf{x}_2, \mathbf{x}_1, \mathbf{g})$ can now be thought of as a normally distributed random variable:

$$\begin{aligned} f(\mathbf{x}_2, \mathbf{x}_1, \mathbf{g}) &\sim N(0, \sigma_e^2) \\ \sigma_e^2 &= J_{\mathbf{x}_2}\Sigma_{\mathbf{x}}J_{\mathbf{x}_2}^\top + J_{\mathbf{x}_1}\Sigma_{\mathbf{x}}J_{\mathbf{x}_1}^\top + J_{\mathbf{g}}\Sigma_c J_{\mathbf{g}}^\top\,. \end{aligned} \tag{8}$$

Figure 1 illustrates the error introduced by the linear approximation. The ground truth was obtained by randomly sampling the camera center and first correspondence according to the assumed normal distributions. For each sample, the possible exact projections were accumulated.

The likelihood of any point $\mathbf{x}_i$ satisfying the constraint, assuming that $\mathbf{x}_1$ is an inlier corrupted by noise, can now be computed by taking the residual of Equation (6) where $\mathbf{x}_2 = \mathbf{x}_i$ and plugging it into Equation (8). The second point is randomly selected proportional to these likelihoods.

4.3 Sampling the Third Correspondence

Even if a given location and two correspondences constitute an over-constrained pose estimation problem, we found it is still better to consider a third correspondence because of the uncertainty in the location measurement. We give here an

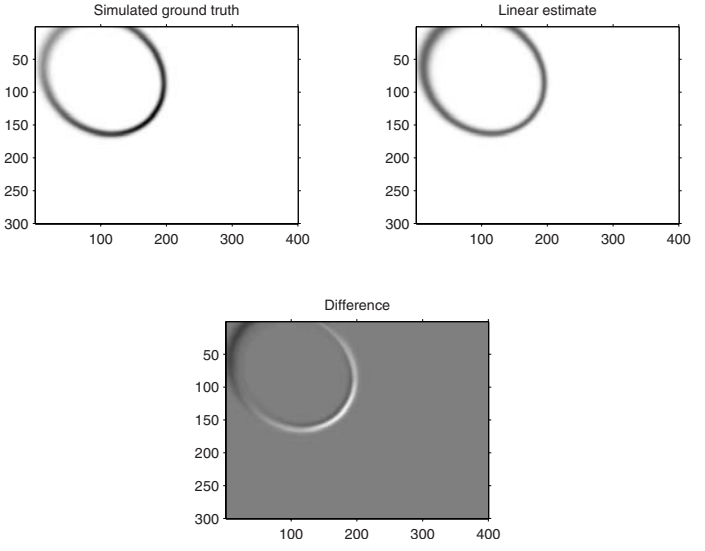

Fig. 1. An example of the linear approximation of the correspondence likelihood. The biggest error happens close to the edges of the image. Light areas of the difference map indicate where the linear approximation under estimates the probability.

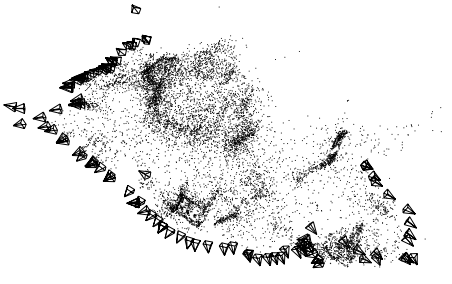

Fig. 2. The scene used in the experiments contains 99 cameras and 10000 world points. The world point visibility in the cameras is based on the original matching information in the original reconstruction.

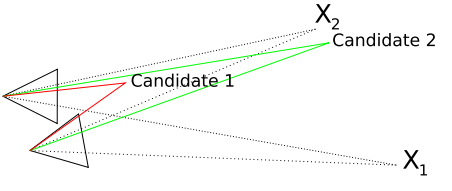

Fig. 3. The sensitivity of the projection of the third point depends on its location relative to the first two points and the camera. Points such as Candidate 1 that are far from the first points are sensitive to camera location. Candidate 2, on the other hand, is close to one of the first two selected points and is less sensitive to camera location.

approximation of the covariance for the projection $\mathbf{x}_3$ of a selected third point $\mathbf{X}_3$ that works well in practice and keeps the computation tractable.

We first assume that the biggest contributor to the variance comes from the camera location uncertainty and all other sources are neglected. We compute a current estimate for the camera pose by taking for the moment the GPS location measurement $\mathbf{g}$ as camera center, and by using the first two correspondences to estimate the rotation $\mathbf{R}$. Intuitively, as shown in Figure 3, when the third point $\mathbf{X}_3$ is close to the two first points $\mathbf{X}_1$ and $\mathbf{X}_2$, the covariance $\Sigma_{\mathbf{x}_3}$ of $\mathbf{x}_3$ will be small. When it is moved away from these points, the covariance will increase. We therefore use the following approximation for $\Sigma_{\mathbf{x}_3}$:

$$\Sigma_{\mathbf{x}_3} = \kappa_1^2 \kappa_2^2 \mathbf{W}^\top \Sigma_c \mathbf{W} \tag{9}$$

where

$$\begin{aligned} \kappa_1 &= \min\left(\frac{\|\mathbf{X}_3 - \mathbf{X}_1\|}{\|\mathbf{X}_1 - \mathbf{g}\|}, \frac{\|\mathbf{X}_3 - \mathbf{X}_2\|}{\|\mathbf{X}_2 - \mathbf{g}\|}\right)^2 \\ \kappa_2 &= \min\left(\frac{1}{\|\mathbf{X}_1 - \mathbf{g}\|}, \frac{1}{\|\mathbf{X}_2 - \mathbf{g}\|}\right)^2 \\ \mathbf{W} &= \mathbf{R}^\top \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}. \end{aligned} \tag{10}$$

The matrix $\mathbf{W}$ maps the covariance of the camera location to the image plane.

We still have to explain how we compute the rotation matrix $\mathbf{R}$ that appears in the third row of (10). The camera maps the 3D points $\mathbf{X}_1$ and $\mathbf{X}_2$ to their respective 2D locations $\mathbf{x}_1$ and $\mathbf{x}_2$ and its rotation $\mathbf{R}$ must satisfy

$$\begin{aligned} \mathbf{R}\hat{\mathbf{X}}_1 &= \hat{\mathbf{x}}_1 \\ \mathbf{R}\hat{\mathbf{X}}_2 &= \hat{\mathbf{x}}_2 \\ \mathbf{R}(\hat{\mathbf{X}}_1 \times \hat{\mathbf{X}}_2) &= \hat{\mathbf{x}}_1 \times \hat{\mathbf{x}}_2\,, \end{aligned} \tag{11}$$

where $\hat{\mathbf{X}}_i = \frac{\mathbf{X}_i - \mathbf{g}}{\|\mathbf{X}_i - \mathbf{g}\|}$ and $\hat{\mathbf{x}}_i = \frac{\mathbf{x}_i}{\|\mathbf{x}_i\|}$. The last equation in (11) comes from the properties of a rotation matrix.

This can be written as a linear system, and solved in the least-squares sense to obtain the elements of the rotation matrix. We force the solution $\mathbf{R}'$ to a proper rotation matrix by computing the singular value decomposition $\mathbf{R}' = \mathbf{U}\mathbf{D}\mathbf{V}^\top$ and then taking $\mathbf{R} = \mathbf{U}\mathbf{V}^\top$.

A summary of our method is given in Algorithm 1. This version has no other stopping criterion than a limited time budget. The function *cost* is the cost function defined in (2). Functions *EllipseLikelihood* and *PointLikelihood* will randomly draw a correspondence index according to the likelihoods defined in Sections 4.2 and 4.3 respectively. Function *PossiblePoses* returns the four solutions of the P3P problem.

Input: A set of N possible correspondences and location information $\mathbf{g}$
Output: Camera pose estimate
while *time left* **do**
  $i_1 := Uniform(1..N)$;
  $i_2 := EllipseLikelihood(\mathbf{g}, i_1, \mathbf{x}_1, ..., \mathbf{x}_N)$;
  $i_3 := PointLikelihood(\mathbf{g}, i_1, i_2, \mathbf{x}_1, ..., \mathbf{x}_N)$;
  $\Theta := PossiblePoses(\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \mathbf{x}_{i_3})$;
  foreach $\theta \in \Theta$ **do**
    if $cost(\theta, \mathbf{x}_1, ..., \mathbf{x}_N, \mathbf{g}) < c_b$ **then**
      $c_b := cost(\theta, \mathbf{x}_1, ..., \mathbf{x}_N, \mathbf{g})$;
      $\theta^* := \theta$;
    end
  end
end

Algorithm 1. Overview of pose estimation

5 Experiments

5.1 Setup

We reconstructed a real 3D scene from 99 images with GPS measurements using our own reconstruction pipeline based on SURF features [20] and the sparse bundle adjustment library described in [21]. The resulting reconstruction contains over 20000 world points. 10000 of the world points were randomly selected for the test set. The resulting scenario is shown in Figure 2.

To analyze the effect of keypoint localization noise and GPS noise, we first corrected the keypoint locations in the images to match the reconstructed world exactly. This created a noise free reconstruction with perfectly known camera poses, which nevertheless represents a real world scenario.

We then added noise to the keypoint locations. Inliers were corrupted with Gaussian noise with a standard deviation of 5 pixels in the original 800×600 pixel images used to capture the scene. With a given probability a keypoint is treated as an outlier, in which case its location is randomly drawn from the uniform distribution over the image. We tests with 4 different outlier ratios from 10% to 70%.

The GPS measurement was generated by adding noise to the camera location in the reconstruction, which after the corrected projections is effectively noiseless. The noise was drawn from the three dimensional Gaussian distribution and normalized to a fixed length. Tests were run with 6 different GPS offsets from 0 to 5 meters. The distribution used for the GPS error in the cost function and in sequential sampling had a standard deviation of 5 meters.

This results in a total of 24 noise scenarios. Three methods were tested under these conditions:

1. Standard RANSAC to optimize the compound cost function.
2. The pose prior method of [1] modified as described below.
3. The proposed sequential sampling method.

The original method described in [1] does not assume known correspondences and performs an exhaustive search. It was adapted to the Random Sampling

Consensus framework as follows. Starting from the given pose prior, the pose estimate and its covariance are updated according to the Kalman filter rules as new points are selected. The selection of the next point is done according to the likelihoods defined by the reprojection error obtained using the current pose estimate. The reprojection error covariance is obtained by propagating the pose estimate covariance using the Jacobian of the projection function.

The pose prior method requires a full pose. To be fair, it was tested with realistic noise which corresponds to the kind of orientation estimate that might be obtained from sensors embedded in a mobile device. The rotation matrix was corrupted by random noise until a rotation matrix was obtained where the mean angle between the axis of the original rotation matrix and the corrupted rotation matrix was between 20 and 25 degrees.

A sample run consists of 10 iterations of the algorithm to estimate the pose of a randomly selected image from the set. A test comprises of 1000 sample runs of an algorithm under specific noise conditions.

5.2 Results

The three histograms of the left of Figure 4 show the results obtained with the three different methods for various GPS errors when there are only 10% outliers. In that case, the three methods perform about the same.

The proposed GPS method, however, is most valuable when the images are difficult to match. In practice, urban scenes have many repetitive structures, for example the windows of a building, and to guarantee the existence of the real correspondence, multiple hypotheses from feature matching should be retained,

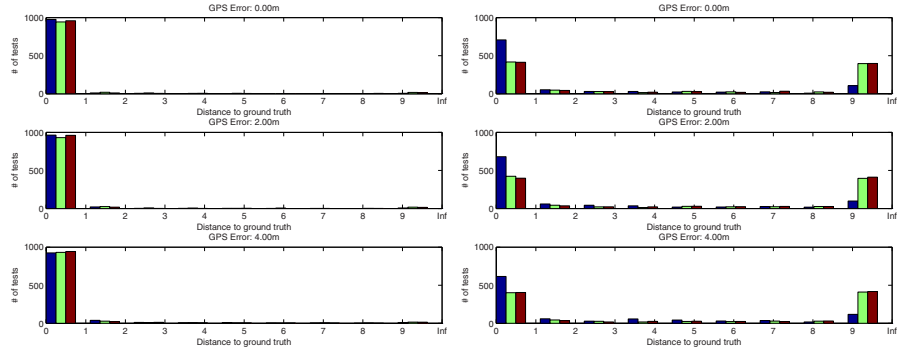

Fig. 4. Comparison with RANSAC and the modified pose prior method of [1] for two different ratios of outliers, and three levels of GPS errors. The bars in the stacks correspond, from left to right, to the proposed method (blue), the modified method of [1] (green), and RANSAC (red). **Left:** For small ratios of outliers, the three methods perform about the same. **Right:** For large ratios of outliers, which correspond to more realistic scenarios, our method is clearly more accurate.

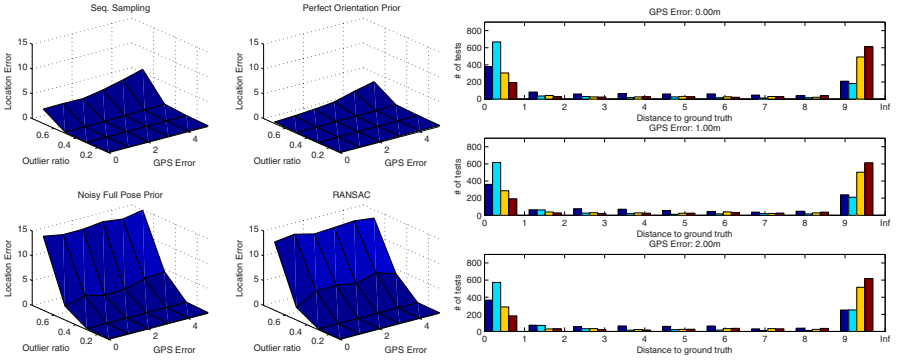

Fig. 5. Left: Median distance to ground truth location for different noise conditions and for all the methods. In this test we also included, for comparison, the modified pose prior method of [1] with no error in the orientation. It can be seen that RANSAC and pose prior with realistic noise in the pose both tend to break down after 50% outliers. Although our method shows degraded performance as the GPS error approaches values unlikely according to the assumed covariance, it still outperforms previous methods. Of course, the full pose prior method with perfect orientation performs very well as with perfect orientation measurement there is little to optimize. **Right:** The histograms of distance to ground truth location for 70% outliers for the proposed method and the full pose prior with different levels of noise in the orientation. The bars in the stacks are from left to right: the proposed method (blue), full pose prior with maximum 5 (cyan), 5-10 (yellow) and 10-15 (red) degrees of error. As is to be expected, with no error in the location and very little error in the orientation, the full pose prior method works very well. When the GPS error is increased, the performance of the proposed method approaches that of the method with nearly perfect orientation. The full pose prior with noisy orientation does not perform as well as the proposed method. It can be observed that if the orientation measurement is available, it must have error less than 5 degrees before the switch to full pose prior method is justified.

resulting in a large number of outliers. As shown by the histograms on the right of Figure 4, our method efficiently takes advantage of the GPS data, and outperforms the other methods.

Figure 5 (left) summarizes the results of more experiments. We also include the results obtained with the modified method of [1] provided by the exact orientation. This version performs remarkably well, unfortunately currently no sensors are able to provide such accuracy on the orientation. It can be seen that RANSAC and pose prior with realistic noise in the pose both tend to break down after 50% outliers.

Finally, we tested our method against the full pose prior method with different levels of noise in the orientation and location measurements. The test case contained 70% outliers and is based on 1000 sample runs. The results are shown in Figure 5 (right). We tested against three different orientation noise cases: where the average angluar error to the axes of the ground truth rotation was less than 5 degrees, when it was between 5 and 10 degrees and between 10 and

15 degrees. Obviously, with nearly perfect prior pose information the full pose prior method performs extremely well. It can be observed, however, that as the GPS error is increased the performance difference becomes smaller.

The results show that if the orientation measurement is available, it has to have an error less than 5 degrees for it to be useful. The proposed method, which uses only GPS, outperforms the full pose prior method with 5-10 degrees of error in the orientation.

6 Conclusion

We showed how GPS information can be used to guide sampling in a RANSAC setting to estimate inliers of the pose estimation problem. This novel sequential sampling method was shown to effectively guide the sampling towards the correct solution.

In the experiments, the method shows clear performance advantage when the number of outliers is high. In real world applications, extremely high outlier ratios commonly occur when multiple hypotheses from feature matching are retained. In the case of repeated patterns, multiple hypotheses can lead to multiple consensus sets only one of which represents the correct pose. The use of GPS effectively resolves this ambiguity and the proposed method does this efficiently.

It should be noted, however, that the evaluation of the likelihoods for each candidate match in steps 2 and 3 is roughly equivalent to one evaluation of the objective function. One iteration of the proposed algorithm in a naive implementation therefore equals to roughly three iterations of standard RANSAC in terms of CPU time. This means, unfortunately, that in practice it is usually faster to not apply the GPS based weighting on the candidate matches unless the outlier ratio is very high. It might be possible to develop more advanced selection strategies which would avoid full evaluation of the likelihoods for each point.

References

1. Moreno-Noguer, F., Lepetit, V., Fua, P.: Pose priors for simultaneously solving alignment and correspondence. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part II. LNCS, vol. 5303, pp. 405–418. Springer, Heidelberg (2008)
2. Takacs, G., Chandrasekhar, V., Gelfand, N., Xiong, Y., Chen, W.C., Bismpigiannis, T., Grzeszczuk, R., Pulli, K., Girod, B.: Outdoors augmented reality on mobile phone using loxel-based visual feature organization. In: ACM international conference on Multimedia information retrieval, pp. 427–434. ACM, New York (2008)
3. Klein, G., Drummond, T.: Tightly integrated sensor fusion for robust visual tracking. Image and Vision Computing 22 (2004)
4. Carceroni, R., Kumar, A., Daniilidis, K.: Structure from motion with known camera positions. In: CVPR, pp. 477–484 (2006)
5. You, S., Neumann, U.: Fusion of vision and gyro tracking for robust augmented reality registration. In: IEEE Conference on Virtual Reality, pp. 71–78 (2001)

6. Ng, T.K., Kanade, T.: PALM: portable sensor-augmented vision system for large-scene modeling. In: 3-D Digital Imaging and Modeling, pp. 473–482 (1999)
7. Pollefeys, M., Nistèr, D., Frahm, J.M., Akbarzadeh, A., Mordohai, P., Clipp, B., Engels, C., Gallup, D., Kim, S.J., Merrell, P., Sinha, S., Talton, B., Wang, L., Yang, Q., Stewènius, H., Yang, R., Weclh, G., Towles, H.: Detailed real-time urban 3d reconstruction from video. International Journal of Computer Vision (2007)
8. Nister, D.: Preemptive RANSAC for live structure and motion estimation. In: ICCV, vol. 1, pp. 199–206 (2003)
9. Chum, O., Matas, J.: Randomized RANSAC with Td,d test. In: BMVC, pp. 448–457 (2002)
10. Tordoff, B., Murray, D.W.: Guided sampling and consensus for motion estimation. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002. LNCS, vol. 2350, pp. 82–96. Springer, Heidelberg (2002)
11. Guo, F., Aggarwal, G., Shafique, K., Cao, X., Rasheed, Z., Haering, N.: An efficient data driven algorithm for multi-sensor alignment. In: Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications, ECCV (2008)
12. Chum, O., Matas, J.: Matching with PROSAC - progressive sample consensus. In: CVPR, pp. 220–226 (2005)
13. Pylvänäinen, T., Fan, L.: Hill climbing method for random sample consensus methods. In: International Symposium on Visual Computing (2007)
14. Chum, O., Matas, J., Kittler, J.: Locally optimized RANSAC. In: Porceedings of the 25th DAGM Symposium Pattern Recognition, pp. 236–243 (2003)
15. Myatt, D., Bishop, J., Craddock, R., Nasuto, S., Torr, P.H.: NAPSAC: High noise, high dimensional robust estimation — it's in the bag. In: BMVC, pp. 458–467 (2002)
16. Tordoff, B.J., Murray, D.W.: Guided-MLESAC: Faster image transform estimation by using matching priors. IEEE Transactions on Pattern Analysis and Machine Intelligence 27, 1523–1535 (2005)
17. Raguram, R., Frahm, J.M., Pollefeys, M.: A comparative analysis of ransac techniques leading to adaptive real-time random sample consensus. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part II. LNCS, vol. 5303, pp. 500–513. Springer, Heidelberg (2008)
18. Torr, P.H., Zisserman, A.: MLESAC: A new robust estimator with application to estimating image geometry. Computer Vision and Image Understanding 78, 138–156 (2000)
19. Torr, P.H., Davidson, C.: IMPSAC: Synthesis of importance sampling and random sample consensus. IEEE Transactions on Pattern Analysis and Machine Intelligence 25, 354–364 (2003)
20. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: SURF: Speeded up robust features. In: Computer Vision and Image Understanding (2008)
21. Louarkis, M., Argyros, A.: The design and implementation of a generic sparse bundle adjustment software package based on the Levenberg-Marquardt algorithm. Technical Report 340, Computer Science-FORTH, Heraklion, Crete, Greece (2004)