

# Robust Epipolar Geometry Estimation Using Noisy Pose Priors

Yehonatan Goldman<sup>a</sup>, Ehud Rivlin<sup>a</sup>, Ilan Shimshoni<sup>b,\*</sup>

<sup>a</sup>Technion - Israel Institute of Technology, Technion City, Haifa 3200003, Israel

<sup>b</sup>University of Haifa, 199 Aba Khoushy Ave., Mount Carmel, Haifa, Israel

---

## Abstract

Epipolar geometry estimation is fundamental to many computer vision algorithms. It has therefore attracted a lot of interest in recent years, yielding high quality estimation algorithms for wide baseline image pairs. Currently many types of cameras such as smartphones produce geo-tagged images containing pose and internal calibration data. These include a GPS receiver, which estimates the position, a compass, accelerometers, and gyros, which estimate the orientation, and the focal length. Exploiting this information as part of an epipolar geometry estimation algorithm may be useful but not trivial, since the pose measurement may be quite noisy. We introduce SOREPP (Soft Optimization method for Robust Estimation based on Pose Priors), a novel estimation algorithm designed to exploit pose priors naturally. It sparsely samples the pose space around the measured pose and for a few promising candidates applies a robust optimization procedure. It uses all the putative correspondences simultaneously, even though many of them are outliers, yielding a very efficient algorithm whose runtime is independent of the inlier fraction. SOREPP was extensively tested on synthetic data and on hundreds of real image pairs taken by smartphones. Its ability to handle challenging scenarios with extremely low inlier fractions of less than 10% was demonstrated. It outperforms current state-of-the-art algorithms that do not use pose priors as well as others that do.

## Keywords:

robust estimation, epipolar geometry, pose priors

---

## 1. INTRODUCTION

Epipolar geometry is the intrinsic projective geometry between two views, and it is encoded in the *fundamental matrix*  $F$  [1]. Its estimation is one of the core problems in computer vision and is used as a basic component for stereo matching [2], structure from motion (SfM) [3], vision-based robot navigation [4], and other applications. The epipolar geometry of two images is usually estimated by finding corresponding features in both. This is done first by detecting and matching invariant features using an algorithm such as SIFT [5], followed by the application of a robust estimation method from the RANSAC [6] family.

The main weakness of RANSAC is the necessity to sample a valid set. As the inlier fraction decreases, the probability to sample a valid set drops rapidly, increasing greatly the required number of iterations.

In recent years considerable progress has been made in developing estimation algorithms that tackle these problems. Such algorithms include LO-RANSAC (Local-Optimization RANSAC) [7], PROSAC (Progressive Sample Consensus) [8], MLESAC (Maximum Likelihood Estimation RANSAC) [9], BEEM (Balanced Exploration Exploitation Model Search) [10], BLOGS (Balanced LOcal and Global Search) [11], and recently USAC (Universal RANSAC) [12]. However, scenarios with wide baseline images or small overlapping regions between the images still challenge even the current state-of-the-art algorithms due to the low inlier fractions. Figure 1 shows several such challenging image pairs.

The *fundamental matrix*  $F$  is constructed from the camera parameters only, independent of the scene. The intrinsic parameters are mainly the focal lengths of the cameras and their principal points, and the extrinsic parameters which describe their relative pose (camera position and orientation). Focal length data already appears in the metadata of standard cameras under the EXIF format. Thus, the internal calibration matrix  $K$

---

\*Corresponding author

Email address: ishimshoni@is.haifa.ac.il  
(Ilan Shimshoni)



Figure 1: Wide baseline images for which the epipolar geometry could not be estimated using current algorithms but was successfully estimated by SOREPP. Left: *Open* set. Right: *Urban* set. Top: The reference image with several control points marked. Bottom: The target image after matching with the control points and their corresponding estimated epipolar lines.

can be estimated using the focal length and by estimating the principal point as the center of the image. Image distortions even though exist are ignored. Under these assumptions algorithms that use this data exist [3, 13, 14].

In this case, the fundamental matrix is simplified to the *essential matrix*  $E$ . Nowadays the pose data is also becoming available as smartphones with built-in sensors such as a compass, accelerometers, gyros and a GPS receiver become popular. Using these sensors, whose values can be readily measured, the camera pose can be estimated. There exist applications that place in the EXIF image header the position and orientation of the camera [15, 16]. Thus, it is very easy to obtain this information.

In principle, knowing the pose and the internal calibration parameters would make vision-based epipolar geometry estimation methods unnecessary, since the fundamental matrix is computed from these known quantities. But this is not the case in practice since the accuracy of these sensors is low. For example, typical compass azimuth errors can exceed  $7^\circ$ . Therefore

epipolar geometry estimation contaminated with such errors can be considered as a search problem in the parametric space of the relative pose. Yet, the noisy pose data is valuable and can be used in order to constrain the search and focus it.

This paper tackles the following problem: how to effectively exploit pose priors for epipolar geometry estimation, while taking into account the fact that the pose priors are noisy or even partly incorrect. The paper focuses on images taken by Smartphones. The images may differ greatly from one another, leading to low inlier fractions.

To handle these difficulties, we propose a novel estimation algorithm called SOREPP (Soft Optimization method for Robust Estimation based on Pose Priors). SOREPP optimizes an M-estimator cost function designed for robustness to putative outlier correspondences, while the pose priors are used to initialize the optimization and regularize the solution. Knowledge of the expected amount of noise in the pose parameters is used to limit the search to a part of the parameter space. Because, in contrast to RANSAC, explicit detection of

inliers is not required, a solution can be found even in extreme conditions, typically where the inlier fraction is around 10%. The algorithm is simple yet effective. It exhibits fast runtime and does not depend on the inlier fractions, making it attractive for real-time applications. SOREPP achieves a notable improvement over the other methods both in runtime and accuracy for very challenging image pairs, such as those in Figure 1. It could be used for example as part of a Structure from Motion (SfM) application in which a large number of images of a scene are taken and in the first part of the algorithmic pipeline are matched. Using SOREPP many more image pairs will be matched correctly and more efficiently. As a result, the recovered structure and motion are much more accurate. The implementation of SOREPP and a Matlab mex wrapper are available at [17].

The paper has three main contributions. First, it addresses the problem of epipolar geometry estimation when pose prior data is available. Unlike previous works we take into account all available pose information and especially the expected noise level based on a physical model. In the paper we apply the method to smartphone based sensors but it can be easily generalized to other sensor types. The second contribution is the expansion of the robust optimization framework that is usually used to refine a solution, enabling it to be directly applied to the raw data with low inlier fractions and high levels of pose uncertainty. As far as we know this is the first time that this approach has been taken in low inlier fraction cases, instead of RANSAC. Finally, we introduce an extensive experimental setup for testing the algorithm and comparing it to the state-of-the-art. With a relatively low amount of work on single images, a large number of image pairs are generated. This enables us to test and compare different algorithms on hundreds of image pairs.

The rest of the paper is organized as follows. Section 2 reviews the main methods for epipolar geometry estimation, with or without pose prior data. In Section 3 we give details about how pose is measured by smartphones and review epipolar geometry. Section 4 introduces SOREPP, our proposed robust algorithm. Experimental results on synthetic data are described in Section 5, while results on real image pairs are presented in Section 6. Section 7 concludes the paper.

## 2. RELATED WORK

### 2.1. Image Based Only Methods

Most robust algorithms are RANSAC based. PROSAC [8] uses priors on the quality of the putative correspondences in order to guide the sampling

and increase the chance to sample a valid set. LO-RANSAC [7] handles the problem that even an outlier-free set may lead to inaccurate estimation, by locally improving the estimated model when the best model so far is found. Recently the USAC [12] was introduced. It integrates PROSAC and LO-RANSAC together with SPRT (Sequential Probability Ratio Test) accelerated hypotheses test. This way an incorrect RANSAC hypothesis can be discarded without considering all the putative correspondences. BEEM [10] uses minimal sets of only two correspondences, by creating three synthetic points around correspondences in the two images according to their scale and orientation, estimated by the SIFT descriptor, and assuming that they match. Thus, each match yields four matches. This decreases considerably the number of RANSAC iterations that are required for epipolar geometry estimation. BLOGS [11] seeks to minimize a soft threshold cost function, and combines global search that is based on Joint Feature Distribution, and local search in order to further improve the current estimation.

The a-contrario method evaluates probabilistically whether a suggested model is true. This method can be used as part of a RANSAC algorithm to evaluate suggested solutions. The method was used successfully for estimating epipolar geometry in [18]. It was extended for repeated structures [19] and scenes with multiple moving objects in [20].

Another category of algorithms are the iterative methods [21], which search for the solution in the parameter space, usually searching in the fundamental matrix space. Methods that work in the relative pose domain also exist [14, 22]. In [14] a branch and bound strategy is used in order to find the optimal solution. The main drawback of this method is its computational cost, making it impractical for real time applications. In [22] an entropy-based cost function is used to achieve robustness to outliers. This method is built for vision-based robot navigation, where the images being matched are successive video frames. Thus, they are similar and yield a high fraction of inliers.

### 2.2. RANSAC Methods that Exploit Pose Priors

Much work has been done on exploiting pose priors, in particular in the robotic and navigation communities. Some papers integrate vision and an inertial measurement unit (IMU) in order to reduce drift [23], to choose images [24], or to constrain a structure-from-motion solution [25]. However, not much work has been done regarding exploiting pose prior data in order to enhance the epipolar geometry estimation itself. The existing methods assume that an IMU is attached to the camera.

The method described in [26] uses IMU measurements in order to limit the correspondence search area to be only at a user defined distance around the corresponding epipolar line. This accelerates the computation and reduces the number of incorrect matches. The method works well in low pose noise scenarios. However, as the noise level increases, so does the width of the region around the epipolar line where the correct correspondence lies, covering large portions of the image. This makes filtering impractical in noisy conditions.

Other methods [27, 28] suggest using the known gravity vector, which is equivalent to two rotation angles, reducing the degrees of freedom of the solution from five unknowns to only three. RANSAC is applied with a correspondence set of size three. A similar method which assumes that the full rotation is known and uses a minimal set of two correspondences is described in [29]. The main limitation of these methods is that they take the values of the measured orientation angles as hard constraints, and are therefore sensitive to noise in them. Moreover, any existing position priors are ignored.

### 2.3. Bundle Adjustment

Bundle adjustment (BA) is the gold standard method for estimating camera poses and the scene from correspondences. It is based on optimizing a cost function that is built from reprojection errors. Bundle adjustment has many variants, the important ones in the context of the paper are variants that apply robustified error models, use epipolar geometry-based distances instead of reprojection errors, “structure-less” variants which estimate only the camera poses, and enforcement of prior constraints [30, 31, 32]. BA is commonly used as a final step, after most of the outliers have been removed and good values for the camera poses have to be given to initialize the process [1, 3]. The core of SOREPP may be considered as a BA-variant, designed solely for two images, which is structure-less, based on a robustified cost function of epipolar constraints, applies pose prior regularization, and initialized by it. However, the general framework of SOREPP expands the BA framework and significantly improves the performance relative to the core optimization scheme. This, together with the robustness of the function to be optimized, enables to apply SOREPP on the raw input, without removing first the outliers or improving the noisy pose priors.

## 3. Preliminaries

### 3.1. Pose Measurements in Smartphones

The pose of a platform is composed of its orientation and position. We use East-North-Up (ENU) coordinates to define the position. The orientation is described uniquely by a rotation matrix that can be decomposed to three Euler angles. A typical smartphone has several sensors used to determine these six pose parameters with the following qualitative accuracy:

**The Position** is usually calculated using a GPS receiver. Its typical accuracy under good reception conditions can be accurate to within a few meters. Under difficult GPS signal visibility conditions, such as in a dense urban environment, the errors may become significantly larger. In the paper we ignore any possible dependency between the position noise of the two cameras and assume that any such noise is independent.

**The Leveling Angles**, also known as pitch and roll angles, define the platform’s orientation relative to the tangent plane of the earth at its current position. They can be easily calculated using accelerometers measuring the gravity vector. The force of gravity ensures quite accurate results in the order of  $1^\circ$  even for cheap, insensitive accelerometers such as the ones used in smartphones.

**The Azimuth** is currently the most challenging parameter to measure. Azimuth is usually measured by means of a compass which detects the direction to the magnetic north, but this method is quite inaccurate. The magnetic field itself is affected by high-voltage electric wires and metals in its vicinity. In our experiments we experienced errors exceeding  $7^\circ$ .

We used SitisMobile’s “GeoCam” application [15] in order to record images with the attached pose. The pose information is recorded in the EXIF image header. Similar applications such as “GeoCam Free” from Wazer [16] also exist.

### 3.2. Epipolar Geometry

This section is based on [1] and reviews the main concepts used in our work. The epipolar constraint is formulated by the following equation:

$$\tilde{p}_2^T E(s) \tilde{p}_1 = 0, \quad (1)$$

where  $E$  is the essential matrix which is a function of the relative pose  $s$  and  $\tilde{p}_1$ ,  $\tilde{p}_2$  are calibrated corresponding points in images 1 and 2, respectively.  $E$  is constructed from the relative pose of the cameras as follows:

$$E = R_r [t_r]_{\times}, \quad (2)$$

where  $R_r$  is the relative rotation matrix,  $t_r$  is the translation vector, and the symbol  $[x]_\times$  represents the cross-product matrix. When the pose is measured independently for each camera, we define  $R_1, R_2$  and  $t_1, t_2$  as the pose of the cameras in a global coordinate system. The relative pose is calculated as follows:

$$R_r = R_2 R_1^T, \quad t_r = R_1(t_2 - t_1). \quad (3)$$

The Euler angles composing  $R_r$  are the azimuth  $\psi_r$ , the pitch  $\theta_r$ , and the roll  $\phi_r$ . The representation of the Euler angles has in general singular points when  $\theta_r = \pm 90^\circ$ . Since the relative pitch of two cameras that view the same scene is mostly smaller, they are safe to use, as indicated in [30]. We describe the relative translation vector using polar coordinates, defining  $\alpha$  to be the relative horizontal angle and  $\beta$  to be the relative vertical angle:

$$\alpha = \arctan 2(t_r(x), t_r(y)), \quad \beta = \arcsin\left(\frac{t_r(z)}{\|t_r\|}\right). \quad (4)$$

We define the vector of unknowns  $s$  to be composed of these five angles, ignoring the distance between the cameras, which is not represented by the epipolar geometry:

$$s \triangleq (\psi_r \quad \theta_r \quad \phi_r \quad \alpha \quad \beta)^T. \quad (5)$$

The input vector  $s_0$  is built from the prior values. Note that the relative pose is asymmetric. Therefore, matching image 2 to image 1 is not identical to matching image 1 to image 2.

The covariance matrix of  $s_0$  has a central role in SOREPP. We assume that the  $6 \times 6$  global pose covariance matrices of each camera are known and denote them as  $\Sigma_1$  and  $\Sigma_2$ . One can approximate these matrices by creating diagonal matrices from the noise approximated in Section 3.1. Under the independent noise assumption, the  $5 \times 5$  relative pose covariance matrix  $\Sigma_{s_0}$  can be linearly approximated from the cameras' covariance matrices:

$$\Sigma_{s_0} \approx J_1 \Sigma_1 J_1^T + J_2 \Sigma_2 J_2^T, \quad (6)$$

where  $J_1$  and  $J_2$  are the Jacobians, composed of the derivatives of the relative pose angles by each camera's pose parameters. These Jacobians are computed by taking the derivatives of (3) (4) and the standard function which computes the Euler angles from a rotation matrix.

Most epipolar geometry estimation algorithms seek to reduce the distances between points and their corresponding epipolar lines. A common distance function is

the Sampson distance [1], which approximates the geometric distance. The signed Sampson distance of the  $k$ -th correspondence is defined as follows:

$$d(k, s) = \frac{\tilde{p}_2(k)^T E(s) \tilde{p}_1(k)}{\sqrt{l_{1,x}(k)^2 + l_{1,y}(k)^2 + l_{2,x}(k)^2 + l_{2,y}(k)^2}}, \quad (7)$$

where  $l_{1,x}, l_{1,y}$  and  $l_{2,x}, l_{2,y}$  are the first two components of the corresponding epipolar lines.

#### 4. OUR ALGORITHM

We will now introduce our proposed algorithm, SOREPP. SOREPP uses pose priors to focus the search in the relevant parameter space by initializing a standard optimization scheme. Thus, global search is avoided. The cost function being optimized is a differential and robust version of the standard, number-of-inliers cost function used by RANSAC. It works on all the putative correspondences, without detecting the inliers explicitly, which might be difficult in low inlier fraction cases. As a result, SOREPP is able to handle low inlier fractions. To do so, we have to define a robust cost function that can handle low inlier fractions while having a large basin of attraction to handle large pose errors.

SOREPP is given as input an estimate of the relative pose  $s_0$  of the two cameras, and its uncertainty is described by a covariance matrix  $\Sigma_{s_0}$ . A set of putative correspondences  $\Omega_{all}$  is given, together with the weight of each correspondence  $w(\cdot)$ , which is a rough approximation of the prior probability of being an inlier. Two of the many options that might be used to define such weights are described in Section 4.1.

We define the following Gaussian score for every putative correspondence  $k$ :

$$g(k, s) = \exp\left(\frac{-d(k, s)^2}{2\sigma_h^2}\right). \quad (8)$$

This function is a type of an M-estimator. It applies a soft threshold with a parameter  $\sigma_h$  on the Sampson distance associated with each correspondence. For the true relative pose  $s_{true}$ , inliers would receive high scores, i.e., close to 1, while outliers far away from the model would receive low scores, i.e., close to 0. These scores determine the preliminary target function that SOREPP will seek to minimize:

$$v = \sum_{k \in \Omega_{all}} \tilde{w}(k) (1 - g(k, s)), \quad (9)$$

where  $\tilde{w}(k)$  are the weights  $w(k)$  after they have been normalized:

$$\tilde{w}(k) = \frac{w(k)}{\sum_{l \in \Omega_{all}} w(l)}, \quad (10)$$

and  $v$  is the score of a specific relative pose  $s$  over all the putative correspondences and is also in the range  $[0, 1]$ . This score function has two important characteristics. Due to the differentiability of the function, inliers affect the score even when the current estimate of  $s$  is far from the correct one and thus can help the optimization process to converge to the correct solution. In addition, the maximal penalty of an outlier pair is at most 1, enabling the optimization process to converge even when a large number of outliers exist in the data. Still, due to the non-convexity of (9), optimizing this function may be difficult in noisy conditions. In order to regularize the solution, we define a measure of the proximity of the solution to the initial pose by the following Mahalanobis distance:

$$\lambda(s) = \frac{1}{|s|} \sqrt{\delta s^T \Sigma_{s_0}^{-1} \delta s}, \quad (11)$$

where  $|s|$  is the size of  $s$  and  $\delta s \triangleq s - s_0$ . We add this regularization term to (9) in order to define the following minimization problem:

$$\hat{s} = \arg \min_s \left\{ c \left( \sum_{k \in \Omega_{all}} \tilde{w}(k) (1 - g(k, s)) \right) + (\lambda(s))^2 \right\}, \quad (12)$$

where  $c$  is a weighting parameter, which is fixed to the value of 5 in the experiments. This equation combines the objective to minimize the Sampson distances for as many putative correspondences as possible while keeping the solution close to the pose prior.

As the function to be optimized is not convex, it will fail to converge if the solution is beyond the function's basin of attraction. To deal with this problem the optimization can be performed starting from several initial conditions as will be explained in Section 4.2.

The minimization can be done using any standard optimization method. In our implementation we use Levenberg-Marquardt [33]. Once the algorithm converges, a hard threshold is applied to select the inliers. They serve only as an output of the algorithm.

This simple optimization procedure is the heart of SOREPP algorithm. The algorithm is simple and fast, as well as robust to low inlier fractions and significant pose noise. In the next subsections we will delve deeper into the algorithm, elaborating on its various components.

#### 4.1. Correspondence Weights

The weights  $w(k)$  are one of the inputs to SOREPP. SOREPP can use any probability-like values. Our basic weights are built from the following simple formula

based on the SIFT ratio test [5]:

$$w(k) = 1 - \frac{d_{1NN}^2(k)}{d_{2NN}^2(k)}, \quad (13)$$

where  $d_{1NN}(k)$ ,  $d_{2NN}(k)$  are the distances from the first and second nearest neighbors of the  $k$ -th correspondence.

A different way to define putative correspondences and weights is to follow BLOGS [11]. The similarity is calculated by the inner product of the descriptor vectors. Putative correspondences are the feature pairs that get the highest similarity score in both directions of matching: a feature from image 1 is compared to all features in image 2, and vice versa. The weights are calculated by the following formula [11]:

$$w(k) = \left(1 - \exp^{-q(k)}\right)^2 \left(1 - \frac{q_r(k)}{q(k)}\right) \left(1 - \frac{q_c(k)}{q(k)}\right), \quad (14)$$

where  $q(k)$  is the highest score for the  $k$ -th correspondence,  $q_r(k)$  is the second highest score in one direction of matching and  $q_c(k)$  is the second highest score in the other direction. The first term of (14) is the similarity based component.

#### 4.2. Pose Region Search

In order to further increase the robustness to noisy pose measurements, the parameter space around the input pose priors is searched. This is done by evaluating the energy of (9) at a sample pattern, where each relative pose parameter is sampled at several values around the prior value. The result is a small five-dimensional hypercube with a target function value at each entry. In that hypercube local peaks are found, creating a set of initial candidate guesses for the optimization scheme. The optimization process of (12) is performed for up to the  $m$  best candidates. The user defined parameter  $m$  controls the number of candidates as well as the input pose and is usually in the range of 0 to 5, where  $m = 0$  means starting only at the pose prior  $s_0$ . In all these cases,  $\delta s$  is always defined relative to  $s_0$ . This component is illustrated in Figure 2.

The samples are spread uniformly over the space determined by  $2\sigma_{s_0}$  for every angle. The samples include only 3 values for  $\theta_r$ ,  $\phi_r$ , and  $\beta$ , 5 values for  $\alpha$ , and, 7 values for  $\psi_r$ . The number of samples per parameter were set according to its variance since some parameters have a larger uncertainty than others.

Even though this sampling method is very simple and sparse, it yields good results, because one of the samples usually falls into the large basin of attraction of (12).

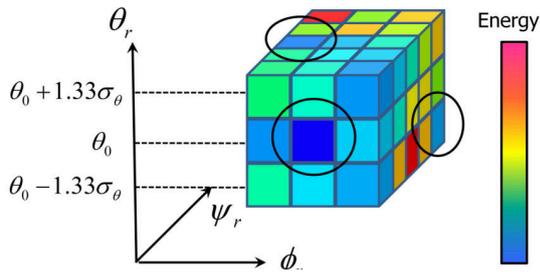


Figure 2: Illustration of the *Pose Region Search* component. The energy is calculated on a sparse grid based on the input pose and the expected noise levels. For a few local minima (shown in black circles) the optimization process is applied. In the figure, only three of the five relative pose parameters are shown.

Note that the computational cost is low. The cost function is evaluated several hundred times, but the optimization scheme is applied only up to  $m + 1$  times. In Section 6.2 the effect on the quality of the results on choosing different values of  $m$  is tested.

#### 4.3. Solution Refinement

One of the main characteristics of SOREPP is that all the putative correspondences contribute to the solution. The main advantage of it is the ability to bypass the need to explicitly find inliers, which may be difficult when inlier fractions are low. However, this characteristic is also a main drawback: outliers also participate in the estimation and skew it. The most harmful outliers are those which randomly obtain small Sampson distances with respect to the current solution. They are usually further than the inliers, but close enough to have significant weights  $g$ .

To deal with this problem we modify  $\sigma_h$  by the following procedure, common for M-estimators. For a solution with energy lower than  $v_{threshold}$  of (9),  $\sigma_h$  is decreased and the solution is estimated again by optimizing (12), starting from the current estimation result  $\hat{s}$ . This way the weight of the correspondences that are close to the solution increases at the expense of the outliers, which are assumed to be further from the solution. We use the following values:  $v_{threshold} = 0.65$  and reduce  $\sigma_h$  by a factor of two. In real experiments the value of the energy is usually very low or very high. Therefore the performance of the algorithm does not depend on the exact value of  $v_{threshold}$ .

#### 4.4. Multiple Possible Poses

The calculations in SOREPP are performed using the essential matrix  $E$  calculated from  $s$ . Each such matrix represents four possible relative poses  $s$ , as explained

in [1, Chapter 9.6]. Usually only one of them falls within the pose priors' uncertainty region. When however more than one possible interpretation lies within the region, the interpretation selected is the one for which the largest number of reconstructed 3D points lie in front of both cameras.

## 5. EXPERIMENTS ON SYNTHETIC DATA

The next two sections describe extensive experimental evaluations performed in various conditions. In the experiments described in this section, an analysis of the sensitivity of SOREPP to pose noise and inlier fractions based on synthetic data is described, as well as a comparison to some state-of-the-art algorithms. In the next section three large datasets consisting of 442 image pairs were used to evaluate SOREPP performance for real life difficult conditions. The comparison with other methods is naturally unfair, as SOREPP uses data that is unavailable to some of them. Still, it demonstrates the advantage of using pose priors, the challenge that these datasets pose, and the importance of handling pose noise properly.

### 5.1. Experimental Setup

Experiments on synthetic data allow to analyze the results in the pose domain, since the true pose of each camera is known exactly. It also removes any unmodeled effects such as inaccurate intrinsic camera parameters, radial distortions, correlated pose noise and more. The way the simulations have been performed is described below.

#### 5.1.1. Evaluation Setup

Following a similar setup to [27], the set was composed of 400 putative correspondences which were located at a depth of 50% of the scene, and the inliers were contaminated with normally distributed noise with a standard deviation of 1 pixel. The distance between the cameras was 10% of the distance to the scene and two camera configurations were used: side by side and one in front of the other. The second camera was randomly rotated around each axis in a way that the world points were still visible. We simulated a typical smartphone camera by choosing a field of view of  $59.6^\circ \times 46.3^\circ$  and resolution of  $2048 \times 1536$  pixels.

We simulated a representative scenario. In this scenario, the cameras were located 100 meters one from the

other. The pose data of each camera  $i \in \{1, 2\}$  was contaminated with independent normally distributed noise

$$\begin{pmatrix} \sigma_{\psi_i} & \sigma_{\theta_i} & \sigma_{\phi_i} & \sigma_{x_i} & \sigma_{y_i} & \sigma_{z_i} \end{pmatrix}^T = \begin{pmatrix} 5^\circ & 1^\circ & 1^\circ & 5 & 5 & 5 \end{pmatrix}^T. \quad (15)$$

The weights of the correspondences were randomly drawn from two distributions, given in [10] for inliers and outliers. They are based on kernel density estimation for SIFT ratio test values taken from real images.

The analysis was performed for increasing pose noise levels up to the full noise level of (15), and different inlier fractions. For each pose noise level and inlier fraction 100 trials were conducted, each with a random orientation of camera 2, random feature point noises and random correspondence weights.

### 5.1.2. Evaluation Method

We evaluated the performance of the different algorithms in two ways: in the relative pose domain, by comparing the estimated relative pose to the ground truth, and in the image domain. The comparison in the relative pose domain was performed by calculating the angular difference between the estimated translation vector and orientation to the ground truth. The orientation error was defined by the smallest angle that can be used in order to bring the estimated orientation to the true one. When more than one relative pose exists, the one closest to the ground truth value is chosen.

The comparison in the image domain was performed as follows. The pure inliers (without the added noise) served as control points. For each trial, we measured the distance between these control points on image 2 to the corresponding epipolar lines, which were calculated using the control points on image 1 and the estimated relative pose, and choose the maximal error as a quantitative measure. This is done in order to ensure that the estimation is accurate over most of the image. If this value was smaller than 15 pixels, we defined the estimated relative pose as “accurate” or “successful”.

### 5.1.3. Implementations and Running Parameters

SOREPP was implemented in C++ using OpenCV. The implementation and a Matlab mex wrapper are available at [17]. For SOREPP, we set  $m = 5$  according to the analysis that will be described in Section 6.2. The input pose noise parameters were set to the values of (15). The inlier threshold was set to 6 pixels for all the algorithms, including  $\sigma_h$  as the soft inlier threshold of SOREPP. We used USAC [12] in two modes: the first used the five-point algorithm, where the implementation was taken from [3], and the second mode

used our own implementation of the closed form three-point algorithm suggested by [28]. In both modes, the local optimization method was used, applying a non-linear Levenberg-Marquardt optimization for a few iterations over all the five relative pose parameters. USAC parameters were adopted to low inlier fractions by reducing the SPRT  $\epsilon$  parameter to 0.05 and increasing the iterations limit to 850,000 for the five-point mode and 20,000 for the three-point mode, both are supposed to be sufficient for very low inlier fractions.

BLOGS and BEEM were not run for the simulated data since the main goal of these experiments is to test the performance of algorithms with very different basic assumptions and since they make similar assumptions as the five-point USAC they produce similar results. Moreover, BLOGS and BEEM deal with uncalibrated cameras yielding epipolar geometry and not relative pose directly. Thus, they were only reported for the real image pairs.

Bundle adjustment was initialized by the pose prior. It used epipolar constraints with a robustified cost function and pose priors regularization, which is, in fact, the same inner optimization function as SOREPP (12). It is referred as “BA-Epi” in the rest of the paper. All random algorithm results were naturally averaged by the large number of trials and additionally averaged over 5 runs. For real images, the focal length was set to the value in the EXIF header, and the principal point to the image center. The evaluation was conducted on a Windows 7 64bit, Intel i5 4GB computer, using only a single core.

### 5.2. Results

The results for a high inlier fraction of 80% are shown in Figure 3. As expected the accuracy of the pose prior data is low, resulting in low success rate in the image domain. Among the different algorithms being compared, SOREPP yields the most accurate relative pose estimation. Bundle adjustment performs well for low pose noise levels, but is sensitive to increased noise, since in such cases the initial errors are beyond its basin-of-attraction. USAC with the three-point algorithm suffers from a similar phenomenon, because it takes the prior leveling angles as hard constraints. USAC with the five-point algorithm (followed by a non linear refinement) yields the closest results to SOREPP. Both succeeded to solve the epipolar geometry for most trials, as seen in (a3) and (b3). However, analyzing the relative pose domain reveals that the accuracy of SOREPP is higher even for high noise levels. This is explained by the fact that there is a small uncertainty region around the true relative pose, where different relative poses in

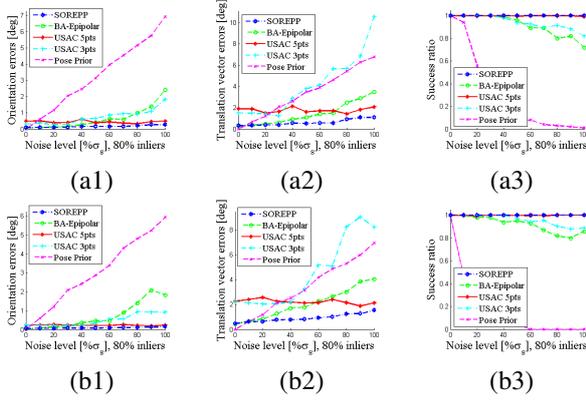


Figure 3: Results on synthetic data, distant cameras, 80% inliers. Rows: (a): Side-by-side configuration, 80% inliers. (b): One-in-front-of-the-other configuration, 80% inliers. Columns: (1): Orientation errors. (2): Translation vector errors. (3): Success ratio, based on the the maximal distance between the control points and the corresponding epipolar lines. All angles are given in degrees.

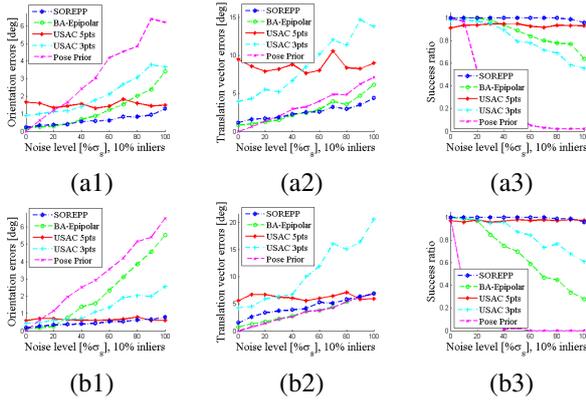


Figure 4: Results on synthetic data, distant cameras, 10% inliers. Rows: (a): Side-by-side configuration. (b): One-in-front-of-the-other configuration. Columns: (1): Orientation errors. (2): Translation vector errors. (3): Success ratio, based on the the maximal distance between the control points and the corresponding epipolar lines. All angles are given in degrees.

that region all explain the epipolar geometry in the image domain. The advantage of SOREPP is that this region and the uncertainty region of the pose priors are in general uncorrelated, therefore their intersection yields a smaller uncertainty region, where SOREPP’s solution is found.

The results for a low inlier fraction of 10% are shown in Figure 4. SOREPP succeeded to estimate the epipolar geometry for most of the trials, when a few failures occurred in the highest noise level. There is a difference however in the quality of the estimation of the

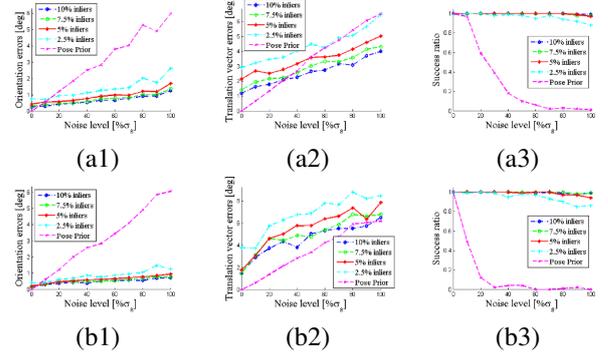


Figure 5: Results of SOREPP on synthetic data, distant cameras, extremely low inlier fractions. Rows: (a): Side-by-side configuration. (b): One-in-front-of-the-other configuration. Columns: (1): Orientation errors. (2): Translation vector errors. (3): Success ratio, based on the the maximal distance between the control points and the corresponding epipolar lines. All angles are given in degrees.

rotation and translation components in the two experiments. This is due to the fact that in the one-in-front-of-the-other configuration changes in the pose parameters have a smaller effect on the epipolar geometry than in the side-by-side configuration. The increased number of outliers reduces the size of the basin-of-attraction of the optimization process. This harms the performance of BA, as seen clearly by the graphs, but does not affect SOREPP, which overcomes the reduced basin-of-attraction by the *Pose Region Search* component. Even in the worst cases, the accuracy of SOREPP is similar to the one of USAC with the five-point algorithm. In side-by-side configurations, the advantage of SOREPP is significant. USAC with the five-point algorithm performs quite well. Still the recovered pose (translation and orientation) parameters are less accurate than SOREPP for both scenarios. Their values are much more accurate (twice as accurate). This is especially true for noise levels below  $50\% \sigma_s$ .

Due to the weights of the putative correspondences, USAC usually converged in the first iterations. Therefore, USAC’s average runtime for the 10% inlier fraction case and the 80% inlier fraction case was 0.5 seconds per image pair. SOREPP’s average runtime was 0.08 seconds per image pair in all cases.

### 5.3. Extremely Low Inlier Fractions

The following experiments are intended to analyze the ability of SOREPP to handle extremely low inlier fractions, and to find its breakdown point. The results are shown in Figure 5. For inlier fractions as low as 5%, SOREPP is still able to estimate the epipolar geometry

try successfully in most of the cases, including for high pose data noise levels, in the sense that the epipolar lines do pass close enough to the corresponding points, as shown in (a3) and (b3). These results demonstrate the large basin of attraction of the SOREPP function. The relative pose estimation reveals that the orientation is estimated accurately, achieving average errors in the order of  $1^\circ$ . The translation vector estimation accuracy is lower, sometimes lower than the noisy pose prior data. This is an estimation effect, where SOREPP chooses the parameters that best fit the model of epipolar geometry and minimize Sampson distances.

For a inlier fraction of 2.5% the ability of SOREPP to solve the epipolar geometry problem decreases. For such a low inlier fraction, the signal to noise ratio between inliers and outliers is too low. The quality of the estimations of the rotation matrix and the translation vector and the success rate decrease as the pose noise increases.

## 6. EXPERIMENTS ON REAL IMAGES

In this section three large datasets consisting of 442 image pairs were used to evaluate SOREPP performance for real life difficult conditions. Since existing images which were used previously to analyze epipolar geometry estimation algorithms do not contain pose measurements, new datasets had to be created. The datasets, including the images, the measured pose of each camera, the control points, and code that applies SOREPP on them, are all available at [17]. The last experiment compares the estimated relative pose to ground truth over a few image pairs.

### 6.1. Experimental Setup

The setup data for the experiments on real images is similar to the setup for the experiments on synthetic data where it is possible. The differences are detailed below.

#### 6.1.1. Evaluation Images

The evaluation data is composed of three datasets collected at different locations. In all cases the SitisMobile’s “GeoCam” application [15] was used supplying the pose associated with each image. *Archi* was taken by a LG G4 smartphone. It presents a relatively easy scenario, most of the images with many inliers. However, since the scene is composed of objects located at variable places, it is not trivial to find the correct solution which is accurate on all of them. *Open* and *Urban* were taken by a Samsung Galaxy SII smartphone. The datasets present challenging scenarios with wide baseline images, small overlapping regions, scale changes,

Dataset	<i>Archi</i>	<i>Open</i>	<i>Urban</i>
Im. Pairs	110	224	108
Cam dist	30 – 530	40 – 440	15 – 200
Camera	LG G4	Galaxy SII	Galaxy SII
Res	$2048 \times 1536$	$2048 \times 1536$	$2048 \times 1232$
Fov X	$54.5^\circ$	$59.6^\circ$	$59.6^\circ$
Az Diff	$\leq 17^\circ$	$\leq 25^\circ$	$\leq 70^\circ$
Init Errs	$\leq 390$	$\leq 375$	$\leq 980$
$\tau$	15	15	15

Table 1: Characteristics of the different datasets. *Im. Pairs* are the image pair quantities, *Cam dist* are the distances between the cameras in meters, *Fov X* are the cameras’ fields-of-view in the x axis, *Init Errs* are the maximal errors calculated directly by the pose prior data in pixel units, and  $\tau$  is the success threshold.

and nondescript objects that make feature matching difficult. Under these conditions the inlier fractions decrease often to less than 10%. Noisy pose measurements are an additional challenge. The characteristics of the datasets are given in Table 1 and representative image pairs are shown in Figures 1,7-9.

#### 6.1.2. Evaluation Method

Since ground truth pose data is unavailable, we evaluate the estimated epipolar geometry based on the image domain only. Each dataset consists of several images of the same scene taken from different locations. A few control points have been defined in the scene and manually marked in each image, when visible. We tried to scatter the points over the images, depending on the ability to recognize the same locations accurately, and on the overlap regions between images. The image pairs are created from images taken at two different locations, automatically generating a set of connected control points. This way even a small set of images creates a large set of image pairs to be used in the experiments. The control points were used as described in Section 5.1.2. The success threshold  $\tau$  defined for each dataset is seen in Table 1. An algorithm’s performance is evaluated mainly by counting the number of successful image pair matches. Due to the asymmetric nature of the relative pose, we consider an image pair in reverse order to be a different pair.

#### 6.1.3. Implementations and Running Parameters

All the algorithms used the same features, calculated by the implementation of SIFT provided by [34] and adapted to be upright. The algorithms used the same putative correspondences, which are the 400 best ranked BLOGS correspondences (using the full set of putative correspondences decreases the inlier fractions, therefore

Regularization	×	√	×	×	√	√	√
$m$	0	0	1	5	1	5	10
<i>Archi</i>	75	64	98	99	90	92	92
<i>Open</i>	105	100	146	153	162	174	174
<i>Urban</i>	24	21	59	53	55	56	59

Table 2: The contribution of the basic SOREPP components: the regularization term and the *Pose Region Search* component. The left column represents the pure optimization of (9), since both components are disabled. The number of successfully estimated image pairs are compared. Putative BLOGS correspondences were used in this experiment. See text for discussion.

reduces the quality of the results), besides BEEM which defines putative correspondences by the SIFT ratio test and uses it as part of the algorithm. SOREPP was tested with both putative correspondence types.

The algorithms used the same parameters as in the experiments of synthetic data. The pose noise standard deviations were also set to the same values as in the synthetic analysis.

We expanded the analysis to some additional algorithms. One algorithm is a conventional BA that tries to estimate the structure besides the cameras pose by minimizing re-projection errors (referred as “BA-Reproj” in the paper). We used our own implementation with a Gaussian M-estimator as a robust function and with pose priors regularization. Structure points were initialized by intersecting putative correspondences according to the pose priors. Two additional algorithms were BEEM and BLOGS, which estimated the fundamental matrix without taking into account the intrinsic calibration data.

## 6.2. Analyzing SOREPP Components

The contribution of the two basic components of SOREPP was evaluated first: the regularization term in (12), and the *Pose Region Search* component under different numbers of optimization initializations  $m$ . For this experiment we disabled the *Solution Refinement* step described in Section 4.3. Note that when the regularization term of (12) is disabled, then the optimization problem degenerates to minimize (9). If additionally  $m = 0$ , then SOREPP performs only the pure optimization of (9). The combinations where  $m = 0$  are equivalent to BA with epipolar constraints, with or without pose priors regularization.

The results using BLOGS correspondences are summarized in Table 2. These results reveal that the most important component is the *Pose Region Search*. For example, this component more than doubles the number of successfully estimated image pairs on the *Ur-*

Algorithm	<i>Archi</i>	<i>Open</i>	<i>Urban</i>	Runtime
BLOGS	76 (69.1%)	76 (33.8%)	43 (39.8%)	Matlab
BEEM	97 (88.2%)	103 (46%)	45 (41.9%)	0.76
USAC-5pts	77 (70%)	90 (40.3%)	42 (39.1%)	131
USAC-3pts	53 (48.2%)	95 (42.6%)	48 (44.3%)	6
BA-Reproj	5 (4.5%)	73 (32.6%)	11 (10.2%)	Matlab
BA-Epi	64 (58.2%)	100 (44.6%)	21 (19.4%)	0.01
SOREPP-SR	98 (89.1%)	158 (70.5%)	55 (50.9%)	0.09
SOREPP-BL	97 (88.2%)	174 (77.7%)	58 (53.7%)	0.09
Prior Only	2 (1.8%)	44 (19.6%)	2 (1.9%)	–

Table 3: Performance comparison for the different algorithms. Accuracy is measured by counting correct image pairs. *SOREPP-SR* is based on SIFT ratio putative correspondences and *SOREPP-BL* uses BLOGS correspondences. *Runtime*: Average time in seconds for C/C++ implementations. *Prior Only* are reference results in which  $F$  was calculated directly from the measured pose.

*ban* dataset, compared to pure optimization: from less than 25 image pairs to more than 50. Its importance depends on the accuracy of the pose prior data. Usually choosing only the best candidate in addition to the input pose is sufficient for accurate estimation. The regularization term also contributes to the algorithm’s performance. Without regularization the performance decreases on the more challenging datasets, while on the *Archi* dataset it slightly harms the results, since many inliers exist anyway and the pose prior is noisy. The importance of the regularization was also explained in Section 5.2, where it helped to improve the accuracy of the estimated relative pose.

## 6.3. Performance Comparison

The performance of SOREPP is compared to that of several state-of-the-art algorithms in Table 3. The datasets represent a greater challenge than the synthetic data, since in many image pairs the inlier fractions are even lower than 10%, the inliers may not be scattered well on the image, the intrinsic camera parameters are not accurately known, and other real world effects. As expected, epipolar geometry calculated directly from the pose prior data is inaccurate and cannot be used without further image-based processing. For example, on the *Urban* dataset the pose prior data is accurate enough for only 2 image pairs. SOREPP demonstrates its ability to successfully match real images taken by a smartphone when the pose measurements are contaminated with substantial noise and the inlier fractions are low. It outperforms algorithms that do not use pose priors (BLOGS, BEEM and USAC with the five-point algorithm), thus demonstrating their importance. For example, on the *Open* dataset, SOREPP gets the score

of 174 while the best RANSAC based method gets only 103. It also performs better than USAC with the three-point algorithm, achieving only 95 successful estimations on this dataset, demonstrating the advantage of using all the priors and not only the leveling angles.

USAC with the five-point algorithm performed worse than on the synthetic data. This has several explanations. Some of the failures occurred due the low inlier fraction, which were many times lower than 10%. USAC is based on PROSAC, therefore needs less iterations than RANSAC, but if the weights of the putative correspondences are not good enough and no solution is found, it degenerates to RANSAC and needs many iterations. However, taking such a large number of 850,000 iterations is impractical, all the more so if even more iterations are required. Other failures might occur due to real world conditions such as the scattering of the points on the image or the limited accuracy of the intrinsic parameters.

The results emphasize an attribute of pose priors that was not mentioned yet. In low inlier fraction cases, it is difficult to distinguish between the true solution and a false one, since both might have a similar number of detected inliers. Pose priors, where available and used by the estimation algorithm, serve also to reject false solutions, which are typically farther from the pose priors than the true one.

Between the two bundle adjustment variants, the one which is based on epipolar constraints works better, and gives the closest results to SOREPP on average. It is reasonable, since the core of SOREPP is the same optimization process that is used in this BA variant. Yet, both BA variants perform significantly worse than SOREPP, demonstrating the advantage of the SOREPP framework and the importance of the *Pose Region Search* component in order to overcome large pose errors.

The *Urban* dataset is characterized by high pose errors, reducing the advantage of SOREPP. Still, even on this dataset, better results are obtained when pose prior data is used. SOREPP achieves the score of 58, in comparison to 42 – 48 of all the RANSAC based methods. SOREPP performs much better than BA, which can not overcome such high prior noises and has succeeded only on 21 image pairs.

SOREPP’s performance using BLOGS putative correspondences is better than its performance using correspondences that are based on the SIFT ratio test, having the biggest advantage on *Open* dataset, where the score has improved from 158 to 174.

The runtime column in Table 3 reveals that SOREPP is fast, having an average runtime of about 0.1 sec-

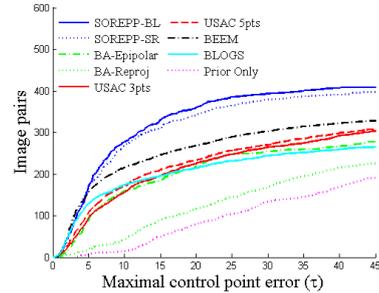
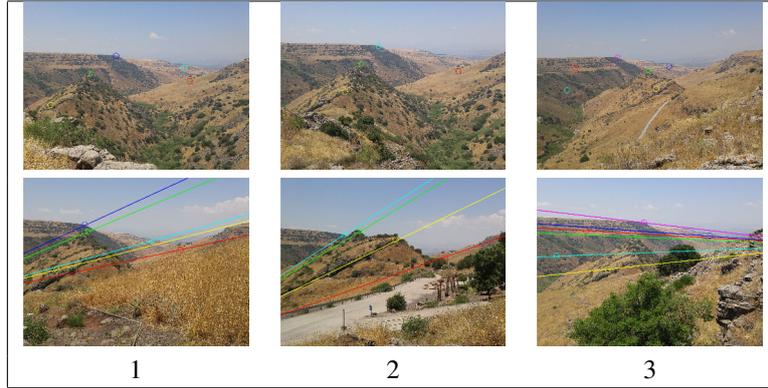


Figure 6: Algorithm performance compared over all the sets. The curves represent the number of correct image pairs as a function of  $\tau$ . See Table 3 for details about the various algorithms. Best viewed in color.

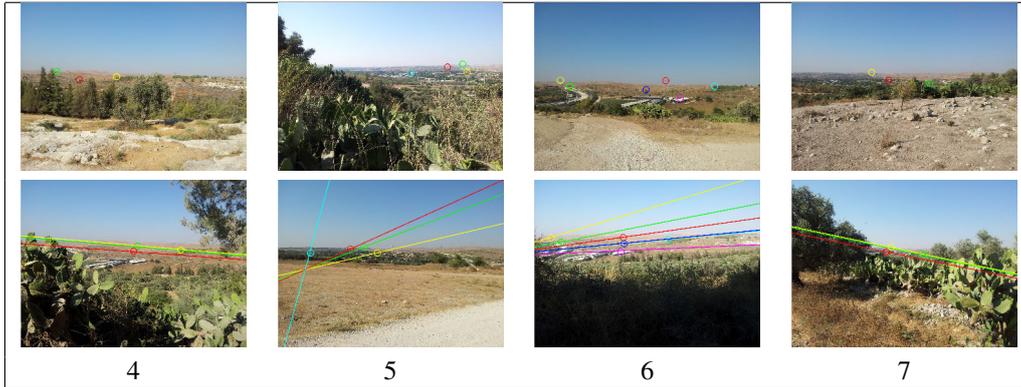
onds per image pair. SOREPP’s runtime is not affected by the low inlier fractions, which slows down all the RANSAC-like algorithms, for example more than 100 seconds for USAC with the five-point algorithm. RANSAC based methods runtime can be reduced by choosing a smaller iteration limit, at the expense of their accuracy. SOREPP’s runtime can be further reduced by better implementation and parallelizing the computation. This can be done, for example, by solving each of the different initializations on a different core.

Figure 6 shows the performance of each algorithm as a function of a continuous threshold  $\tau$ , showing again the low accuracy of the measured pose (*prior only*) while emphasizing the advantage of using it in an estimation algorithm such as SOREPP. For example, for a threshold of 10 pixels, SOREPP obtains the score of 277, while the second best method obtains the score of 215 (about 25% less) and the pose priors gets only 15. Figures 7-9 present some examples from the datasets and compares the performance of the algorithms on them. For each image pair, the left image is the reference image with a small part of the control points marked. The right image is the target image, with the control points and their corresponding epipolar lines estimated using SOREPP. Inliers are counted as detected by SOREPP. This lets us approximate the difficulty of the image pair. SOREPP succeeds in estimating the epipolar geometry while other algorithms have difficulty doing so. For example, in Figure 9(8), SOREPP overcomes prior error of 337.8 pixels and gets an error of 10.6 pixels. The second best method is BLOGS, achieving the maximal error of 26.3 pixels over the control points.



id	#Control points	Prior only	#Inliers (SOREPP detect)	BLOGS	BEEM	USAC 5pts	USAC 3pts	BA Epipolar	SOREPP
1	7	79.4	343 (85.75%)	35.4	23.8	58.9	<b>5.8</b>	87	<b>6.2</b>
2	6	220.3	236 (59%)	<b>11.8</b>	<b>4.6</b>	156.1	390.3	225	<b>12.5</b>
3	8	126.9	271 (67.75%)	26.8	16.6	<b>3.6</b>	126.9	<b>8.7</b>	<b>3</b>

Figure 7: Images from *Archi* dataset matched using some of the algorithms. The *Prior only* column gives reference results calculated directly by the pose prior data. Bold represents successful estimations, where the maximal control point error is smaller than the proper  $\tau$  (15 pixels on this dataset).



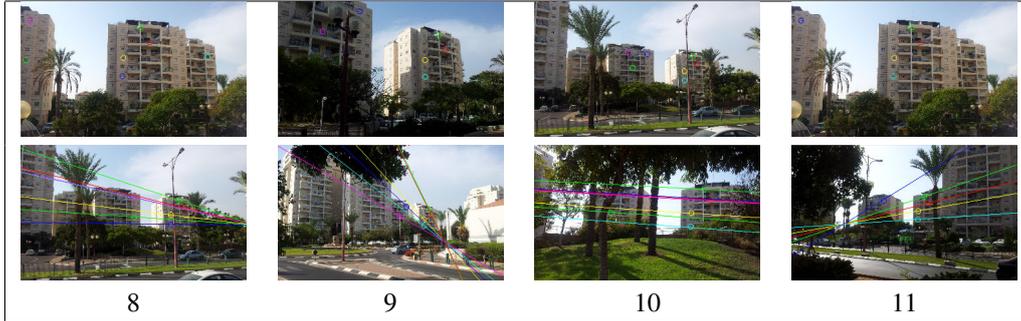
id	#Control points	Prior only	#Inliers (SOREPP detect)	BLOGS	BEEM	USAC 5pts	USAC 3pts	BA Epipolar	SOREPP
4	3	79.5	23 (5.75%)	54.2	59.9	46	20.1	68.1	<b>5.7</b>
5	12	306.1	39 (9.75%)	22	16.4	21.1	19.8	259.4	<b>8.1</b>
6	6	65.6	52 (13%)	<b>10.1</b>	16.9	19.3	<b>15.1</b>	59	<b>10</b>
7	8	41.6	20 (5%)	511.9	19.7	100.6	103.4	40.7	<b>7</b>

Figure 8: Images from *Open* dataset. See Figure 7 for details.  $\tau = 15$  pixels.

#### 6.4. Analyzing the Relative Pose

In the last experiment a few image pairs, with ground truth pose data, were used to analyze and compare the algorithms in the relative pose domain. A few dozens of images were taken around a parking car using a Samsung Galaxy SII smartphone and SitisMobile’s “Geo-

Cam” application [15], supplying the pose associated with each image. The Bundler structure-from-motion software package [3] was used to estimate the cameras’ pose. These poses serve as the ground truth. The algorithms, their parameters, and the evaluation method were all the same as in the synthetic data experiments.



id	#Control points	Prior only	#Inliers (SOREPP detect)	BLOGS	BEEM	USAC 5pts	USAC 3pts	BA Epipolar	SOREPP
8	7	337.8	41 (10.25%)	26.3	188.5	101.9	33.5	313.8	<b>10.6</b>
9	6	170.4	34 (8.5%)	19	25.6	32.7	<b>5.1</b>	189.6	<b>4.9</b>
10	7	217.9	28 (7%)	232	23.4	42.6	30.1	181.4	<b>8</b>
11	5	159.5	28 (7%)	35.8	113.8	20.8	22.3	203.6	<b>9.5</b>

Figure 9: Images from *Urban* dataset. See Figure 7 for details.  $\tau = 15$  pixels.

USAC results were averaged over 5 runs.

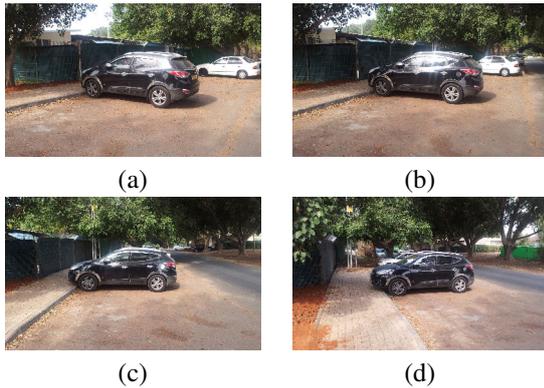


Figure 10: The car images which were used for the relative pose analysis. Image (a) was matched to the three others.

Three image pairs were chosen, representing an increased difficulty level as the baseline grows. The images are shown in Figure 10. The main difficulties of this dataset were the glistening of the car, preventing features from being detected on its planar surfaces and resulting in bad distribution of the putative correspondences, and low quality pose position data due to the proximity of the cameras.

The results are shown in Table 4. In the easy case of images (a) and (b), all the algorithms converged to the true solution and yield similar accuracies. In the moderate difficulty level of images (a) and (c), all algorithms yielded results of similar accuracy. In the difficult case

of images (a) and (d), both USAC variants suffered from low accuracy, probably due to outliers which infiltrated to the estimation, and the bad dispersion of the inliers. The three-point algorithm accuracy was lower, probably due to large prior orientation errors. The epipolar bundle adjustment failed to converge, due to the low inlier fraction and the high pose prior noise. On the other hand, SOREPP demonstrated its ability to overcome this challenge and converged to the true solution, although its accuracy was lower than in the easier examples.

## 7. CONCLUSION

The paper introduced SOREPP, a novel robust algorithm designed to estimate epipolar geometry using pose priors. Pose priors are used to apply an optimization scheme, while their estimated uncertainty is used in order to focus the search in the relevant area in the parameter space. SOREPP is able to deal with extremely low inlier fractions of less than 10% while overcoming significant pose noise, such as an azimuth error of  $7^\circ$ . It does so without slowing down the runtime which, unlike RANSAC-based algorithms, is not sensitive to the inlier fractions. Extensive evaluation on hundreds of image pairs was performed in various conditions, demonstrating the ability of SOREPP to estimate epipolar geometry even under severe, realistic conditions. When the pose priors are supplied, SOREPP outperforms current state-of-the-art methods. In easier cases its performance

Image pair	Azimuth diff	#Inliers (true)	Pose error	Pose Prior	USAC 5pts	USAC 3pts	BA Epi.	SOREPP
(a)(b)	9.5°	385 (96.25%)	R	0.56°	0.37°	0.37°	0.39°	0.35°
			t	80.24°	0.28°	0.28°	0.31°	0.47°
(a)(c)	32.2°	109 (27.25%)	R	1.84°	1.41°	1.58°	1.36°	1.32°
			t	89.43°	3.35°	3.06°	2.15°	2.72°
(a)(d)	51.5°	37 (9.25%)	R	5.47°	7.67°	9.44°	4.14°	2.51°
			t	31.6°	6.8°	7.67°	32.54°	2.48°

Table 4: Estimated pose errors, in the orientation (“R”) and in the translation vector (“t”).

is comparable. We therefore suggest that when pose priors are available the algorithm should be used because of its performance in challenging cases and due to its speed. This is especially useful for realtime applications.

## References

- [1] R. I. Hartley, A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd Edition, Cambridge University Press, 2004.
- [2] D. Scharstein, R. Szeliski, A taxonomy and evaluation of dense two-frame stereo correspondence algorithms, *IJCV* 47 (1) (2002) 7–42.
- [3] N. Snavely, S. Seitz, R. Szeliski, Photo tourism: exploring photo collections in 3D, *ACM Transactions on Graphics (TOG)* 25 (3) (2006) 835–846.
- [4] K. Konolige, M. Agrawal, J. Sola, Large-scale visual odometry for rough terrain, *Robotics Research* 66 (2011) 201–212.
- [5] D. Lowe, Distinctive image features from scale-invariant keypoints, *IJCV* 60 (2) (2004) 91–110.
- [6] M. Fischler, R. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, *Communications of the ACM* 24 (6) (1981) 381–395.
- [7] O. Chum, J. Matas, J. Kittler, Locally optimized RANSAC, in: *German Pattern Recognition Symp.*, 2003, pp. 236–243.
- [8] O. Chum, J. Matas, Matching with PROSAC-progressive sample consensus, in: *CVPR*, 2005, pp. 1:220–226.
- [9] P. H. Torr, A. Zisserman, MLESAC: A new robust estimator with application to estimating image geometry, *Computer Vision and Image Understanding* 78 (1) (2000) 138–156.
- [10] L. Goshen, I. Shimshoni, Balanced exploration and exploitation model search for efficient epipolar geometry estimation, *PAMI* 30 (7) (2008) 1230–1242.
- [11] A. Brahmachari, S. Sarkar, BLOGS: Balanced local and global search for non-degenerate two view epipolar geometry, in: *ICCV*, 2009, pp. 1685–1692.
- [12] R. Raguram, O. Chum, M. Pollefeys, J. Matas, J. Frahm, USAC: A universal framework for random sample consensus, *PAMI* 35 (8) (2013) 2022–38.
- [13] D. Nistér, An efficient solution to the five-point relative pose problem, *PAMI* 26 (6) (2004) 756–770.
- [14] R. Hartley, F. Kahl, Global optimization through searching rotation space and optimal estimation of the essential matrix, in: *ICCV, IEEE*, 2007, pp. 1–8.
- [15] GeoCam from Sitis-Mobile company, available from: <https://play.google.com/store/apps/details?id=ru.sitis.android.geocam&hl=en> (2016).
- [16] GeoCam Free from Wazer Entertainment company, available from: <https://play.google.com/store/apps/details?id=com.myway&hl=en> (2016).
- [17] SOREPP website, <http://is.haifa.ac.il/~ishimshoni/SOREPP/> (2014).
- [18] L. Moisan, B. Stival, A probabilistic criterion to detect rigid point matches between two images and estimate the fundamental matrix, *IJCV* 57 (3) (2004) 201–218.
- [19] F. Sur, N. Noury, M.-O. Berger, Image point correspondences and repeated patterns, Ph.D. thesis, INRIA (2011).
- [20] J. Rabin, J. Delon, Y. Gousseau, L. Moisan, Mac-ransac: a robust algorithm for the recognition of multiple objects, in: *3DPTV*, 2010, p. 051.
- [21] X. Armangué, J. Salvi, Overall view regarding fundamental matrix estimation, *Image and Vision Computing* 21 (2) (2003) 205–220.
- [22] F. Di Corato, L. Pollini, M. Innocenti, G. Indiveri, An entropy-like approach to vision based autonomous navigation, in: *IEEE International Conference on Robotics and Automation (ICRA)*, 2011, pp. 1640–1645.
- [23] J.-P. Tardif, M. George, M. Laverne, A. Kelly, A. Stentz, A new approach to vision-aided inertial navigation, in: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2010, pp. 4161–4168.
- [24] A. Irschara, C. Hoppe, H. Bischof, S. Kluckner, Efficient structure from motion with weak position and orientation priors, in: *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2011, pp. 21–28.
- [25] R. Carceroni, A. Kumar, K. Daniilidis, Structure from motion with known camera positions, in: *CVPR*, Vol. 1, 2006, pp. 477–484.
- [26] D. Wong, M. Hayes, A. Bainbridge-Smith, IMU-aided SURF feature matching for relative pose estimation, in: *International Conference of Image and Vision Computing New Zealand (IVCNZ)*, 2010, pp. 1–6.
- [27] F. Fraundorfer, P. Tanskanen, M. Pollefeys, A minimal case solution to the calibrated relative pose problem for the case of two known orientation angles, *ECCV* (2010) 269–282.
- [28] O. Naroditsky, X. S. Zhou, J. Gallier, S. I. Roumeliotis, K. Daniilidis, Two efficient solutions for visual odometry using directional correspondence, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34 (4) (2012) 818–824.
- [29] L. Kneip, M. Chli, R. Siegwart, Robust real-time visual odometry with a single camera and an IMU, in: *BMVC*, 2011, pp. 1–11.
- [30] B. Triggs, P. F. McLauchlan, R. I. Hartley, A. W. Fitzgibbon, *Bundle adjustment - a modern synthesis*, in: *Vision algorithms: theory and practice*, Springer, 2000, pp. 298–372.
- [31] V. Indelman, R. Roberts, C. Beall, F. Dellaert, Incremental light bundle adjustment., in: *BMVC*, 2012, pp. 1–11.
- [32] M. Tamaazousti, V. Gay-Bellile, S. Collette, S. Bourgeois,

- M. Dhome, Nonlinear refinement of structure from motion reconstruction by taking advantage of a partial knowledge of the environment, in: CVPR, IEEE, 2011, pp. 3073–3080.
- [33] K. Levenberg, A method for the solution of certain problems in least squares, *Quarterly of App. Math.* 2 (1944) 164–168.
- [34] A. Vedaldi, An open implementation of the SIFT detector and descriptor, Tech. Rep. 070012, UCLA CSD (2007).