Dynamic Markers: Optimal control point configurations for homography and pose estimation

Raul Acuna[1], and Volker Willert[1]

***Abstract*— In this paper, we propose a gradient descent optimization method to find optimal configurations of control points for the tasks of homography estimation and subsequent planar pose estimation from a number of $n \geq 4$ noisy (3D-2D) point correspondences. We obtain optimal configurations (for the minimal number of four control points sets (also 5 and 6?) found) by minimizing the first order perturbed solution of the (condition number of the data matrix in the) direct linear transform (DLT) algorithm. This method guarantees points configurations which increases/maximizes the robustness of the DLT homography estimation (that minimize the sensitivity to noise) when noisy correspondences are present. (not only for the DLT transform but as well for other homography estimation methods). Furthermore, it is shown that this optimal control points configurations also increase the performance of state of the art pose estimation methods including the iterative minimization of the reprojection error which is the most accurate algorithm available. (Finally, looking at the statistics of the point configurations, we derive new rotation and scaling invariant coordinate normalizations $T(H)$ as a function of the homography itself. These results shed some light on the influence of the control model points on the accuracy of space resectioning methods, which has been an open problem in the literature so far.**

I. INTRODUCTION

Space resectioning, homography estimation and the PnP problem are some of the most researched topics in the fields of Computer Vision and Photogrammetry. Even though the research on these areas has been wide, there is a surprisingly lack of information regarding the effect of the 3D control point configurations on the accuracy and stability of the estimation methods.

It is clear from the literature (As will be shown in Chapter II) that the control points configurations are relevant. First, it is widely accepted that increasing the number of control points increases the accuracy of the methods in presence of noise (refs?). Second, the spatial configuration of the points affects the estimation. In several studies when simulations are performed to compare methods, great care is given to possible singular points configurations, such as non centered data or near planar cases which are singularities or degenerate cases for certain estimation methods. Additionally, in homography estimation and planar pose estimation methods based on homography it is important to include a normalization step of both the control points and image points.

*This work was sponsored by the German Academic Exchange Service (DAAD) and the Becas Chile doctoral scholarship.

[1]These authors are within the Institute of Automatic Control and Mechatronics, Technische Universität Darmstadt, Germany. (racuna, vwillert)@rmr.tu-darmstadt.de

However, none of the above give an answer to a simple question:

Is there a n-point configuration which can increase the likelihood for a space resectioning method to find the best possible solution?

Many more questions are also connected to this main one:

- May a well conditioned n-points configuration be better than a (an ill-conditioned or) random configuration with more points? When does the increase in number of points outperform the optimal configuration of a fixed number point set?
- A wider separation between points is better?
- Is there a configuration which is better for all camera poses, or on the other hand there is a best point configuration for each pose? Is the optimal point configuration dependent on the homography/camera pose itself?
- If there is an optimal configuration, is it invariant to the intrinsic camera parameters?

The goal of the present work is to define a method that can be used to obtain optimal control points configuration that increase the robustness of the homography estimation and a possible subsequent pose estimation in the presence of noise. This research is relevant both for those interested on the development of algorithms for PnP and space resectioning as well for the community of fiducial markers designers.

We propose the use of a gradient search approach to move the control points in space in order to find the minimum of an optimality condition. The homography was selected as a simple basis for space resectioning, thus restricting the control points configurations to a plane. The optimality condition was defined as the condition number of the A matrix in the DLT method of homography estimation, since it improves the stability of the SVD decomposition and its robustness to noise as it will be described in Chapter **??**.

The paper is structured as follows: In Sec. **??**, state of the art. Sec. **??**, we present our method. Later in Sec. **??** we describe the simulation results, and finally in Sec. **??** we discuss the results and give some conclusions.

II. STATE OF THE ART

A. Space resection and PnP

Camera or space resection is a term used in the field of photogrammetry in which the spatial position and orientation of a photo is obtained by using image measurements of control points present on the photo. A similar concept comes from the computer vision community, where it is known as the Perspective-n-Point (PnP) problem, which in turn

is defined as the process of finding the absolute pose of a calibrated camera in world reference frame given a set of known 3D control points and their 2D camera image measurements.

(PnP is the problem of estimating the pose of a calibrated camera given a set of n 3D points in the world and their corresponding 2D projections in the image)

The main difference between definitions is that traditional PnP assumes a calibrated camera, meanwhile (ref?) in camera resection the camera parameters are assumed as unknown (I think camera resection is also done with calibrated cameras).

PnP can be considered an over-constrained (only for $n \geq 3$) and generic solution to the pose estimation problem from point correspondences. PnP methods can be classified into those which solve for a small and predefined amount of points (n), and those which can handle the general case. The minimal number of points to solve the PnP problem is three.

B. PnP solutions for limited amount of points

The P3P problem has been studied in detail,

it is able to provide up to four solutions for non-collinear points, and in order to find the correct solution additional points are required.

In the case of planar PnP, the P4P case has a unique solution if no 3 points are collinear (Hung et al. 1984).

We look at the number of constraints in terms of corresponding features required to estimate the projective transformation H. A lower bound is available from the number of degrees of freedom and the number of constraints. The matrix H contains 9 entries, but is defined only up to scale. Thus, the total number of degrees of freedom in a 2D projective transformation is 8. Each corresponding 2D point or line generates two constraints on H by Equation $x' = Hx$ and hence the correspondence of four points or four lines is sufficient to compute H (A Survey of Planar Homography Estimation Techniques, 2009).

P3P is a well known solution to the pose estimation problem, however, since it has been proven that pose accuracy usually increases with the number of points [**?**], other PnP approaches that use more points ($n > 3$) are usually preferred. Additionally, the algorithms present limited stability under noisy correspondences, thus many solutions employ outlier rejection methods such as RANSAC.

For the general PnP problem the main aim is to exploit redundancy by using a larger number of correspondences and thus improve the accuracy. The general PnP methods can be broadly divided into whether they are iterative or non-iterative.

C. Iterative PnP Solutions

Iterative approaches formulate the problem as a non-linear least-squares problem and usually they differ in the choice of the cost function to minimize. The cost function is usually associated to an algebraic or geometric error (reprojection error).

The **POSIT** algorithm [**?**] is one of the first iterative solutions. It consists in approximating iteratively to the correct pose by first using an affine camera and then calculating the error introduced by the affine camera assumption to adjust the system. The adjusted system is then used to recalculate the pose.

The **LHM** method [**?**] is one of the best PnP methods to date and probably convergent. The pose is initialized using a weak perspective assumption and then minimizing the object-space collinearity error iteratively. The algorithm operates by successively improving an estimate of the rotation portion of the pose and then estimates an associated translation. The intermediate rotation estimates are always the best orthogonal solution for each iteration. It is globally convergent in the 3D case, however, it is not stable in the planar and quasi-singular cases.

The Procrustes PnP method or **PPnP** [**?**], is an iterative method which casts the PnP problem as an instance of the Orthogonal Procrustes problem in which each measurement may have a different scaling factor. This method tries to reach the best trade-off between speed and accuracy and is significantly easier to implement than other iterative methods.

In order to avoid the risk of local minima, the global optimization method **SDP** [**?**] tackles this by formulating the PnP problem as a semidefinite program with O(n). However the runtime of the algorithm is prohibitive for real-time applications.

The above direct minimization methods [**?**], [**?**], [**?**], [**?**] have the common disadvantage that they return only a single solution for the pose, which might not be the true one. Most of the above mentioned methods can only guaranteed to find a local minima, and the ones that are designed to find a global minima remain highly computing intensive. In general, the major limitation of iterative methods is that they are rather slow and neither convergence nor optimality can be guaranteed, and a good initial guess is usually needed to converge to the right solution.

D. Non-iterative PnP Methods

In order to optimize the computing burden, the non-iterative methods try to reformulate the problem so it may be solved by a potentially large equation system. However, early non-iterative solvers were also computational demanding and worse for larger number of points.

The first efficient and non-iterative O(n) solution was **EPnP** [**?**], which was then later improved by using an iterative method to increase accuracy. EPnP is capable of handling $n >= 4$ and both planar and non-planar configurations. The idea behind EPnP is to represent the n 3D points as a weighted sum of four virtual control points, this means that the PnP problem is reduced to only obtaining the coordinates of these virtual points in camera frame increasing efficiency. One problem of this method is that it minimizes only an algebraic error, it is not stable on cases of pose-ambiguity and only provides one solution.

More recent non-iterative solutions to the PnP problem are based on polynomial solvers trying to achieve linear

performance without the problems of EPnP and with higher accuracy.

The first successful O(n) method is the **DLS** [**?**], the idea is to obtain up to 27 stationary points of the cost function by solving a polynomial equation system of fourth order polynomials, after obtaining the minima, the cost function is evaluated to find the optimal orientation, and the corresponding translation is then computed. This method achieves a least-squares geometric error minimization in linear time, however a Cayley representation of the rotation is used, this parametrization is unfortunately degenerate for all 180 degree rotations around x, y, z axis, reducing the accuracy of the method around this singularities.

Robust PNP or **RPnP** [**?**] is a non-iterative polynomial based solution and the first one that provides more accurate results than iterative algorithms when a low number of points is used $n5$. this method takes a different approach by dividing the PnP problem into several P3P problems, obtaining several fourth order polynomials. The squared sum of the these polynomials is calculated to form a cost function and finally the roots of the derivative of this cost function are found to determine the optimum, obtaining four stationary points. The final solution is the stationary point with the least reprojection error. It is mentioned that not only the amount of points is relevant for the accuracy of the estimation but as well the 3D point configuration, and three broad groups are defined for classification: the ordinary 3D, the quasi-singular and the planar case. The accuracy of RPnP is similar to LHM and it is faster. Nonetheless, this methods can't provide any guarantees on the amount of returned solutions and it is not possible to do a further geometric characterization of its solutions [**?**].

To avoid the singularities present in the DLS method, the **OPnP** (Optimal PnP) was introduced [**?**]. The approach is similar to DLS but the Cayley rotation parametrization is replaced by an unusual non-unit quaternion representation of the rotation matrix, formulating the PnP problem into an unconstrained optimization problem. Up to 40 independent solutions are found by using a two-fold symmetry Grobner Basis solver (avoiding the quaternion sign duality), the candidate solutions are then pruned by using a single damped Newton step. For $n > 6$ the solution is unique and the stationary point with the smallest objective value is returned, in slightly redundant scenarios $n = 4, n = 5$ and for $n = 3$, all remaining minima are returned to the end user. This method doesn't have degenerate cases as in DLS, however, it is still based on an algebraic error, even though the authors point out that their results are comparable to the reprojection error minimization method.

A possible disadvantage of both DLS and OPnP is the amount of stationary points that have to be found in intermediate steps (27 and 40 respectively), this means that each method is in fact calculating far more solutions than a minimal solver, this has been pointed out as a seemingly too high level of complexity by UPnP authors [**?**].

UPnP [**?**] is a linear non-iterative method that generalizes the solution into the NPnP (Non perspective N-point) problem. UPnP employs the object space error without doing convex relaxation techniques, which is supposed to guarantee a geometrical optimum. In contrast to OPnP the authors used normalized unit quaternions to represent rotations. However, just as OPnP a special step is needed to eliminate the sign ambiguity of quaternion, or two-fold symmetry.

Remarkably, one year later in **optDLS** [**?**] a return to the Cayley rotation parametrization used on DLS is proposed, mentioning a simple trick to avoid the singularities and deriving a new optimality condition without Lagrange multipliers. The author demonstrates that the Cayler parametrization is the most compact representation and since it doesn't have the two-fold symmetry problem of the quaternion representations the method is three times faster than OPnP. The experiments performed in this work give optDLS a similar accuracy to OPnP and found that UPnP is actually a suboptimal solution closer to the RPnP method than the OPnP.

More modern approaches to the PnP problem try to face the case when not all the camera intrinsic parameters are available, for example unknown focal length (PnPf). Many of these methods are generalization of regular PnP methods [**?**], [**?**], [**?**], [**?**].

The methods described until this point assume that the observations are equally accurate and free of erroneous correspondences which is not the case. To overcome this, other PnP methods try to include directly into the pose estimation an algebraic outlier rejection scheme which improves the accuracy for a large number of points with noisy correspondences, once again these methods are usually an extension of standard PnP methods [**?**], [**?**], [**?**].

E. Planar pose estimation

Planar pose estimation, or PPE, is a space resectioning problem which involves the process of recovering the relative pose of a plane with respect to a camera's coordinate frame from a single image measurement. In general, there are two main ways of solving a PPE problem, by calculating the model-plane to image-plane homography transformation and then extracting the pose from the homography matrix, this is known as homography decomposition [**?**], [**?**], or by using a set of points in the plane as the measurement with a special case of the PnP methods (planar PnP).

Of the iterative planar PnP methods, the **RPP-SP** [**?**] is the most relevant. It is designed on the assumptions that there is either one (the correct) minimum or there are two local minima of the reprojection error depending on the actual configuration. The method requires an initial pose calculation which is estimated using the LHM method, and then a second solution is obtained which corresponds to a local minimum of the reprojection error with respect to a 1-DoF rotation.

Recently, the Infinite Plane pose estimation **IPPE** [**?**] presents a non-iterative and fast method capable of providing two geometrical related solutions without any artificial degeneracies. It is based in a homography estimation obtained by some other method, the idea is to exploit redundancy in the homography coefficients since a noisy homography will be better at estimating the transform at some regions of the

plane than others. The pose is solved by a non-redundant PDE using first order transform information at a point on the model which is well approximated by the centroid of the points on the model plane. This can be be thought of as "solving pose using transform information within an infinitesimally-small region about a single point on the model plane" [**?**]. IPPE is more accurate than homography decomposition methods and in the majority of cases has better performance and is faster than modern PnP methods, all steps only need floating point operations and it doesn't require any additional numerical libraries.

In general Planar PnP method outperform the best homography decomposition methods when noise is present. Additionally, homography decomposition methods only provide a single solution in contrast to modern planar-PnP methods.

F. Homography estimation

The homography estimation is a key part of the homography decomposition methods and the IPPE algorithm. The standard linear algorithm for homography estimation is the Direct Linear Transform (DLT) [**?**], which was improved later in [**?**] using an orthogonalization step. For both methods the normalization of the measurements is a key step to improve the quality of the estimated homography. As stated in [**?**]: "*Data normalization is an essential step in the DLT algorithm. It most not be considered optional*".

However, the calculation of homographies using normalization has some disadvantages [**?**]. First, the coordinate normalization matrices are calculated from noisy measurements and thus are sensitive to outliers, and second, for a given measurement of control points the noise affecting each point is independent of the others, however, in normalized measurements this independence is removed with the additional consequence that the errors in the normalized matches won't be i.i.d. Gaussian noise anymore [**?**]. A method is proposed in [**?**] which tries to overcome this problems by avoiding the normalization step and using instead a Taubin estimator which in the end produced similar results than the normalized one.

G. Control points configurations

It is pointed out in [**?**], [**?**] that the 3D point configurations have an influence in the local minima of the PnP problem. In the RPnP paper [**?**] a broad classification of the control points configurations into three groups is presented. The classification is based on the Rank of the 3x3 matrix M^TM, where $M = [X_1, X_2, X_3, ..., X_n]^T$, X_i is the coordinate of control point i and n is the amount of control points. The defined groups are: 1) Ordinary 3D case, when the $Rank(M^TM) = 3$ and the smallest singular value of M^TM is different to zero. 2) Planar case, when the $Rank(M^TM) = 2$ and 3) Quasi-singular case, when the $Rank(M^TM) = 3$ and the ratio of the smallest eigenvalue to the largest one is very small (< 0.05).

In EPnP it is shown that if the control points are taken from the "*uncentered data*" or the region where the image projections of the control points cover only a small part of the image, the stability of the compared methods greatly degrades. In RPnP it is elaborated that based on the previous classification this "*uncentered data*" is a configuration that lays between the "ordinary 3D case and the planar case.

Some assumptions about the influence of the control points configurations are also present in IPPE. Through statistical results the authors found out that the accuracy for the 4-point case decreases if the points are uniformly sampled from a given region, some of the configurations will lead to a worse conditioning of the homography estimation problem, in their case they avoided this problem by selecting the corners of the region as the positions for the control points and then refer the reader to the Chen and Suter paper [**?**], were the analysis of the stability of the homography estimation to 1st order perturbations is presented, in this analysis it is clear that the error in homography estimate is dependent of the singular values of the A matrix in the DLT transform.

On the same line, in [**?**] a way of visualizing homography uncertainty for different camera poses is presented. It is proved by simulations that different poses of the camera are more stable for the homography estimation process. Different poses of the camera for static control points can be seen as well as having a static camera and moving the control points instead.

In absence of noise the PnP methods and homography estimation methods return the true solution. The problem is the image and model noise propagation in the pose estimation process. If it were possible to select the control points it would be obvious to select those which increase the robustness of the algorithms in presence of noise. It is relevant that even with the great amount of research in this field there is not much information about the influence of the control point configurations and no clear answer to the best possible point configuration is given besides the vague suggestions and classifications presented above.

In this paper we propose a simple approach to find optimal control points configurations. We assume that there must exist a metric which represents the robustness of a given point configuration to noise for homography estimation and PnP methods. The problem is then to find a suitable metric and then use it as cost function which has to be minimized by moving the control points in the space. The location of the minima for a given number of points should be the one that increases the robustness of the estimation methods versus noise.

III. Optimal point configuration for homography estimation

We shortly summarize the *golden standard* optimization methods for pose estimation:

General point configuration for pose estimation:

Given a 3D-2D point correspondence with i-th 3D point P_i with world W coordinates $\mathbf{X}_i^W = [X_i^W, Y_i^W, Z_i^W]^T \in \mathbb{R}^3$ and its corresponding projection p_i onto a planar calibrated camera[1] with normalized image coordinates $\mathbf{x}_i = [x_i, y_i]^T \in \mathbb{R}^2$ the

[1]Assuming the calibration matrix $\mathbf{K} \in \mathbb{R}^{3\times3}$ to be known, the homogeneous image coordinates in pixel $\overline{\mathbf{x}}_i' = [x_i', y_i', 1]^T$ can be transformed to homogeneous normalized image coordinates in metric units $\overline{\mathbf{x}}_i = \mathbf{K}^{-1}\overline{\mathbf{x}}_i'$.

relation between these points is given by the relative pose $g = (\mathbf{R}, \mathbf{T}) \in SE(3)$ (Euclidean transformation) between world W and camera C frame $\mathbf{X}_i^C = \mathbf{R}\mathbf{X}_i^W + \mathbf{T}$ followed by a projection π whereas $\mathbf{x}_i = \pi(\mathbf{X}_i^C) = [X_i^C/Z_i^C, Y_i^C/Z_i^C]$.

This leads to the relation:

$$\mathbf{x}_i = \pi(\mathbf{X}_i^C) = \pi(\mathbf{R}\mathbf{X}_i^W + \mathbf{T}). \quad (1)$$

Including additive noise $\boldsymbol{\varepsilon}_i = [\varepsilon_i, \zeta_i]^T$ on the error-free image coordinates $\mathbf{x}_i$ we get noisy measurements of the image coordinates $\tilde{\mathbf{x}}_i = \mathbf{x}_i + \boldsymbol{\varepsilon}_i$. Thus, we can solve for the reprojection error $\|\boldsymbol{\varepsilon}_i\|_2^2$ of each point which is a squared 2-norm. Minimizing the squared 2-norm of all points for the optimal pose $(\hat{\mathbf{R}}, \hat{\mathbf{T}})$ leads to the following least-squares estimator

$$(\hat{\mathbf{R}}, \hat{\mathbf{T}}) = \text{argmin}_{\mathbf{R},\mathbf{T}} \sum_{i=1}^{n} \|\boldsymbol{\varepsilon}_i\|_2^2, \quad n \geq 3. \quad (2)$$

Iterative gradient descent optimization of (2) leads to the most accurate pose estimation results in the literature so far also for planar point configurations and is often used as the reference (refs?).

Planar point configuration for pose estimation:

If the world coordinates are all on a plane -homography 8 degrees of freedom, each point two constraints

Here we answer: Why we use homography estimation as the basis?

We start with the golden standard DLT algorithm.

A. Background and notation

Here the basic equations of the camera and the DLT algorithm are presented

B. The matrix condition number

C. Gradient descent approach for finding the optimal point configuration

IV. Simulations

Simulation description:

Metrics for 3D rotatitons: [**?**]

A. Gradient descent results for homography estimation

- Use 4 points as the base for the errors - Compare different homography methods

- Show a typical run for: - Fronto-parallell configuration - Inclined view (30 degrees or similar)

- Show different typical point configurations for 4,5,6,7 and 8 points

- Compare different number of points: - Average percentage of improvement versus initial random configuration

B. Gradient descent results for pose estimation

Explain the motivation (why are we also interested in pose estimation, even when we are improving the A matrix of the DLT transform)

- 4 points compared with different methods

V. Conclusions and Future Work

Mention that EPNP is considered not good for less that 6 points, but when we use the ideal point configuration for the pose then the results with 4 points are equally good.

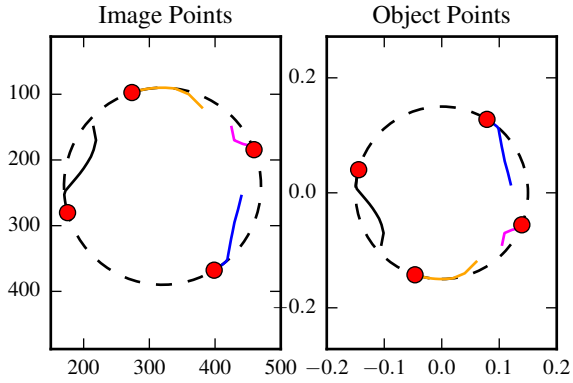

Fig. 1: Movement of the points in image and object coordinates during gradient descent for the Fronto-Parallel configuration. The red dots mark the final point configuration.

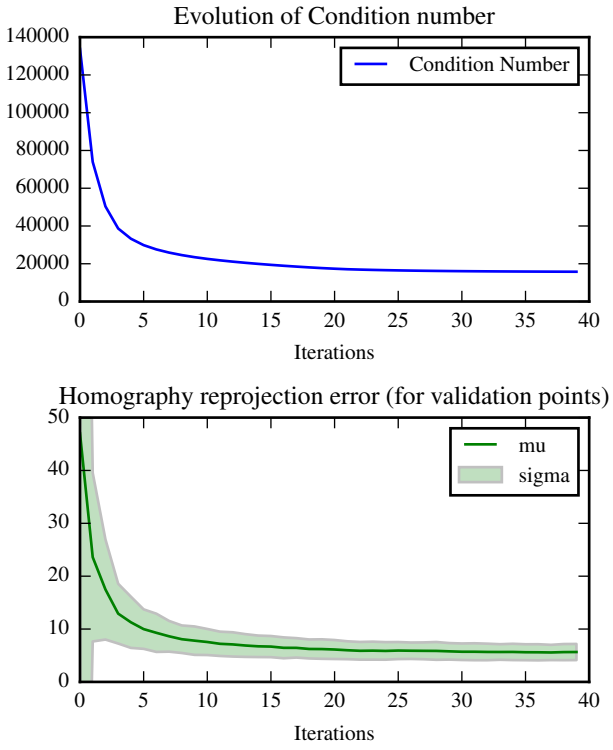

Fig. 2: Evolution of the condition number and the homography reprojection error during gradient descent (Fronto-Parallel configuration).

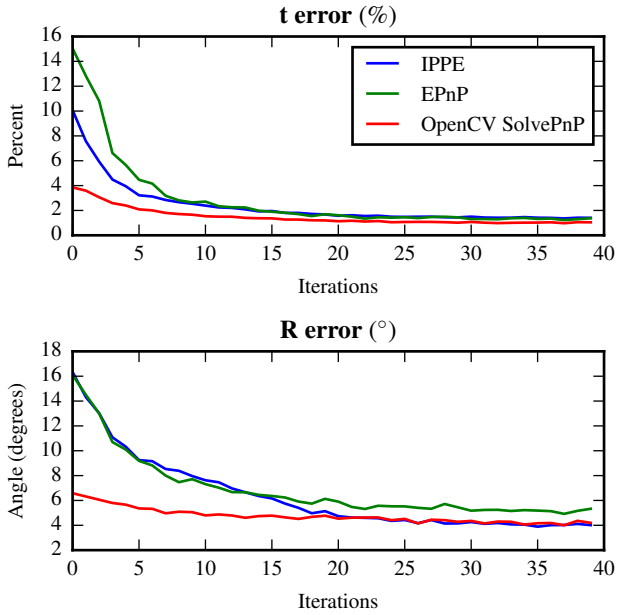

Fig. 3: Evolution of the pose estimation during gradient descent (Fronto-Parallel configuration).

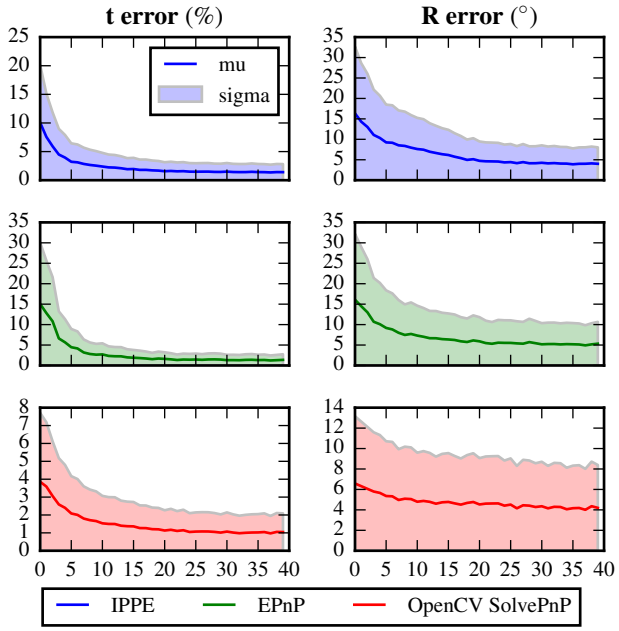

Fig. 4: Evolution of the pose estimation during gradient descent for each algorithm including standard deviations (Fronto-Parallel configuration).

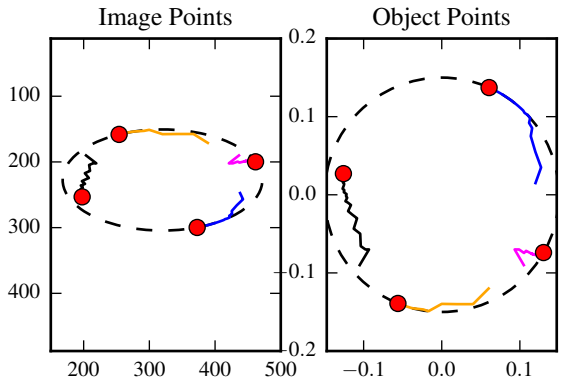

Fig. 5: Movement of the points in image and object coordinates during gradient descent for the inclined configuration. The red dots mark the final point configuration.

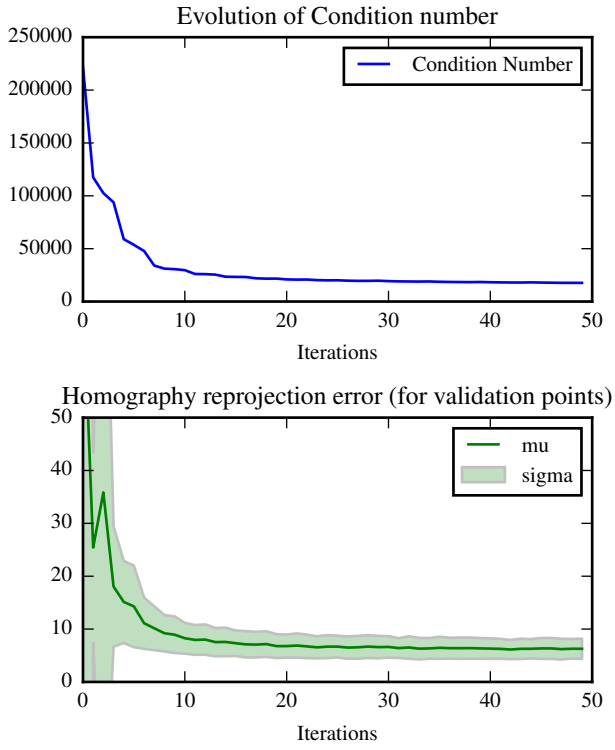

Fig. 6: Evolution of the condition number and the homography reprojection error during gradient descent (inclined configuration).

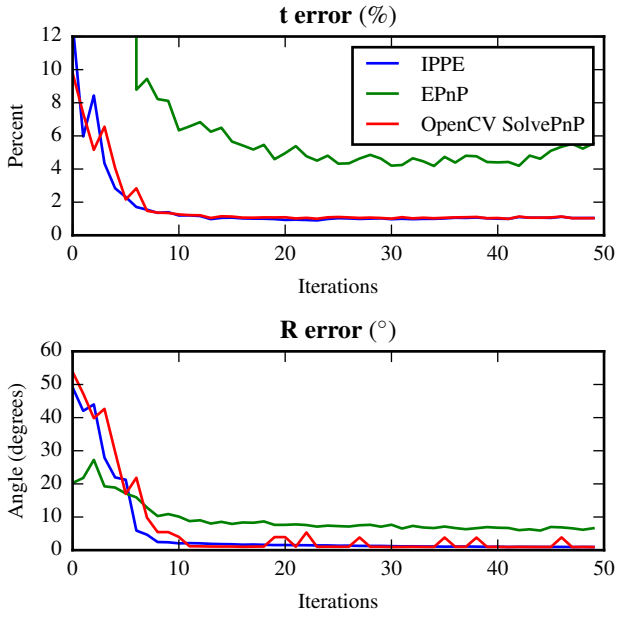

Fig. 7: Evolution of the pose estimation during gradient descent (inclined configuration).

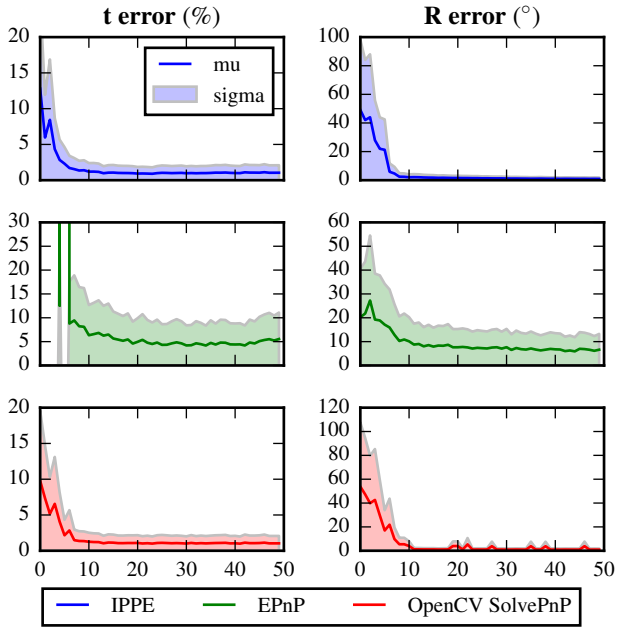

Fig. 8: Evolution of the pose estimation during gradient descent for each algorithm including standard deviations (inclined configuration).