Renormalization method and unbiased statistical estimators: an alternative implementation to the EPnP algorithm.

Iván Maldonado Zambrano, Javier Flavio Vigueras Gómez and Arturo Hernández Aguirre

Abstract—In this paper, we deal with the Perspective-n-Point problem (PnP): the estimation of the pose of a calibrated camera from $n$ 3D to 2D point correspondences. In the state of the art, there exists a solution, named EPnP, with linear computational complexity $O(n)$; nevertheless, the solution is computed as a linear combination of four eigenvectors of a linear system design matrix and is not always easily and straightforward determined.

Based on Kanatani’s renormalization method and unbiased statistical estimators, we suggest an alternative implementation to the EPnP algorithm, efficient and adequate to minimize the image reprojection error criteria. The performance of the approach has shown to be better in the presence of noise compared with the classical Minimum Least Squares method (MLS); the latter is notoriously non-robust to noisy measurements and the estimations can be biased.

Our method takes into account the statistical distribution of the data and the eventual errors in the camera parameters, and is straightforwardly computed from the null-space of a given linear system. It is important to mention that our method does not affect the EPnP's computational complexity, being more accurate and robust than the original version, as it is shown in the experiments.

Index Terms—camera pose, renormalization, unbiased statistical estimators.

1 INTRODUCTION

In computer vision, as well as in the photogrammetry industry, the visual pose estimation problem is defined as estimating the position and the orientation of a camera into an environment. In order to compute the camera pose in a robust way, sometimes it is necessary to assume that some information is known: (i) the intrinsic camera parameters, (ii) the location of some landmarks in the world and (iii) the correspondence between landmarks and their images. Under these assumptions, the pose estimation problem is also called the Perspective-n-Point problem (PnP), such that the geometry of $n$ feature points is used.

The PnP problem has been widely studied and it is one of the most important research subjects in photogrammetry, model-based localization and landmark-based robot navigation. Several solutions have been developed for three or more correspondences in coplanar [1], [2] and non-coplanar [3], [4], [5], [6], [7], [8] configurations. The approaches may be also classified into: (i) iterative [1], [7], [8], [9], often requiring a good initialization; and (ii) non-iterative [3], [4], [5], [6], trying to give a closed-form solution without an initial approximation. Computation time is always an important factor in designing and selecting a PnP solver, because many applications require the problem to be solved in real-time.

Recently, Noguer et al. introduced the non-iterative Efficient Perspective $n$ Points method (EPnP) [7]; certainly, one of the fastest and most accurate non-iterative methods, robust to noise, stable, and having the ability to handle pose computation of hundreds of points in real-time. The success of the EPnP lies in its strategy of representing the location of landmarks in the world as a linear combination (thus, a weighted sum) of some control (real or virtual) points. The use of control points reduce the number of unknowns to treat and allows to compute the camera pose in $O(n)$ time, regardless of the amount of feature points used in the application. EPnP method builds a homogenous system of linear equations and search for the right linear combination of vectors describing the null space of a defined $(12 \times 12)$ matrix which provides the best and final pose estimation.

Motivated by [7] and observing that the EPnP’s linear combination is necessary and compulsory in their approach to reduce the errors due by the presence of noisy data, we suggest an improved PnP implementation based on K. Kanatani’s [10] renormalization method and unbiased statistical method, arose from the study of noise and its characterization for linear concurrence and incidence problems, like the PnP. These methods allow us to include statistical information from data in an iterative way. Despite of the iterativity of the suggested process, our new approach provides a strict unique solution as fast, accurate and robust as the original EPnP linear combination.

1.1 Outline of the article

In section 2 we introduced the EPnP method. In section 3, we explained and introduced the Kanatani’s Renormalization method and the Unbiased Statistical method as an alternative to the EPnP algorithm. In section 4, we show the implementation of our new approach. In the last section, we show some experiments showing that our approach is as fast, accurate and robust as the original EPnP algorithm.

2 THE EPnP APPROACH

As in Noguer, et al. [7], we assumed that the 3D coordinates and the 2D image projections of $n$ given reference points are known. The goal is to retrieve their 3D world coordinates in...
the camera coordinate system and, consequently, the rigid 3D transformation (orientation and translation) that aligns both reference systems.

Contrary to the classical camera pose methods, the EPnP algorithm uses a special parameterization (section 2.1) expressing the 3D point coordinates as a weighted sum of virtual control points, 4 for general non-coplanar configurations and 3 for planar configurations. In this article we just focus in the general configuration, nevertheless, an analog approach for planar configurations may be easily deduced by using only 3 control points to describe the 3D plane passing through. The main idea of implementing control points is to reduce the number of unknow variables to compute (12 for general configuration and 9 for planar configuration), instead of 3n as traditional approaches do.

The PnP problem is then simplified to obtain the camera coordinates system from only 4 (or 3) virtual control points. From n 3D world coordinates and their 2D image projections, EPnP creates a homogenous system and obtains the best solution from a linear combination $\beta_i, i = 1, \ldots, 4$ of the null space of a $2n \times 12$ (or $2n \times 9$) $M$ matrix. Applying an optimization technique as EPnP gives priority to solve the system $M^T M x = 0$, considering that the most time consuming step of the whole application is directly proportional to the number of reference points used (thus, it is an $O(n)$ complexity).

In the following subsections we introduce the parameterization used by the EPnP method for general configurations, summarizing the EPnP algorithm and showing that the given results are biased.

### 2.1 Parameterization of the EPnP in the general case

Noguer, et al. suggested in [7] that n 3D world coordinates (represented with a $w$ superscript) and their respective camera coordinates (represented with a $c$ superscript) can be described in terms of a weighted sum of 4 virtual control points $c_i, j = 1, \ldots, 4$. Considering that the relation holds for every pair of points we have

$$p_i^w = \sum_{j=1}^{4} \alpha_{ij} c_j^w, \text{ with } \sum_{j=1}^{4} \alpha_{ij} = 1, \quad (1)$$

and,

$$p_i^c = \sum_{j=1}^{4} \alpha_{ij} c_j^c. \quad (2)$$

Coefficients $\alpha_{ij}$ are the key of the whole method and it is demonstrated in [7] that generating the linear system of equations is a linear time process. The selection of control points can be arbitrary (but strictly non-coplanar), but in the experiments it is recommended to form a vector basis.

### 2.2 EPnP solution as a linear combination of vectors

Assuming that the camera is calibrated, the intrinsic parameters $A$ are known as well as the 2D image projections $[u_i, v_i, 1]^T, i = 1, \ldots, n$ and their respective coefficients $\alpha_{ij}$ computed from the 3D coordinates as earlier was referred. Matrix $A$ groups the focal coefficients $(f_u, f_v)$ and the principal point position $(u_c, v_c)$, such that the unique unknowns of our expressions below are the 3D points camera coordinates $[x_j^c, y_j^c, z_j^c]^T$ for control points $c_j$.

$$\forall i, w_i \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = A p_i^c = A \sum_{j=1}^{4} \alpha_{ij} c_j^c. \quad \text{(3)}$$

more elaborated,

$$\forall i, w_i \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \begin{bmatrix} f_u & 0 & u_c \\ 0 & f_v & v_c \\ 0 & 0 & 1 \end{bmatrix} \sum_{j=1}^{4} \alpha_{ij} \begin{bmatrix} x_j^c \\ y_j^c \\ z_j^c \end{bmatrix}. \quad \text{(4)}$$

If projective parameters $w_i, i = 1, \ldots, n$ are substituted by $\sum_{j=1}^{4} \alpha_{ij} z_j^c$, two linear equations for each reference point are produced:

$$\sum_{j=1}^{4} \alpha_{ij} f_u x_j^c + \alpha_{ij} (u_c - u_i) z_j^c = 0, \quad (5)$$

$$\sum_{j=1}^{4} \alpha_{ij} f_v y_j^c + \alpha_{ij} (v_c - v_i) z_j^c = 0. \quad (6)$$

If equations (5) and (6) are put into an array, a homogenous linear system $M x = 0$ is obtained and can be solved with the DLT (Direct Linear Transform) method [11]. Solving for the $M$ matrix of $2n \times 12$ size is not as efficient as solving for $M^T M$ matrix of $12 \times 12$ size because the number of equations is reduced in the second case, producing the least square solution on the point reprojection errors.

To give a final solution $x = [c_1^T, c_2^T, c_3^T, c_4^T]^T$, the EPnP algorithm solves a small number of quadratic equations in order to compute $\beta_i, i = 1, \ldots, N$ and express $x$ as a linear combination of $N$ eigenvectors $v_i$ of $M^T M$ (defining its null space):

$$x = \sum_{i=1}^{N} \beta_i v_i. \quad \text{(7)}$$

Noguer, et al. in [7] never use values of $N$ greater than 4 and the final solution is computed for all $N$ and keep the one that yields the smaller reprojection error:
error = \sum_i \text{dist}^2 \left( A [R \{t\} p_i^w, \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} \right). \quad (8)

2.3 Biased solution
Kenichi Kanatani [10] introduced a model that describes the noise behaviour and allowed him to perform an statistical analysis on the incidence-concurrence linear problems. With this model, Kanatani showed that the linear least square solutions are in general statistically biased and derived a scheme to unbias the estimations. In the EPnP algorithm, the least square solution is compatible with $M^T M$. If we observe the behaviour of $\beta_i, i = 1, \ldots, N$ computed by the EPnP algorithm, there exist a predominance of one of the four eigenvectors and it is not necessarily the one that correspond to the smallest eigenvalue as we could have thought, Fig. 1. This behaviour indicates for very high noise levels the solution to the linear system ceased to belong to the lower eigenvalue and for that reason the need to implement a linear combination of four eigenvectors to get a better approximation.

3 Kanatani’s Statistical Analysis
Consider $n$ collinear image points $p_i, i = 1, \ldots, n$ in homogeneous coordinates and $N$-vectors $m_i, i = 1, \ldots, n$ of each point when there is no noise. We know that (Theorem 2.8 in [10]) the common line of multiple collinear image points of $N$-vectors $m_i, i = 1, \ldots, n$ is robustly computed and represented as the unit eigenvector of the moment matrix:

$$M = \sum_{i=1}^{N} W_i m_i m_i^T.$$ \quad (9)

where $W_i$ are positive constants and the smallest eigenvalue of $M$ is zero.

3.1 Model of noise
Kanatani assumed that noise affects each image point $p_i, i = 1, \ldots, n$ in an isotropic way, and perturbs each $N$-vector as $m_i' = m_i + \Delta m_i$. Considering that noise expectation $E[\Delta m_i]$ is equal to 0 and that the difference of $\Delta m_i$ with respect to $m_i$ is small, the covariance matrix $V[m_i]$ characterize the error behaviour:

$$V[m_i] = E[\Delta m_i \Delta m_i^T]. \quad (10)$$

3.2 Renormalization Method
Minimum least squares (MLS) approaches always tries to solve:

$$\sum_{i=1}^{N} W_i (p_i, m_i')^2 \rightarrow \text{min.} \quad (11)$$

The inner product $(p_i, m_i)$ in the absence of noise is equal to 0 but in the presence of a perturbation the error $\epsilon_i = (p_i, m_i') = (p_i, \Delta m_i)$ acquires a mean equal to 0 and a variance

$$\sigma^2 = E[\epsilon^2] = (p_i, V[m_i] p_i). \quad (12)$$

1. Given the point $p$ in homogeneous coordinates $(x, y, 1)$, its $N$-vector $m = \pm N[p] = \frac{p}{\|p\|}$, with $\|p\| = \sqrt{x^2 + y^2 + 1}$.

Obeying a standard Gaussian distribution with variance $u_i = \epsilon_i/\sigma_i$, Eq. (11) is equivalent to:

$$\sum_{i=1}^{N} u_i^2 = \sum_{i=1}^{N} \frac{(p_i, m_i')^2}{(p_i, V[m_i] p_i)} \rightarrow \text{min.} \quad (13)$$

with,

$$W_i = \frac{1}{(p_i, V[m_i] p_i)} \quad (14)$$

It is important to observe that large weights correspond to reliable data with small covariance matrices, whereas small weights correspond to unreliable data with large covariance matrices.

3.3 Unbiased statistical estimator
A perturbation on each $N$-vector $m_i \rightarrow m_i + \Delta m_i$ causes a perturbed moment matrix:

$${\hat{M}} = \sum_{i=1}^{N} W_i (m_i + \Delta m_i)(m_i + \Delta m_i)$$ \quad (15)

$$= \sum_{i=1}^{N} W_i (m_i m_i^T) + \sum_{i=1}^{N} W_i (\Delta m_i \Delta m_i^T)$$ \quad (16)

$$= M + \Delta M$$ \quad (17)

where the expectation of $\Delta M$ is not zero:

$$E[\Delta M] = \sum_{i=1}^{N} W_i E[\Delta m_i \Delta m_i^T] = \sum_{i=1}^{N} W_i V[m_i]. \quad (18)$$

while on the contrary of

$$\sum_{i=1}^{N} W_i (m_i \Delta m_i^T) = \sum_{i=1}^{N} W_i (\Delta m_i m_i^T) = 0. \quad (19)$$

because $E[\Delta m_i] = 0$.

From Eq. (18), the following optimal unbiased estimation scheme was derived by K. Kanatani [10] (pag. 293):

Proposition 1: The common line fitted to image points of $N$-vectors $m_i$ with covariance matrices $V[m_i]$ estimated as the unit eigenvector of

$$M = \sum_{i=1}^{N} W_i \left( m_i m_i^T - V[m_i] \right) \quad (20)$$

for the smallest eigenvalue is estitically unbiased.

4 EPnP approach with Renormalization and Unbiased Statistical Estimator
With Renormalization and Unbiased Statistical Estimator, the solution to the EPnP method become the eigenvector that correspond to the null space of an unbiased $M$ matrix and not the linear combination of null eigenvectors of a biased $M$ matrix, as originally. Considering that computing $M^T M$ is the most expensive part, our new approach computational complexity is as good as the original one.
We assume that we know the 3D point coordinates, their 2D image reprojectations and that our camera is calibrated, as originally stated. For our approach, we must add and consider that 3D coordinates are affected by noise and assume that it is characterized by a normal distribution with zero mean and known standard deviation $\sigma$. The error in $M$ is caused by the propagation of the standard deviation $\sigma$ presented in all the 3D coordinates. The Kanatani’s statistical analysis succeed to avoid error is known and just lies on that multiply the solution vector.

The error propagated matrix is [from Eq. (13)] given by:

$$M = \sum_{i=1}^{n} (W_i(\vec{v}_i^T \vec{v}_i) + W_i(\vec{v}_i^T \vec{v}_i - V [\vec{v}_i^T \vec{v}_i]).$$

6) Applying singular value descomposition to $M$, the solution to the homogenous linear system $MX = 0$ always correspond to the eigenvector with the lowest eigenvalue. The final solution is gotten when vector $x = [c_1, c_2, c_3, c_4]^T$ converges.

7) Using the 3D observed points and their respective image projection Eq.(2), the camera pose estimation $(R, t)$ is computed.

8) For 3D coordinates and 2D image projection, the reprojection error Eq.(8) is estimated with $(R, t)$ parameters.

6 RESULTS

In this section, we show that the implementation of the Normalization method and The Statistical Unbiased method conserved the $O(n)$ complexity and is as accurate and robust as the original EPnP method [7].

Our experiments were conducted using synthetic data and were generated using different number of reference points $n$ and standard deviation $\sigma$ with the purpose of recreating potential situations where there is little information and a certain amount of noise in the data. Our programs were implemented in MATLAB Version 7.4 and our results were generated averaging 300 runs.

The EPnP algorithm code used into the experiments was obtained from http://cvlab.epfl.ch/~fmoreno and, as Noguer et al. cited in their article [7], we determined the relative error rotation $R$ and translation $t$ by:

$$E_{rot} (%) = ||q_{true} - q|| / ||q||,$$

$$E_{trans} (%) = ||t_{true} - t|| / ||t||.$$

where $R_{true}$ and $t_{true}$ are the true camera rotation and translation respectively, $q$ and $q_{true}$ are the normalized quaternions corresponding to the rotation matrix as well as $t$ is the computed translation vector.

6.1 Time of process

As mentioned above, our method is an iterative procedure and it was necessary to know how this affects the complexity of the EPnP. Conducting some experiments with different number of reference points $n$ and error $\sigma$ we found that the time of process
of WEPnP and WUEPnP is directly proportional to \( n \) and the standard deviation \( \sigma \) presented does not affect this property for our two approaches. Nevertheless, for the EPnP method, the time of process is directly affected by \( \sigma \) because at the final step of the classical procedure it is necessary to compute the coefficients of a linear combination of vectors, which is an iterative process too, and this is critical for high noise levels, as it is shown in Fig.(2). The complexity of the WEPnP and WUEPnP methods does not rise with \( n \) because it is not necessary to compute such an adequate linear combination.

### 6.2 Reprojection, rotation and translation error

For the next experiment, we used reprojection, rotation and translation errors as measurements of comparison between EPnP, WEPnP, and WUEPnP. For 300 runs, \( n = 12 \) and \( n = 20 \) with \( \sigma = 5 \), it is evident that the mean reprojection error between EPnP and our approaches are very similar, and increases accordingly to the number of reference points \( n \), see Figs.(3a and 3b).

In Fig.(3c) EPnP works satisfactorily with a small number of reference points \( n \). Contrary to WEPnP and WUEPnP that requires a big enough set of reference points to estimate the error behavior. However, the experiments were conducted for a very high noise level (\( \sigma = 5 \)), when nice interest point detectors may currently offer an average noise level smaller than a half pixel.

### 7 Conclusions

WEPnP and WUEPnP, despite of being iterative procedures, are as fast, accurate and robust as the original EPnP. The only disadvantage with WEPnP and WUEPnP in comparison with EPnP is that fail for a very small number of observed points (less than 9 for our experiments), being 4 points the minimum number of required observations.

As a future work, we pretend to estimate automatically the standard deviation \( \sigma \) from data, the covariance matrices for the rotation and translation parameters, and add incertitude information about the camera calibration.

### References


