



Robust Pose Estimation Based on Maximum Correntropy Criterion

Qian Zhang, Badong Chen

► To cite this version:

Qian Zhang, Badong Chen. Robust Pose Estimation Based on Maximum Correntropy Criterion. 17th IFIP International Conference on Artificial Intelligence Applications and Innovations (AIAI), Jun 2021, Hersonissos, Crete, Greece. pp.555-566, 10.1007/978-3-030-79150-6_44 . hal-03287697

HAL Id: hal-03287697

<https://inria.hal.science/hal-03287697>

Submitted on 15 Jul 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Robust Pose Estimation Based on Maximum Correntropy Criterion

Qian Zhang and Badong Chen^(✉)[0000-0003-1710-3818]

Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University,
Xi'an, Shaanxi, People's Republic of China
zhangqian2018@stu.xjtu.edu.cn
chenbd@mail.xjtu.edu.cn

Abstract. Pose estimation is a key problem in computer vision, which is commonly used in augmented reality, robotics and navigation. The classical orthogonal iterative (OI) pose estimation algorithm builds its cost function based on the minimum mean square error (MMSE), which performs well when data disturbed by Gaussian noise. But even a small number of outliers will make OI unstable. In order to deal with outliers problem, in this paper, we establish a new cost function based on maximum correntropy criterion (MCC) and propose an accurate and robust correntropy-based OI (COI) pose estimation method. The proposed COI utilizes the advantages of correntropy to eliminate the bad effects of outliers, which can enhance the performance in the pose estimation problems with noise and outliers. In addition, our method does not need an extra outliers detection stage. Finally, we verify the effectiveness of our method in synthetic and real data experiments. Experimental results show that the COI can effectively combat outliers and achieve better performance than state-of-the-art algorithms, especially in the environments with a small number of outliers.

Keywords: Pose estimation · Orthogonal iterative (OI) algorithm · Maximum correntropy criterion (MCC).

1 Introduction

Pose estimation is widely used in augmented reality, robotics and navigation. It determines the orientation and position of the perspective camera through n known 3D reference points in the object coordinates and their imaging 2D points. Pose estimation is also known as the perspective- n -point (PnP) problem in computer vision, which has been studied for several decades.

In the past few years, there have been many methods to solve the PnP problem. One of the typical PnP methods is the orthogonal iterative (OI, also referred to as LHM) algorithm that calculates the orthogonal rotation matrix directly by minimizing the collinearity error in object space and has been proven to be globally convergent [21]. Later works apply the OI algorithm to different scenarios, such as augmented reality [22] and unmanned aerial vehicle systems [8] [28],

mainly studying how to provide a closer initial guess to obtain accurate pose estimation. In addition, other works improve this algorithm to adapt to complex noise environments. GOI [14] uses the inverse covariance matrix to describe the uncertainty of the feature points. SoftOI [9] introduces the Assign algorithm to determine the correspondences. Although these methods have a certain ability to suppress the influence of outliers, they are time-consuming.

Almost all of the above methods build their cost function based on the minimum mean square error (MMSE). MMSE can provide optimal solution under Gaussian noise, but it is sensitive to outliers. Recently, the maximum correntropy criterion (MCC) is shown to perform well in dealing with outliers [19]. MCC uses the characteristics of a kernel function to weaken the bad influence of outliers and non-Gaussian noise, ensuring the performance of the algorithm. In the present paper, we propose a correntropy-based OI (COI) method. The new method is robust to outliers without requiring the additional outliers detection stage. In particular, COI can achieve desirable performance for the case of small number of outliers, which is a very common situation in practical applications.

2 Related Works

For PnP problems with a small set of correspondences ($n \leq 5$), we can obtain a closed-form solution through P3P [15], P4P [3] or P5P [27] problems. These methods have higher calculation speed, but they are more sensitive to image noise and outliers. Generally, we roughly divide PnP methods into two categories: non-iterative and iterative.

For non-iterative methods, $EPnP$ [23] is an efficient solution that controls the computational complexity to $O(n)$ for the first time, and then Lepetit et al. [16] combined it with the Gauss-Newton algorithm to further improve accuracy. In addition, many non-iterative methods utilize polynomial solving techniques. For example, the direct least squares (DLS) [13] method analytically determines the roots of the polynomial system with the multiplication matrix. However, DLS applies Cayley-Gibbs-Rodriguez (CGR) to parameterize rotation, which would cause its degradation at 180 degrees [29]. In order to remove this limitation, $ASPnP$ [30] and $OPnP$ [29] use quaternion representation to parameterize the rotation and retrieve all stationary points through Gröbner basis technique.

Compared with non-iterative methods, iterative methods are usually more robust to noise. The classic iterative method is based on optimizing the image space collinearity error. In addition, Dementhon [7] proposed an iterative method (POSIT) based on proportional orthographic projection, but did not prove the convergence. Moreover, SoftPOSIT [6] improves POSIT to the unknown correspondences situations. SoftPOSIT can deal with a variety of noise data, but it is time-consuming because all matches need to be tried during the computing process. Later, Lu [21] developed a globally convergent OI algorithm that minimizes the object space collinearity error, which is the most accurate iterative method currently.

So far, most PnP methods suppose that image points are disturbed by Gaussian noise, which means 3D-2D correspondences do not contain outliers. However, image feature detection and matching methods, such as SIFT [20], SURF [2] and ORB [25], would be affected by illumination, scaling and occlusion, so that feature points may be mismatched in practical applications. These mismatches would seriously reduce the accuracy of pose estimation. The traditional solution combines the P3P algorithm with the RANSAC [12] scheme to reject outliers in a preliminary stage, and then executes another PnP method on the remaining inliers. However, the RANSAC scheme requires many trials to obtain the best results. Later, Ferraz et al. [11] proposed a very fast solution (REPP nP) to integrate outliers rejection schemes into the linear formula of the PnP solution. However, REPP nP is not suitable for small sets of correspondences [17]. Zhou et al. [31] proposed a robust 1-point RANSAC-based (R1PP nP) method, which introduces a soft weighting mechanism to distinguish between inliers and outliers by setting the inliers threshold. In addition, Li et al. [17] proposed a new cost function based on the l_q -norm for the PnP problem. This method can handle a high percentage of outliers, but it needs to provide a good initial value, which is difficult to achieve in practical applications. In recent years, many studies have shown that the correntropy can effectively deal with outliers, so we introduce MCC into the pose estimation algorithm. Compared with the above methods, our method can automatically suppress the effects of outliers and especially perform better in environments with a small number of outliers.

3 MCC-Based Pose Estimation

In this section, we briefly review the OI algorithm and MCC and propose a pose estimation algorithm based on correntropy and the OI algorithm.

3.1 Problem Formulation and OI Algorithm

Assume that the camera internal calibration is known, and then we give n non-collinear 3D points $\mathbf{p}_i = (x_i, y_i, z_i)^t, i = 1, \dots, n, n \geq 3$ in an object coordinate system and their corresponding camera coordinates $\mathbf{q}_i = (x_i, y_i, z_i)^t$. The relation between them can be written as:

$$\mathbf{q}_i = \mathbf{R}\mathbf{p}_i + \mathbf{t}, \quad (1)$$

where $\mathbf{R} = (\mathbf{r}_1^t, \mathbf{r}_2^t, \mathbf{r}_3^t)^t$ and $\mathbf{t} = (t_x, t_y, t_z)^t$ are the rotation matrix and the translation vector, respectively. The image point $\mathbf{v}_i = (u_i, v_i, 1)^t$ is the projection of \mathbf{p}_i on the normalized image plane. Under an ideal perspective camera model, the imaging relation can be expressed as:

$$\begin{cases} u_i = \frac{\mathbf{r}_1^t \mathbf{p}_i + t_x}{\mathbf{r}_3^t \mathbf{p}_i + t_z} \\ v_i = \frac{\mathbf{r}_2^t \mathbf{p}_i + t_y}{\mathbf{r}_3^t \mathbf{p}_i + t_z}, \end{cases} \quad (2)$$

or

$$\mathbf{v}_i = \frac{1}{\mathbf{r}_3^t \mathbf{p}_i + t_z} (\mathbf{R} \mathbf{p}_i + \mathbf{t}), \quad (3)$$

which is called the collinearity equation. Moreover, collinearity can be considered as the orthogonal projection of the camera coordinate point \mathbf{q}_i on \mathbf{v}_i should coincide with \mathbf{q}_i itself, which is expressed as follows:

$$\mathbf{R} \mathbf{p}_i + \mathbf{t} = \mathbf{V}_i (\mathbf{R} \mathbf{p}_i + \mathbf{t}), \quad (4)$$

where $\mathbf{V}_i = \mathbf{v}_i \mathbf{v}_i^t / (\mathbf{v}_i^t \mathbf{v}_i)$ is the line-of-sight projection matrix. Based on collinearity in object space, the OI algorithm formulates the pose estimation problem as minimizing the following sum of squared object-space collinearity error:

$$E(\mathbf{R}, \mathbf{t}) = \sum_{i=1}^n \|e_i\|^2 = \sum_{i=1}^n \|(\mathbf{I} - \mathbf{V}_i)(\mathbf{R} \mathbf{p}_i + \mathbf{t})\|^2. \quad (5)$$

Since the above objective function is quadratic in \mathbf{t} , we can assume that the rotation matrix \mathbf{R} is fixed, then the optimal value of \mathbf{t} can be obtained by:

$$\mathbf{t}(\mathbf{R}) = \frac{1}{n} \left(\mathbf{I} - \frac{1}{n} \sum_{i=1}^n \mathbf{V}_i \right)^{-1} \sum_{i=1}^n (\mathbf{V}_i - \mathbf{I}) \mathbf{R} \mathbf{p}_i. \quad (6)$$

We can compute the optimal \mathbf{R}^* iteratively through the following steps: First, suppose that $\mathbf{R}^{(k)}$ is the estimated value for the k -th iteration, so $\mathbf{t}^{(k)} = \mathbf{t}(\mathbf{R}^{(k)})$, $\mathbf{q}_i^{(k)} = \mathbf{R}^{(k)} \mathbf{p}_i + \mathbf{t}^{(k)}$. Then $\mathbf{R}^{(k+1)}$ is determined by solving the following absolute orientation problem [21]:

$$\mathbf{R}^{(k+1)} = \arg \min_{\mathbf{R}} \sum_{i=1}^n \left\| \mathbf{R} \mathbf{p}_i + \mathbf{t} - \mathbf{V}_i \mathbf{q}_i^{(k)} \right\|^2, \quad (7)$$

The translation is $\mathbf{t}^{(k+1)} = \mathbf{t}(\mathbf{R}^{(k+1)})$. This process is repeated until the termination condition is reached.

3.2 Maximum Correntropy Criterion

In previous studies, MCC has been successfully used in the field of signal processing, machine learning and point set registration [4] [5] [10]. Unlike MMSE, which is sensitive to complex noise, MCC has been proven to be robust to outliers [19]. In information theoretic learning, correntropy [24] is a nonlinear measure of similarity between two random variables X and Y :

$$V_\sigma(X, Y) = E[\kappa_\sigma(X, Y)], \quad (8)$$

where $\kappa_\sigma(\cdot)$ is a positive kernel function and σ is the kernel width. We take Gaussian kernel, which is the most widely used, as the kernel function in this paper:

$$\kappa_\sigma(x, y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right). \quad (9)$$

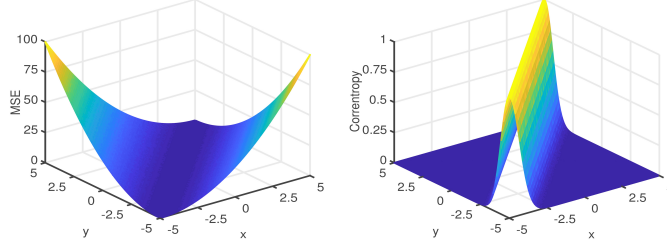


Fig. 1. Mean square error (MSE) (left) and correntropy (right) in the joint space

In practice, only a finite number of samples $(x_i, y_i)_{i=1}^N$ can be obtained, then correntropy can be estimated as:

$$\hat{V}_{N,\sigma}(X, Y) = \frac{1}{N} \sum_{i=1}^N \kappa_{\sigma}(x_i, y_i). \quad (10)$$

Usually, we hope that the estimated \hat{X} is as close to the desired X as possible. Therefore, it is necessary to maximize the correntropy between X and \hat{X} , as follows:

$$\arg \max_{\omega} \hat{V}_{\sigma}(X, \hat{X}), \quad (11)$$

which is called MCC, where ω is the unknown parameter. As shown in Fig. 1, MSE is a quadratic function in the joint space. For MSE, the value increases quadratically for the points away from the $x = y$ line, so it amplifies the influence of non-Gaussian noise and outliers on the mean value. However, correntropy has its maximum value along the $x = y$ line, and the value decreases exponentially away from this line. In other words, MCC considers more about the data near $x = y$ and has advantages in dealing with outliers.

3.3 Orthogonal Iterative Algorithm Based on MCC

It is well known that OI pose estimation algorithm is one of the most accurate PnP iterative methods. However, OI is based on MMSE, so it is not satisfactory in the situations with outliers. In order to solve this problem, we introduce MCC to improve the ability to deal with outliers. First, the correntropy-based cost function can be formulated as:

$$(\mathbf{R}^*, \mathbf{t}^*) = \arg \max_{\mathbf{R}, \mathbf{t}} \sum_{i=1}^n \exp\left(-\frac{\|(\mathbf{I} - \mathbf{V}_i)(\mathbf{R}\mathbf{p}_i + \mathbf{t})\|^2}{2\sigma^2}\right). \quad (12)$$

Notice that (12) is very similar to the form of (5). Inspired by the iterative process of OI algorithm [21], we assume that a fixed rotation matrix \mathbf{R} is given

and take the partial derivative of (12) with respect to \mathbf{t} . We have:

$$\frac{\partial \sum_{i=1}^n \exp(-\|e_i\|^2/2\sigma^2)}{\partial \mathbf{t}} = \sum_{i=1}^n g(e_i) \frac{(\mathbf{V}_i - \mathbf{I})(\mathbf{R}\mathbf{p}_i + \mathbf{t})(\mathbf{I} - \mathbf{V}_i)}{\sigma^2}, \quad (13)$$

where $g(e_i) = \exp(-\|(\mathbf{I} - \mathbf{V}_i)(\mathbf{R}\mathbf{p}_i + \mathbf{t})\|^2/2\sigma^2)$ and $e_i = (\mathbf{I} - \mathbf{V}_i)(\mathbf{R}\mathbf{p}_i + \mathbf{t})$.

Considering that $g(e_i)$ is the exponential term and changes little in the iteration process, we replace $\mathbf{R}^{(k-1)}$ with \mathbf{R} and $\mathbf{t}^{(k-1)}$ with \mathbf{t} , so the k -th iteration kernel term is $g(e_i) = g(\mathbf{R}^{k-1}, \mathbf{t}^{k-1})$. Setting (13) to zero, one can get a closed form solution of \mathbf{t} as follows:

$$\mathbf{t} = \left[\sum_{i=1}^n g(e_i)(\mathbf{I} - \mathbf{V}_i) \right]^{-1} \sum_{i=1}^n g(e_i)(\mathbf{V}_i - \mathbf{I})\mathbf{R}\mathbf{p}_i, \quad (14)$$

then (14) can be rewritten as:

$$\mathbf{t} = \left(\sum_{i=1}^n g(e_i)(\mathbf{V}_i - \mathbf{I})\mathbf{R}\mathbf{p}_i + \sum_{i=1}^n g(e_i)\mathbf{V}_i\mathbf{t} \right) / \sum_{i=1}^n g(e_i). \quad (15)$$

We substitute \mathbf{t} in the cost function (12) with (15) and define :

$$\begin{cases} \mathbf{q}_i(\mathbf{R}) = \mathbf{V}_i(\mathbf{R}\mathbf{p}_i + \mathbf{t}) \\ \mathbf{p}'_i = \mathbf{p}_i - \left(\sum_{i=1}^n g(e_i)\mathbf{p}_i \right) / \sum_{i=1}^n g(e_i) \\ \mathbf{q}'_i(\mathbf{R}) = \mathbf{q}_i(\mathbf{R}) - \left(\sum_{i=1}^n g(e_i)\mathbf{q}_i(\mathbf{R}) \right) / \sum_{i=1}^n g(e_i), \end{cases} \quad (16)$$

then we get the following simplified objective function:

$$(\mathbf{R}^*, \mathbf{t}^*) = \arg \max_{\mathbf{R}, \mathbf{t}} \sum_{i=1}^n \exp \left(-(\|\mathbf{R}\mathbf{p}'_i - \mathbf{q}'_i(\mathbf{R})\|^2)/2\sigma^2 \right). \quad (17)$$

Inspired by the singular value decomposition (SVD) method [1], we first construct a matrix \mathbf{M} :

$$\mathbf{M} = \sum_{i=1}^n \mathbf{q}'_i(\mathbf{R})g(e_i)\mathbf{p}_i^t, \quad (18)$$

which can be decomposed into $\mathbf{M} = \mathbf{U}\Sigma\mathbf{V}^t$. We calculate the estimation of $\mathbf{R}^{(k)}$ as follows:

$$\mathbf{R}^{(k)} = \mathbf{V}\tilde{\mathbf{I}}\mathbf{U}^t, \quad (19)$$

and determine $\tilde{\mathbf{I}}$ according to the value of $\det(\mathbf{V}\mathbf{U}^t)$. If $\det(\mathbf{V}\mathbf{U}^t) = 1$, we need $\tilde{\mathbf{I}} = \mathbf{I}_3$; and if $\det(\mathbf{V}\mathbf{U}^t) = -1$, we need $\tilde{\mathbf{I}} = \text{diag}(1, 1, -1)$. It can be seen from mathematics that when $\sigma \rightarrow \infty$, COI degenerates to the OI algorithm.

Here we make a summary of the method flow.

1. For a given set of 3D-2D points, the initial estimate of $\mathbf{R}^{(0)}$ can be computed using EPnP, then we can calculate $\mathbf{t}^{(0)}$ by (6).

2. We use the initial $\mathbf{R}^{(0)}$ and $\mathbf{t}^{(0)}$ to start the first iteration. \mathbf{R} and \mathbf{t} can be calculated iteratively from (19) and (14), respectively.

3. The second step needs to be repeated until the reprojection error converges or reaches the maximum number of iterations.

4 Experimental Results

In this section, we show the accuracy and stability of COI in both synthetic and real data situations. Our method is implemented in MATLAB.

4.1 Synthetic Experiments

In synthetic experiments, we compare our method against recent PnP solutions, including EPnP+GN [16], LHM [21], EPnP+LHM, RPnP [18], DLS [13], OPnP [29], ASPnP [30], SDP [26], EPPnP [11], REPPnP [11] and R1PPnP [31].

We assume a virtual well-calibrated camera with a resolution of 2048×1152 pixels, focal length of 1500 and principal point in the image center. Then 3D-2D correspondences are generated randomly. For the ordinary-3D case, the reference 3D points are uniformly distributed into the interval $[-2, 2] \times [-2, 2] \times [4, 8]$. For the quasi-singular case, the interval is $[1, 2] \times [1, 2] \times [4, 8]$. Finally, the ground-truth translation \mathbf{t}_{true} is set as the mean value of the reference 3D points. The ground-truth rotation matrix \mathbf{R}_{true} is generated randomly. For our method, the kernel width σ experimentally select from 0.1 to 2 with a step 0.5, besides, we use 10^5 to represent $\sigma \rightarrow \infty$. For the metric errors, the rotation error between \mathbf{R}_{true} and the estimated \mathbf{R} in degree is measured as (20), and then the translation error is defined as (21).

$$e_{rot}(\text{deg}) = \max_{k=1}^3 \arccos(\mathbf{r}_{k,true} \cdot \mathbf{r}_k) \times 180/\pi \quad (20)$$

$$e_{trans}(\%) = \|\mathbf{t}_{true} - \mathbf{t}\|/\|\mathbf{t}\| \times 100, \quad (21)$$

where $r_{k,true}$ and r_k are the k -th column of \mathbf{R}_{true} and \mathbf{R} , respectively.

All results of experiments are the average of 500 independent experiments.

Outlier-Free Synthetic Situations Most PnP methods can only deal with outlier-free situations. Their performance are significantly reduced with a small number of outliers. Although our method is mainly aimed at outliers, in order to verify the universality of COI, it is compared with other methods in outlier-free situations.

The first and second columns in Fig. 2 show the results of varying the number of correspondences n from 10 to 100. The Gaussian noise with a standard deviation of $\delta = 5$ pixels is added to image points. Then, in the second set of experiments, the number of correspondences is fixed as $n = 30$. We vary the level of Gaussian noise from 2 to 20 pixels, as shown in the third and fourth columns in Fig. 2. The ordinary case and quasi-singular case results indicate

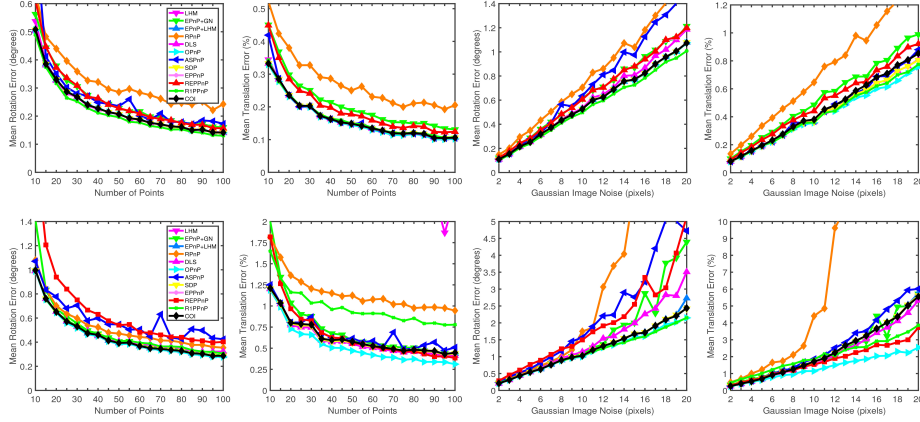


Fig. 2. Synthetic experiments without outliers, varying the number of corresponding points and the level of Gaussian noise, ordinary case (first row) and quasi-singular case (second row)

that some methods can accurately estimate the rotation matrix, and so does COI. However, the translation estimation of our method is not so accurate in quasi-singular case. LHM cannot handle quasi-singular case, while LHM+EPnP is still stable in the same situations. Furthermore, when using LHM to optimize the EPnP solution, the results have better accuracy than that of Gauss-Newton optimization. The accuracy of RPnP and ASPnP decreases significantly when the Gaussian noise is large.

Robustness to Outliers In this experiments, the number of correct matches between image points and object points is fixed as $n_{in} = 30$, then a Gaussian noise with standard deviation of $\delta = 10$ pixels is added to image points. In addition, we randomly generate corresponding points $n_{out} = n \times p_{out}$ as outliers, where p_{out} is the true percentage of outliers and $n = n_{in} + n_{out}$. For outliers, their image points are randomly distributed across the entire image plane. Fig. 3 shows the experimental results of varying the p_{out} from 5 to 50. The inliers threshold of R1PPnP is set as $h = 10$. Most PnP algorithms cannot correctly estimate the pose when there are outliers in the correspondences, only COI, REPPnP and R1PPnP maintain good accuracy.

In order to further demonstrate the robustness of our method to outliers and noise, we compare COI with REPPnP and R1PPnP. In addition, COI is also compared with some traditional methods, such as RANSAC+P3P+OPnP, RANSAC+P3P+ASPnP, RANSAC+P3P and RANSAC+RP4P+RPnP methods. For COI, the kernel width σ experimentally select from 0.1 to 1 with a step 0.1 and 1.2 to 3.5 with a step 0.2, besides, we use 10^5 to represent $\sigma \rightarrow \infty$.

Fig. 4 (a) shows the mean errors for increasing the percentage of outliers from 5 to 50. Then, the percentage of outliers p_{out} is fixed at $p_{out} = 20$. We add

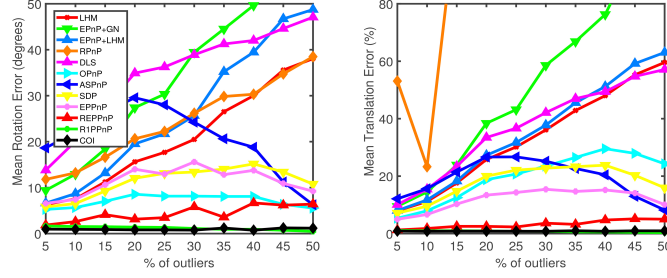


Fig. 3. Synthetic experiments without outliers, varying the percentage of outliers

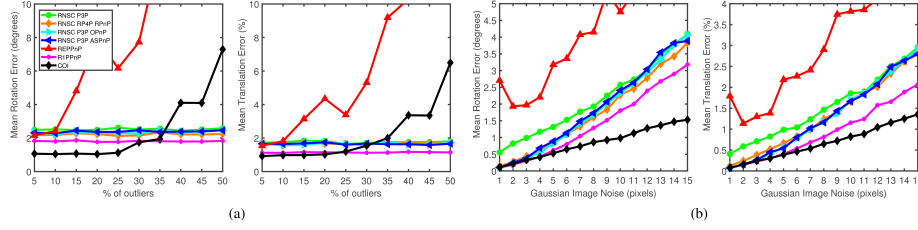


Fig. 4. Synthetic experiments with outliers, the number of inlier points was fixed as $n_{in} = 30$. (a) varying the percentage of outliers (b) varying the level of Gaussian noise

different levels of Gaussian image noise from 1 to 15, as shown in Fig. 4 (b). The performance of REPPnP is not as good as expected. This may be because in our experimental setting, the number of inliers is small, but REPPnP is more suitable in situations with a large number of correspondences. The RANSAC-based methods perform well, but they need to combine with another PnP method or set suitable inliers threshold to distinguish inliers and outliers, which is more complicated to use. However, COI can automatically suppress the impact of outliers and perform better than other methods when the percentage of outliers is less than 30%. Moreover, Fig. 4 (b) indicates that COI can achieve stronger robustness to Gaussian noise under the same outliers percentage.

4.2 Real Data Experiments

In real data experiments, we use PnP methods to determine the rotation and translation of the mixed reality device relative to the RGBD camera at a certain moment, and obtain the transformation between these two device coordinate systems. First, the RGBD data is obtained by an Intel RealSense D435 depth camera with an image resolution of 640×480 . Then we use a mixed reality device, the Microsoft HoloLens, to capture current scene image with a resolution of 2048×1152 . SIFT algorithm is used to compute image features and initial matches. Finally, the correspondences are obtained by brute force method. Cause the correspondences are affected by many external factors, there may be some mismatches in the results.

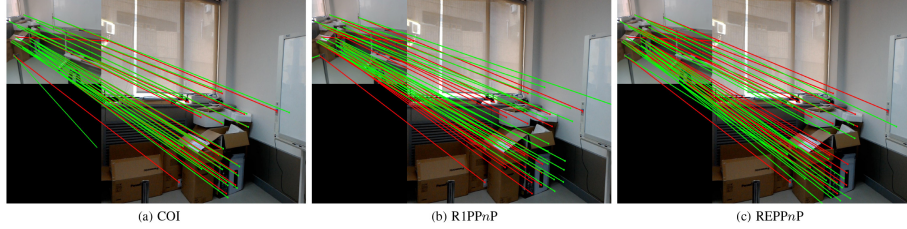


Fig. 5. Real image experiments, we use the estimated rotation and translation to re-project these correspondences. The pose estimation methods from left to right are COI, R1PPnP and REPPnP. (The red lines and green lines represent the original matches and reprojected matches, respectively.)

Fig. 5 shows a set of sample images, with 23 pairs of corresponding points and about 3 pairs of outliers. Outliers are basically due to missing depth data or mismatches. For R1PPnP, the inliers threshold is set as $h = 5$. The reprojection results show that the result of COI is closest to the original matches, which means that COI well overcomes the effects of outliers. However, REPPnP has a rather messy result after reprojection. Moreover, due to the effects of outliers, the results of R1PPnP have a tendency of collective translation.

5 Conclusion

The OI algorithm is one of the most accurate iterative pose estimation algorithms, but it is not robust enough to combat outliers. In this paper, we propose a robust pose estimation algorithm based on MCC to complement the shortcomings of the OI algorithm. The proposed COI utilizes its non-linear similarity measure and performs well when dealing with outliers and noise. Synthetic and real data experiments indicate that this method has ability to combat noise and performs better in the situations with a small number of outliers. These advantages are at the cost of introducing a free parameter σ . We will further study how to select this parameter in future work.

Acknowledgment. This work was supported by the National Natural Science Foundation of China (No. 61976175, 91648208) and Fundamental Research Funds for the Central Universities (No. xzy022020045).

References

1. Arun, K.S., Huang, T.S., Blostein, S.D.: Least-squares fitting of two 3D point sets. *IEEE Trans. Pattern Anal. Mach. Intell.* **9**(5), 698–700 (1987)
2. Bay, H., Tuytelaars, T., Van Gool, L.: SURF: Speeded up robust features. In: *Computer Vision – ECCV 2006*. pp. 404–417. Springer Berlin Heidelberg (2006). https://doi.org/10.1007/11744023_32

3. Bujnak, M., Kukelova, Z., Pajdla, T.: A general solution to the P4P problem for camera with unknown focal length. In: 2008 IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–8 (2008). <https://doi.org/10.1109/CVPR.2008.4587793>
4. Chen, B., Xing, L., Liang, J., Zheng, N., Príncipe, J.C.: Steady-state mean-square error analysis for adaptive filtering under the maximum correntropy criterion. *IEEE Signal Process. Lett.* **21**(7), 880–884 (2014)
5. Chen, B., Xing, L., Zhao, H., Zheng, N., Príncipe, J.C.: Generalized correntropy for robust adaptive filtering. *IEEE Trans. Signal Process.* **64**(13), 3376–3387 (2016)
6. David, P., Dementhon, D., Duraiswami, R., Samet, H.: SoftPOSIT: Simultaneous pose and correspondence determination. *Int. J. Comput. Vis.* **59**(3), 259–284 (2004)
7. Dementhon, D.F., Davis, L.S.: Model-based object pose in 25 lines of code. *Int. J. Comput. Vis.* **15**(1), 123–141 (1995)
8. Ding, M., Wei, L., Wang, B.: Vision-based estimation of relative pose in autonomous aerial refueling. *Chinese J. Aeronaut.* **24**(6), 807–815 (2011)
9. Dong, H., Sun, C., Zhang, B., Wang, P.: Simultaneous pose and correspondence determination combining softassign and orthogonal iteration. *IEEE Access* **7**, 137720–137730 (2019)
10. Du, S., Xu, G., Zhang, S., Zhang, X., Gao, Y., Chen, B.: Robust rigid registration algorithm based on pointwise correspondence and correntropy. *Recognit. Lett.* **132**, 91–98 (2020)
11. Ferraz, L., Binefa, X., Moreno-Noguer, F.: Very fast solution to the PnP problem with algebraic outlier rejection. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition. pp. 501–508. IEEE Computer Society (2014). <https://doi.org/10.1109/CVPR.2014.71>
12. Fischler, M.A., Bolles, R.C.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* **24**(6), 381–395 (1981)
13. Hesch, J.A., Roumeliotis, S.I.: A direct least-squares (DLS) method for PnP. In: 2011 International Conference on Computer Vision. pp. 383–390. IEEE (2011). <https://doi.org/10.1109/ICCV.2011.6126266>
14. Huo, J., Zhang, G., Cui, J., Yang, M.: A novel algorithm for pose estimation based on generalized orthogonal iteration with uncertainty-weighted measuring error of feature points. *J. Mod. Opt.* **65**(3), 331–341 (2018)
15. Kneip, L., Siegwart, R., Scaramuzza, D.: A novel parametrization of the perspective-three-point problem for a direct computation of absolute camera position and orientation. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition. pp. 2969–2976. IEEE Computer Society (2011). <https://doi.org/10.1109/CVPR.2011.5995464>
16. Lepetit, V., Moreno-Noguer, F., Fua, P.: EPnP: An accurate $O(n)$ solution to the PnP problem. *Int. J. Comput. Vis.* **81**(2), 155–166 (2009)
17. Li, J., Hu, Q., Zhong, R., Ai, M.: Exterior orientation revisited: A robust method based on l_q -norm. *Amer. Soc. Photogrammetry* **83**(1), 47–56 (2017)
18. Li, S., Xu, C., Xie, M.: A robust $O(n)$ solution to the perspective-n-point problem. *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(7), 1444–1450 (2012)
19. Liu, W., Pokharel, P.P., Principe, J.C.: Correntropy: Properties and applications in non-gaussian signal processing. *IEEE Trans. Signal Process.* **55**(11), 5286–5298 (2007)
20. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **60**(2), 91–110 (2004)

21. Lu, C.P., Hager, G.D., Mjolsness, E.: Fast and globally convergent pose estimation from video images. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(6), 610–622 (2000)
22. Ma, J., Zhou, Y., Liu, W., Hao, Q.: More improved robust orthogonal iterative algorithm for pose estimation in AR. In: Zhou, L. (ed.) *International Symposium on Photoelectronic Detection and Imaging 2007: Image Processing*. vol. 6623, pp. 642 – 652. International Society for Optics and Photonics, SPIE (2008), <https://doi.org/10.1117/12.791589>
23. Moreno-Noguer, F., Lepetit, V., Fua, P.: Accurate non-iterative $\mathcal{O}(n)$ solution to the PnP problem. In: *2007 IEEE 11th International Conference on Computer Vision*. pp. 1–8. IEEE (2007). <https://doi.org/10.1109/ICCV.2007.4409116>
24. Principe, J.C.: *Information Theoretic Learning: Renyi’s Entropy and Kernel Perspectives*. Springer, New York (2010)
25. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: ORB: An efficient alternative to sift or surf. In: *2011 International Conference on Computer Vision*. pp. 2564–2571 (2011). <https://doi.org/10.1109/ICCV.2011.6126544>
26. Schweighofer, G., Pinz, A.: Globally optimal $\mathcal{O}(n)$ solution to the PnP problem for general camera models. In: *Proceedings of the British Machine Vision Conference*. pp. 1–10. BMVA Press (2008). <https://doi.org/10.5244/C.22.55>
27. Triggs, B.: Camera pose and calibration from 4 or 5 known 3d points. In: *Proceedings of the Seventh IEEE International Conference on Computer Vision*. vol. 1, pp. 278–284 (1999). <https://doi.org/10.1109/ICCV.1999.791231>
28. Zhang, W., Xu, G., Cheng, Y., Liu, T.: Research on orthogonal iteration algorithm of visual pose estimation for UAV landing. In: *2018 IEEE CSAA Guidance, Navigation and Control Conference (CGNCC)*. pp. 1–6. IEEE (2018). <https://doi.org/10.1109/GNCC42960.2018.9019144>
29. Zheng, Y., Kuang, Y., Sugimoto, S., Åström, K., Okutomi, M.: Revisiting the PnP problem: A fast, general and optimal solution. In: *2013 IEEE International Conference on Computer Vision*. pp. 2344–2351. IEEE (2013). <https://doi.org/10.1109/ICCV.2013.291>
30. Zheng, Y., Sugimoto, S., Okutomi, M.: ASPnP: An accurate and scalable solution to the perspective-n-point problem. *IEICE Trans. Inf. Syst.* **96**(7), 1525–1535 (2013)
31. Zhou, H., Zhang, T., Jagadeesan, J.: Re-weighting and 1-point ransac-based PnP solution to handle outliers. *IEEE Trans. Pattern Anal. Mach. Intell.* **41**(12), 3022–3033 (2019)