A NEW EM ALGORITHM FOR 2D-3D POINT CLOUD REGISTRATION WITH PROBABILISTIC DATA ASSOCIATION

Younes Boutiyarzist[1,2,3], *Samy Labsir*[2,4], *Jean-Yves Tourneret*[2,5], *François Vincent*[1] *and Philippe Salmon*[3]

[1] ISAE-SUPAERO, Univ. Toulouse, Toulouse, France. firstname.lastname@isae-supaero.fr
[2] TéSA laboratory, Toulouse, France. firstname.lastname@tesa.prd.fr
[3] Collins Aerospace, Blagnac, France. firstname.lastname@collins.com
[4] IPSA, Toulouse, France. firstname.lastname@ipsa.fr
[5] IRIT-ENSEEIHT, Univ. Toulouse, Toulouse, France. firstname.lastname@toulouse-inp.fr

ABSTRACT

This work studies a new Expectation-Maximization (EM) algorithm for solving the 2D-3D registration problem, which consists of estimating the position and orientation of a camera using a 3D map and a 2D image of the same scene. This algorithm associates each image feature coordinate to one vector of the 3D map using the pinhole camera model or to a class of outliers, making the registration robust to the presence of abnormal image features. It iteratively improves the camera pose by estimating the associations between the image features and the 3D map coordinates (using a robust mixture model) and minimizing the reprojection errors between the image and map points. Experimental results demonstrate that the proposed EM algorithm achieves competitive results in both absolute position and orientation compared to the Iterative Closest Point (ICP) approach.

Index Terms— robust estimation, camera pose estimation, 2D-3D registration, mixture model.

1. INTRODUCTION

Registration of 2D-3D point clouds, commonly known as camera pose estimation, consists of estimating the position and orientation of a camera relative to a reference frame by analyzing image features and their correspondences with 3D reference points. This task is critical for applications such as augmented reality [1] [2], 3D reconstruction [3] [4], robotics [5] [6] and medical imagery [7] [8]. In the aerospace domain, camera pose estimation plays a critical role in navigation and automated landing systems, enabling precise alignment and control in dynamic and challenging environments [9].

In controllable environment or/and good weather condition, camera pose estimation involves three key steps: detecting image features or markers [10–12], establishing correspondences between these features and known 3D reference points [13], and solving the Perspective-n-Point (PnP) problem to compute the camera's position and orientation [14]. The PnP problem is classically addressed by minimizing the reprojection error between the 3D points in the world reference frame and their 2D correspondences in the image.

In uncontrollable environments or adverse weather conditions, feature associations can become ambiguous or unreliable, significantly limiting the effectiveness of traditional registration methods [15] [16]. Leveraging prior knowledge of a 3D model, such as road positions (and/or fixed landmarks) that can be generated through the fusion of different data sources (e.g., LiDAR and radar), can improve the robustness and accuracy of the camera pose estimation. In this context, the association between the detected image features and the 3D model is unknown, and feature detectors usually generate outliers making this association complicated. Iterative Closest Point (ICP) algorithms [8] provide a solution for correspondence-free registration problems. However, point clouds are never perfectly aligned, especially when they are generated by different sensors, leading to inaccuracies due to noise or false/out-of-date data [17].

To address the challenges related to 2D-3D registration, this paper studies a new Expectation-Maximization (EM) algorithm that performs a robust probabilistic association between the two point clouds. It is structured as follows: Section 2 introduces a statistical model for camera pose estimation in the presence of outliers and derives an EM algorithm for estimating its parameters via maximum likelihood (ML) estimation. Section 3 evaluates the performance of the resulting robust EM algorithm using various experiments on synthetic data. Conclusions and future works are reported in Section 4.

2. PROPOSED EM ALGORITHM

2.1. Problem Formulation

Consider two point clouds of the same scene. The first point cloud consists of n 2D noisy position vectors detected in an image, denoted as $\boldsymbol{x}_1^i \in \mathbb{R}^2$, where $i = 1, \ldots, n$. The second point cloud consists of m 3D position vectors denoted as $\boldsymbol{x}_2^j \in \mathbb{R}^3$, where $j = 1, \ldots, m$. When the association between these two point clouds is known, the relationship between each vector $\boldsymbol{x}_1^i$ and its 3D point $\boldsymbol{x}_2^j$ can be defined using a pinhole camera model [3]:

$$\begin{bmatrix} \boldsymbol{x}_1^i \\ 1 \end{bmatrix} = \begin{bmatrix} u_1^i \\ v_1^i \\ 1 \end{bmatrix} = \frac{\mathbf{K}\left(\mathbf{R}\boldsymbol{x}_2^j + \mathbf{t}\right)}{\left(\mathbf{R}\boldsymbol{x}_2^j + \mathbf{t}\right)_3} + \begin{bmatrix} \boldsymbol{e}_{i,j} \\ 0 \end{bmatrix}, \tag{1}$$

where the camera intrinsic matrix is defined as

$$\mathbf{K} = \begin{bmatrix} \alpha_x & 0 & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{bmatrix},$$

with α_x and α_y the focal lengths in pixels along the x and y directions and (x_0, y_0) the principal point. Note that the notation $(.)_3$ refers to the third component of a vector. The orientation and position of the camera are defined by a matrix $\boldsymbol{R} \in \mathbb{SO}(3)$ (the Special Orthogonal Group) and $\boldsymbol{t} \in \mathbb{R}^3$ that need to be estimated. The model error $\boldsymbol{e}_{i,j}$ is supposed to follow a Gaussian distribution $\mathcal{N}(\mathbf{0}_2, \sigma^2 \boldsymbol{I}_2)$, with zero mean and covariance matrix $\sigma^2 \boldsymbol{I}_2$.

2.2. A robust registration model for 2D-3D registration

This section introduces a 2D-3D registration model for cases where the associations between the 2D and 3D points are unknown and potential outliers are present in the scene. A uniform distribution defined in the image area is introduced to account for the presence of outliers resulting from the detection algorithm [18]:

$$p(\boldsymbol{x}_1^i|\text{outlier}) = \frac{1}{\Delta}\mathbb{1}_I(\boldsymbol{x}_1^i), \tag{2}$$

where Δ denotes the image area in square pixels and $\mathbb{1}_I$ is the indicator function defined on the image area I.

To tackle the data association problem, inspired by [17], we introduce a matrix $\boldsymbol{A} \in \{0,1\}^{n\times(m+1)}$, where each element $a_{i,j}$ indicates the association between the detected 2D feature $\boldsymbol{x}_1^i$ and the 3D reference point $\boldsymbol{x}_2^j$, i.e., $a_{i,j}=1$ when $\boldsymbol{x}_1^i$ is associated with $\boldsymbol{x}_2^j$, $a_{i,j}=0$ if $\boldsymbol{x}_1^i$ and $\boldsymbol{x}_2^j$ are not associated (with $i,j=1,...,m$) and $a_{i,m+1}=1$ if $\boldsymbol{x}_1^i$ is an outlier in the first point cloud. Since each 2D point of the first point cloud is associated with at most one 3D point of the second point cloud, each row of $\boldsymbol{A}$ satisfies the constraint $\sum_{j=1}^{m+1} a_{ij}=1$. The prior distribution for the different associations is defined as follows:

$$P(a_{i,m+1}=1)=\rho_o,\ P(a_{i,j}=1)=\frac{1-\rho_o}{m}, j=1,...,m, \tag{3}$$

where m is the number of 3D points, and ρ_o is the proportion of outliers in the set of measurements $\boldsymbol{x}_1$. This prior reflects a non-informative distribution for the feature associations (i.e., each 2D feature $\boldsymbol{x}_1^i$ has the same probability of being asssociated with any of the m 3D points $\boldsymbol{x}_2^j$), and ρ_o is the outlier probability, which has to be adjusted by the user to control the robustness of the model.

The registration problem consists of estimating the rotation matrix $\boldsymbol{R}$ and the translation vector $\boldsymbol{t}$ (that are related to the orientation and position of the camera), i.e., $\boldsymbol{\Theta}=(\boldsymbol{R},\boldsymbol{t})$, from the 2D measurements $\boldsymbol{X}_1=\{\boldsymbol{x}_1^1,...,\boldsymbol{x}_1^n\}$ and the 3D vectors $\boldsymbol{X}_2=\{\boldsymbol{x}_2^1,...,\boldsymbol{x}_2^m\}$ for an unknown latent association matrix $\boldsymbol{A}$.

2.3. Likelihood and Complete Likelihood

The marginal likelihood of the proposed model is defined as:

$$\mathcal{L}(\boldsymbol{X_1}|\boldsymbol{\Theta},\boldsymbol{X_2}) = \sum_{\boldsymbol{A}\in\boldsymbol{\Psi}} p\left(\boldsymbol{X_1},\boldsymbol{A}\mid\boldsymbol{\Theta},\boldsymbol{X_2}\right), \tag{4}$$

where $\boldsymbol{\Psi}$ denotes the set of all valid association matrices. The total number of valid associations is $(m+1)^n$, which makes the MLE computationally intractable. To address this challenge, we investigate a new EM algorithm to estimate $\boldsymbol{\Theta}$ based on the so-called complete likelihood defined as:

$$\begin{aligned}\mathcal{L}_c(\boldsymbol{X_1},\boldsymbol{A}|\boldsymbol{\Theta},\boldsymbol{X_2}) &= p(\boldsymbol{X_1}\mid\boldsymbol{A},\boldsymbol{X_2},\boldsymbol{\Theta})P(\boldsymbol{A}\mid\boldsymbol{X_2},\boldsymbol{\Theta})\\ &= \prod_{i=1}^{n}\prod_{j=1}^{m}\left[p(\boldsymbol{x}_1^i|\boldsymbol{x}_2^j,\boldsymbol{\Theta})P(a_{i,j})\right]^{a_{i,j}}\\ &\times\prod_{i=1}^{n}\left[p(\boldsymbol{x}_1^i|\text{outlier})P(a_{i,m+1})\right]^{a_{i,m+1}}.\end{aligned} \tag{5}$$

2.4. EM algorithm

The EM algorithm is an iterative approach used to compute maximum likelihood estimates in models with latent variables. The algorithm alternates between Expectation (E) and Maximization (M) steps defined from the complete likelihood [19]. For the proposed model, the $(t+1)$-th iteration is defined as:

- E-step: Compute $Q(\boldsymbol{\Theta}\mid\boldsymbol{\Theta}^{(t)})$, the expected value of the complete log-likelihood given the observed data and the current parameter estimate $\boldsymbol{\Theta}^{(t)}$:

$$Q(\boldsymbol{\Theta}\mid\boldsymbol{\Theta}^{(t)}) = \mathbb{E}_{\boldsymbol{A}|\boldsymbol{X_1},\boldsymbol{X_2},\boldsymbol{\Theta}^{(t)}}\left[\log\mathcal{L}_c(\boldsymbol{X_1},\boldsymbol{A}\mid\boldsymbol{\Theta},\boldsymbol{X_2})\right], \tag{6}$$

- M-step: Update the parameter estimate by maximizing $Q(\boldsymbol{\Theta}\mid\boldsymbol{\Theta}^{(t)})$ with respect to $\boldsymbol{\Theta}$:

$$\boldsymbol{\Theta}^{(t+1)} = \arg\max_{\boldsymbol{\Theta}} Q(\boldsymbol{\Theta}\mid\boldsymbol{\Theta}^{(t)}). \tag{7}$$

The complete log-likelihood for the proposed model results from (5):

$$\begin{aligned}\log\mathcal{L}_c(\boldsymbol{X_1},\boldsymbol{A}\mid\boldsymbol{\Theta},\boldsymbol{X_2}) &= \sum_{i=1}^{n}\sum_{j=1}^{m} a_{i,j}\log p(\boldsymbol{x}_1^i\mid\boldsymbol{x}_2^j,\boldsymbol{\Theta})\\ &+\sum_{i=1}^{n} a_{i,m+1}\log p(\boldsymbol{x}_1^i\mid\text{outlier})\\ &+\sum_{i=1}^{n}\sum_{j=1}^{m} a_{i,j}\log\left(\frac{1-\rho_o}{m}\right)+\sum_{i=1}^{n} a_{i,m+1}\log(\rho_0).\end{aligned} \tag{8}$$

E-Step (Expectation)

A ray casting algorithm [20] is used to determined the set of visible features of the 3D map $\boldsymbol{X_2}$ at a given camera pose $\boldsymbol{\Theta}^{(t)}$. According to Bayes theorem, the conditional distribution of $a_{i,j}\mid\boldsymbol{x}_1^i,\boldsymbol{x}_2^j,\boldsymbol{\Theta}^{(t)}$ can be defined for any $(i,j)\in\{1,...,n\}\times\{1,...,m\}$ as:

$$\gamma_{i,j}^{(t)} = P(a_{i,j}=1\mid\boldsymbol{x}_1^i,\boldsymbol{x}_2^j,\boldsymbol{\Theta}^{(t)}), \tag{9}$$

with

$$\gamma_{i,j}^{(t)} = \frac{p(\boldsymbol{x}_1^i\mid\boldsymbol{x}_2^j,\boldsymbol{\Theta}^{(t)})P(a_{i,j}=1)}{\sum_{j=1}^{m} p\left(\boldsymbol{x}_1^i\mid\boldsymbol{x}_2^j,\boldsymbol{\Theta}^{(t)}\right)P(a_{i,j}=1)+\frac{1}{\Delta}P(a_{i,m+1}=1)}. \tag{10}$$

Moreover, the outlier probabilities are defined for $i=1,...,n$ as:

$$\gamma_{i,m+1}^{(t)} = \frac{\frac{1}{\Delta}P(a_{i,m+1}=1)}{\sum_{j=1}^{m} p\left(\boldsymbol{x}_1^i\mid\boldsymbol{x}_2^j,\boldsymbol{\Theta}^{(t)}\right)P(a_{i,j}=1)+\frac{1}{\Delta}P(a_{i,m+1}=1)}. \tag{11}$$

Simple computations resulting from (6), (8) and (10) lead to:

$$Q(\boldsymbol{\Theta}\mid\boldsymbol{\Theta}^{(t)}) = \sum_{i=1}^{n}\sum_{j=1}^{m}\gamma_{i,j}^{(t)}\log p(\boldsymbol{x}_1^i\mid\boldsymbol{x}_2^j,\boldsymbol{\Theta})+K, \tag{12}$$

where K is a constant independent of $\boldsymbol{\Theta}$.

M-Step (Maximization)

The M-step maximizes the function $Q(\boldsymbol{\Theta}\mid\boldsymbol{\Theta}^{(t)})$ defined in (12) with respect to $\boldsymbol{\Theta}=(\boldsymbol{R},\boldsymbol{t})$. This maximization problem reduces to minimizing the weighted reprojection error, i.e.,

$$\arg\min_{\boldsymbol{\Theta}} \overbrace{\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{\gamma_{i,j}}{\sigma^2}\underbrace{\|\boldsymbol{x}_1^i-\pi(\boldsymbol{R}^{(t)}\boldsymbol{x}_2^j+\boldsymbol{t}^{(t)})\|^2}_{r_{i,j}^2(\boldsymbol{\Theta}^{(t)})}}^{L(\boldsymbol{\Theta}^{(t)})}, \tag{13}$$

where π is the projection operation appearing in (1), L denotes the loss function, and $r_{i,j}$ is the residual corresponding to the i,j-th association, leading to the vector of residuals $\boldsymbol{r} = [r_{1,1}, ..., r_{n,m}]$. Given the nonlinear nature of this problem, we adopt a variant of the EM algorithm known as gradient EM, whose convergence has been investigated in many works [21–23]. This approach replaces the M-step of the EM algorithm by one iteration of a gradient algorithm such as the Gauss-Newton algorithm. For the cost function defined in (13), the Gauss-Newton algorithm updates $\boldsymbol{R}$ and $\boldsymbol{t}$ as follows:

$$\boldsymbol{\Theta}^{(t+1)} = \boldsymbol{\Theta}^{(t)} + \Delta\boldsymbol{\Theta}. \tag{14}$$

with

$$\Delta\boldsymbol{\Theta} = -(\boldsymbol{J}^\top \boldsymbol{J})^{-1} \boldsymbol{J}^\top \boldsymbol{r}, \tag{15}$$

where $\boldsymbol{J}$ is the Jacobian matrix of the reprojection residuals with respect to the parameters $\boldsymbol{R}$ and $\boldsymbol{t}$. In this work, the matrix $\boldsymbol{R}$ has been parametrized using the Rodrigues formula [24], which allows the matrix $\boldsymbol{J}$ to be computed. The Gauss-Newton iterations are repeated until convergence, i.e., when $\|\Delta\boldsymbol{\Theta}\|$ is below a predefined threshold τ, or until a maximum number of iterations n_{max} is reached, yielding Algorithm 1.

Algorithm 1 EM Algorithm

Input: map, $\boldsymbol{X_1}$, σ^2, ρ_o, n_{max}, τ, $\boldsymbol{\Theta}^{(0)}$
Output: $\boldsymbol{\Theta}^{(f)}$
while $n_{\text{it}} < n_{\text{max}}$ **and** $\|\Delta\boldsymbol{\Theta}\| > \tau$ **do**
 $\boldsymbol{X_2} \leftarrow$ ray-casting $\left(\text{map}, \boldsymbol{\Theta}^{(t-1)}\right)$
 Expectation step
 for $(i,j) \in \{1, ..., n\} \times \{1, ..., m\}$ **do**
  $\gamma_{i,j} \leftarrow P(a_{i,j} = 1 \mid \boldsymbol{x}_1^i, \boldsymbol{x}_2^j, \boldsymbol{\Theta}^{(t-1)})$ ▷ (10)
 end for
 Maximization step
 for $(i,j) \in \{1, ..., n\} \times \{1, ..., m\}$ **do**
  $r_{i,j} \leftarrow r_{i,j}(\boldsymbol{\Theta}^{(t-1)})$ ▷ (13)
 end for
 $\boldsymbol{\Theta}^{(t)} \leftarrow$ Gauss-Newton $\left(\boldsymbol{r}, \boldsymbol{\Theta}^{(t-1)}\right)$
 $n_{\text{it}} \leftarrow n_{\text{it}} + 1$
end while
$\boldsymbol{\Theta}^{(f)} \leftarrow \boldsymbol{\Theta}^{(t)}$

3. SIMULATION RESULTS

Several experiments were conducted to assess the performance of the proposed EM algorithm. This algorithm requires to adjust three key components: the initial camera pose $\boldsymbol{\Theta}^{(0)} = (\boldsymbol{R}^{(0)}, \boldsymbol{t}^{(0)})$, the noise variance σ^2 (that is related to the uncertainty of the measurements) and the probability of outliers ρ_o. The initial pose $\boldsymbol{\Theta}^{(0)}$ can be obtained from a Kalman filter prediction, with or without incorporating additional sensors. The user must specify σ^2 and ρ_o, which depend on the characteristics of the detection algorithm.

This section studies a crossroad scenario illustrated in Fig. 1. The camera is mounted on a drone and positioned at $[120, 200, 60]$ meters along the x, y, and z axes, with an orientation described by Euler angles $[0°, -60°, -170°]$ representing rotations around these axes. Fig. 1 shows the 3D map with road positions represented as a sparse point cloud. A ray-casting algorithm has been used to determine which points are visible or not by the camera (indicated by blue and black crosses). The corresponding 2D point cloud has been acquired by the fixed camera providing the detections displayed in Fig. 2 (blue and red points).

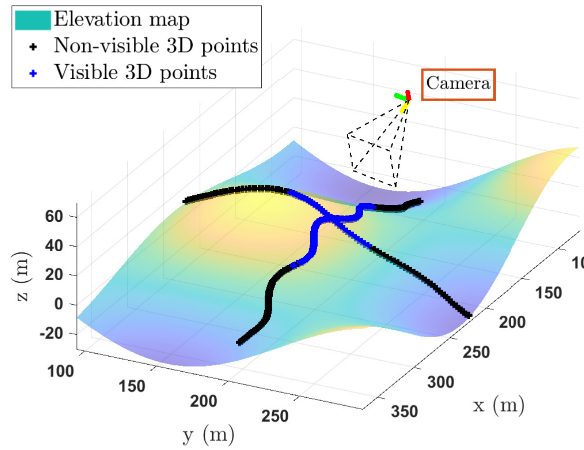


Fig. 1. 3D point cloud for a road map. Blue and black crosses indicate visible and non visible road points from the camera's pose.

3.1. ICP algorithm

The ICP algorithm will be used as a benchmark to solve the 2D-3D point cloud registration problem investigated in this paper. The reprojection error for this algorithm is defined as:

$$\sum_{i=1}^{n} \min_{j=1,...,m} \frac{1}{\sigma^2} \|\boldsymbol{x}_1^i - \pi(\boldsymbol{R}\boldsymbol{x}_2^j + \boldsymbol{t})\|^2, \tag{16}$$

where the reprojection error is computed by finding the closest match for each 3D point. To ensure robustness against outliers, a RANSAC-based approach [25] is incorporated, making the registration process not strongly corrupted by erroneous correspondences or noise. The matrix $\boldsymbol{R}$ and translation $\boldsymbol{t}$ are updated using the Gauss-Newton method, as in the M-step of the EM algorithm.

3.2. Synthetic Dataset

This section compares the performance of the EM and ICP algorithms. The evaluation is conducted on a dataset generated with $n = 200$ nominal 2D observations (blue points), a noise variance of $\sigma^2 = 25$, and an outlier probability of $\rho_o = 10\%$. The initial pose for both algorithms is fixed at a position $\boldsymbol{t} = [125, 195, 65]$ with orientations $\boldsymbol{\phi} = [3°, -57°, -167°]$, as shown in Fig. 2. For the ICP algorithm, the number of iterations was fixed empirically to 100. For the EM algorithm, the maximum number of iterations, n_{max}, and the convergence threshold, τ, are set to 100 and 10^{-3}.

Outliers are detected using the posterior probability of association to the outlier class defined in (11), i.e., a 2D point i is classified as an outlier if its posterior probability is such that $\gamma_{i,m+1} > 0.5$. Figure 2 illustrates the inliers (blue points), outliers (points with orange circles), and the 2D projections corresponding to the estimated poses obtained from both algorithms (green and purple crosses for ICP and EM). Table 1 provides quantitative results in terms of position and orientation mean square errors (MSEs) showing that the EM algorithm is very competitive when compared to ICP.

Fig. 3 finally shows typical evolutions of the loss function (13) and the absolute errors of the pose parameters ((t_x, t_y, t_z) for the position and (ϕ_x, ϕ_y, ϕ_z) for orientation) versus the number of iterations, illustrating the algorithm convergence.

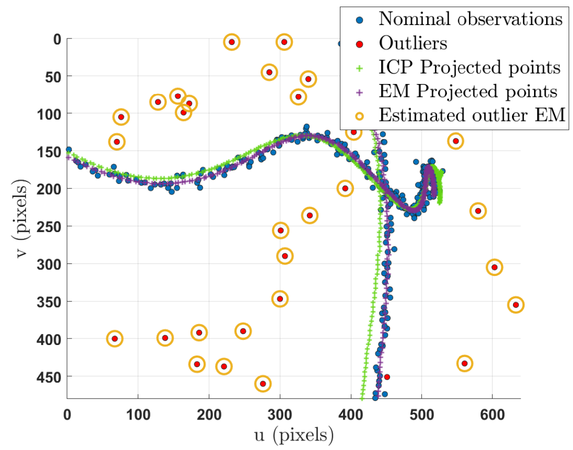


Fig. 2. Comparison between EM and ICP algorithms. Blue and red points indicate the inliers and outliers forming the 2D point cloud. The green and purple crosses are the projections of 3D points using the estimated poses of ICP $\hat{\boldsymbol{\Theta}}_{\text{ICP}}$ and EM $\hat{\boldsymbol{\Theta}}_{\text{EM}}$. Points circled in orange are identified as outliers by the EM algorithm.

Table 1. MSEs for position and orientation

Algorithm	**Position**	**Orientation**
EM	1.82×10^{-2}	2.65×10^{-2}
ICP	13.08	4.15

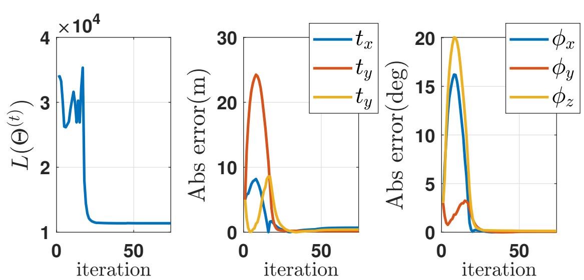


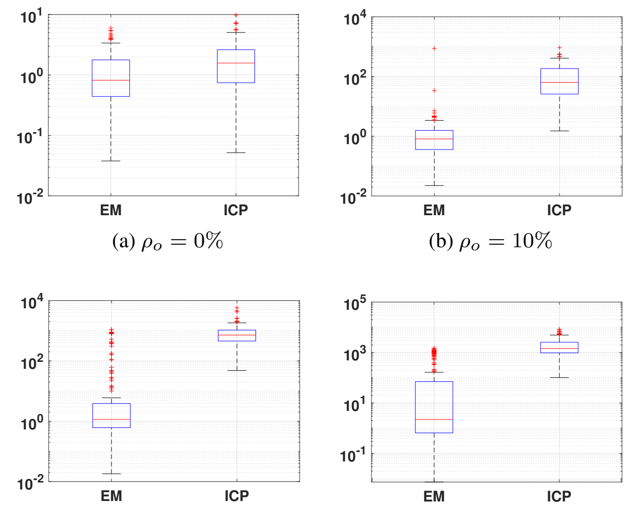
Fig. 3. Loss L and absolute errors for each pose parameter versus the number of iterations of the EM algorithm.

3.3. Monte-Carlo simulations

Monte-Carlo simulations allow the performance of the proposed algorithm to be quantified when varying the hyperparameters. The data were generated as in Section 3.2. A first experiment evaluates the robustness of the proposed EM algorithm to the probability of outliers ρ_o in comparison with RANSAC ICP. Fig. 4 shows that the proposed EM algorithm provides lower MSEs of $\boldsymbol{\Theta}$ than ICP for all values of ρ_o. The higher the proportion of outliers, the larger the performance gap between EM and ICP.

The performance of the EM algorithm for different noise variances is displayed in Fig. 5a (data generated without outliers), confirming that the EM algorithm is very competitive when compared to ICP. The influence of the initial pose $\boldsymbol{\Theta}^{(0)}$ on the EM algorithm is shown in Fig. 5b, showing the MSE of the estimator of Θ for a range of initial poses around the ground truth. The closer the initial pose to the ground truth, the more accurate the estimated pose, as expected.

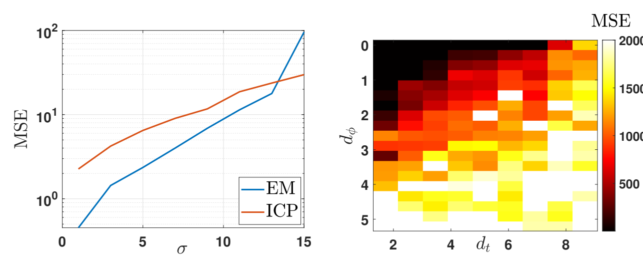
Finally, it is interesting to note that the execution time of the proposed EM algorithm is much smaller than for RANSAC-ICP, by a factor close to 80 (the total execution times for the two algorithms are $t_{\text{EM}} = 0.31$s and $t_{\text{ICP}} = 25.35$s).



(a) $\rho_o = 0\%$ (b) $\rho_o = 10\%$

(c) $\rho_o = 30\%$ (d) $\rho_o = 50\%$

Fig. 4. Boxplots for the MSEs of the vector of positions and orientations $\boldsymbol{\Theta} = (\boldsymbol{t}, \boldsymbol{\phi})$ for different values of ρ_o.



(a) MSE versus σ (b) Influence of $\boldsymbol{\Theta}^{(0)}$

Fig. 5. MSE of the estimates of $(\boldsymbol{t}, \boldsymbol{\phi})$ versus σ (left) and (d_ϕ, d_t) (right), where d_ϕ and d_t are the Euclidean distances between the values of $\boldsymbol{\phi}$ and $\boldsymbol{t}$ for initial and ground truth poses.

4. CONCLUSION

This paper studied a new robust EM algorithm for the registration of 2D and 3D point clouds. A mixture model was considered to model the presence of outliers, and a latent association matrix was introduced to solve the association problem between the two point clouds. The proposed EM algorithm requires to adjust three key parameters: the probability of outliers, the noise variance, and the initial pose. Experimental results demonstrated that the EM algorithm is very competitive when compared to the vanilla ICP algorithm. Including extensions of the ICP algorithm [26] in the comparison would be interesting, even if the generalization of these algorithms to 2D-3D registration is challenging. Future work includes automatic hyperparameter estimation, developing an online EM algorithm for real-time applications and deriving Rao-Cramer bounds to evaluate the optimal estimation performance.

5. REFERENCES

[1] E. Marchand, H. Uchiyama, and F. Spindler, "Pose estimation for augmented reality: a hands-on survey," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, no. 12, pp. 2633–2651, 2015.

[2] F. Ababsa and M. Mallem, "Robust camera pose estimation using 2D fiducials tracking for real-time augmented reality systems," in *Proc. Int. Conf. Virtual-Real. Contin. Appl. Ind. (VRCAI)*, (Singapore, Singapore), 2004.

[3] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, UK: Cambridge university press, 2000.

[4] J. L. Schonberger and J.-M. Frahm, "Structure-From-Motion Revisited," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit. (CVPR)*, (Las Vegas, USA), June 2016.

[5] T. Qin, P. Li, and S. Shen, "VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator," *IEEE Trans. Robot.*, vol. 34, no. 4, pp. 1004–1020, 2018.

[6] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. M. Montiel, and J. D. Tardós, "ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual–Inertial, and Multimap SLAM," *IEEE Trans. Robot.*, vol. 37, no. 6, pp. 1874–1890, 2021.

[7] U. Mitrović *et al.*, "3D–2D registration in endovascular image-guided surgery: evaluation of state-of-the-art methods on cerebral angiograms," *Int. J. Comput. Assis. Radiol. Surg.*, vol. 13, no. 2, pp. 193–202, 2018.

[8] J. Song and M. Ghaffari, "DynaWeightPnP: Toward global real-time 3D-2D solver in PnP without correspondences," *arXiv preprint arXiv:2409.18457*, 2024.

[9] P. Veneruso *et al.*, "Vision-aided approach and landing through AI-based vertiport recognition," in *Proc. Int. Conf. Unmanned Aircraft Syst. (ICUAS)*, (Warsaw, Poland), 2023.

[10] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, vol. 2, pp. 1150–1157, 1999.

[11] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-Up Robust Features (SURF)," *Comput. Vis. Image Underst.*, vol. 110, no. 3, pp. 346–359, 2008.

[12] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, pp. 2564–2571, 2011.

[13] M. Muja and D. G. Lowe, "Fast Matching of Binary Features," in *Proc. Conf. Comput. Robot Vis.*, (Toronto, Canada), 2012.

[14] V. Lepetit, F. Moreno-Noguer, and P. Fua, "EPnP: An accurate O(n) solution to the PnP problem," *Int. J. Comput. Vis.*, vol. 81, Feb. 2009.

[15] C. Valgren and A. J. Lilienthal, "SIFT, SURF & seasons: Appearance-based long-term localization in outdoor environments," *Robot. Auton. Syst.*, vol. 58, no. 2, pp. 149–156, 2010.

[16] S. Duan *et al.*, "Influence of Fog Weather on Automotive Vision Target Detection," in *Proc. Int. Conf. Veh. Control Intell. (CVCI)*, (Nanjing, China), 2022.

[17] G. Agamennoni *et al.*, "Point Clouds Registration with Probabilistic Data Association," in *Proc. Int. Conf. Intell. Robots Syst. (IROS)*, (Daejeon, South Korea), 2016.

[18] J. Lesouple *et al.*, "Robust Hypersphere Fitting from Noisy Data Using an EM Algorithm," in *Proc. European Signal Processing Conference (EUSIPCO)*, pp. 1–5, 2021.

[19] A. P. Dempster *et al.*, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *J. Roy. Stat. Soc. Ser. B (Methodol.)*, vol. 39, no. 1, pp. 1–38, 1977.

[20] S. D. Roth, "Ray casting for modeling solids," *Comput. Graph. and Image Process.*, vol. 18, no. 2, pp. 109–144, 1982.

[21] K. Lange, "A Gradient Algorithm Locally Equivalent to the EM Algorithm," *J. Roy. Stat. Soc. Ser. B (Methodol.)*, vol. 57, pp. 425–437, Dec. 2018.

[22] S. Balakrishnan, M. J. Wainwright, and B. Yu, "Statistical guarantees for the EM algorithm: From population to sample-based analysis," *Ann. Stat.*, vol. 45, no. 1, pp. 77 – 120, 2017.

[23] W. Xu, M. Fazel, and S. S. Du, "Toward global convergence of gradient em for over-parameterized gaussian mixture models," in *Proc. Neural Inf. Process. Syst. (NeurIPS)*, (Vancouver, Canada), 2024.

[24] J. S. Dai, "Euler–Rodrigues formula variations, quaternion conjugation and intrinsic connections," *Mech. Mach. Theory*, vol. 92, pp. 144–152, 2015.

[25] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, p. 381–395, June 1981.

[26] A. Segal, D. Hähnel, and S. Thrun, "Generalized-ICP," in *Proc. Robot.: Sci. Syst.*, (Seattle, USA), 06 2009.