# Accurate Mix-Norm-Based Scan Matching

**Conference Paper** · October 2018

DOI: 10.1109/IROS.2018.8594278

CITATIONS
8

READS
290

**7 authors**, including:

Di Wang
Xi'an Jiaotong University
**25** PUBLICATIONS **43** CITATIONS
SEE PROFILE

Jianru Xue
Xi'an Jiaotong University
**177** PUBLICATIONS **1,208** CITATIONS
SEE PROFILE

Tao Zhongxing
Xi'an Jiaotong University
**9** PUBLICATIONS **8** CITATIONS
SEE PROFILE

Shaoyi Du
Xi'an Jiaotong University
**119** PUBLICATIONS **886** CITATIONS
SEE PROFILE

2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
Madrid, Spain, October 1-5, 2018

# Accurate Mix-Norm-Based Scan Matching

Di Wang, Jianru Xue†, Zhongxing Tao, Yang Zhong, Dixiao Cui,
Shaoyi Du, and Nanning Zheng, *IEEE Fellow*

***Abstract*— Highly accurate mapping and localization is of prime importance for mobile robotics, and its core lies in efficient scan matching. Previous research are focusing on designing a robust objective function and the residual error distribution is often ignored or simply assumed as unitary or mixture of simple distributions. In this paper, a mixture of exponential power (MoEP) distributions is proposed to approximate the residual error distribution. The objective function induced by MoEP-based residual error modelling ensembles a mix-norm-based scan matching (MiNoM), which enhances the matching accuracy and convergence characteristic. Both the parameters of transformation (rotation and translation) and residual error distribution are estimated efficiently via an EM-like algorithm. The optimization of MiNoM is iteratively achieved via two phases: An on-line parameter learning (OPL) phase to learn residual error distribution for better representation according to the likelihood field model (LFM), and an iteratively reweighted least squares (IRLS) phase to attain transformation for accuracy and efficiency. Extensive experimental results validate that the proposed MiNoM outperforms several state-of-the-art scan matching algorithms in both convergence characteristic and matching accuracy.**

## I. INTRODUCTION

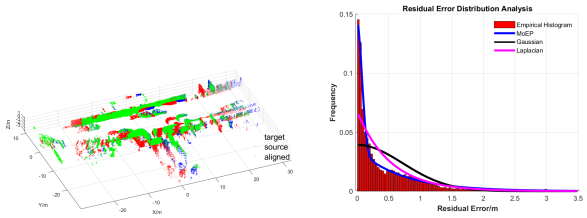
In recent decades, extensive research efforts have been put into exploring highly accurate mapping and localization techniques for mobile robot applications, especially in autonomous driving[1][2]. Without loss of generality, the problem of mapping and localization is to attain the transformation matrices among different coordinate systems, i.e. finding the best transformation hypothesis to align the current local observation with previous observations or a map. This process is also referred as scan matching in fields of mobile robotics.

Numerous algorithms and techniques have been investigated for the scan matching problem, and its core lies in how to deal with partially missing information caused by limitation of sensor's coverage or occlusion and outliers incurred by dynamic objects.

To address the aforementioned issues, many algorithms focus on designing an objective funtion which can prune unreliable data directly, weight data based on matching quality, or introduce sparsity. Several variants of iterative closest point (ICP)[3] are proposed, in which the residual error is based on distance of point-to-point[4], point-to-plane[5], or plane-to-plane[6][7], with correspondences given. Commonly, the distribution of residual error is assumed to be a

*This work is supported by National Natural Science Foundation of China (NSFC) projects 61751308 and 61773311.

*Authors are with Visual Cognitive Computing and Intelligent Vehicle (VCC&IV) Lab, Xi'an Jiaotong University, Xi'an, P.R. China. Corresponding author's email: jrxue@xjtu.edu.cn




(a) (b)

Fig. 1. Modelling residual error distribution accurately is essential to improve the robustness of scan matching. (a) illustrates point clouds from real scenes. (b) illustrates the empirical distribution for the residual error. Here the residual error is defined as point-to-point Euclidean distance in Eq(2).

Gaussian distribution, and the objective function is therefore formulated as sum of squares. Xu et al. [8] proposed a correntropy-based ICP (CICP), which uses maximum correntropy criteria (MCC) as the objective function. Bergstrom et al. [9] utilized robust M-estimation techniques to limit the influence of outliers. Instead of modelling residual error distribution as a Gaussian distribution, Bouaziz et al. [10] proposed a sparse ICP (SICP), which introduces a sparsity-inducing penalty, i.e. imposing a power $p$ $(0 < p \leq 1)$ to the original residual error. It should be noted that the objective function of SICP can be explained as $L_p$-based M-estimator, or the residual error obeys a distribution belonging to the exponential family. The outlier handling mechanism of these algorithms mainly lies in correspondence determination or designing robust distance or weighting, which may lack of efficacy when portion of partially missing information or outliers is large.

To overcome the limitation of modelling residual error as a single distribution, Magnusson et al. [11] proposed normal distribution transformation (NDT), which assumes scan registration as a generative process, i.e. the source data are generated from a stochastic distribution. In NDT's original formulation[12], a mixture of a Gaussian distribution and a uniform distribution is utilized to account for outliers. Wolcott et al. [13] reported a fast multi-resolution scan matcher, which formulated the objective function by mixing several Gaussian distributions with a uniform distribution. However, the parameters of these mixture distributions are fixed or hand-tuned.

The aforementioned scan matching algorithms[3][10][11] can be cast in the framework of residual error modelling, and their difference lies in the assumption of residual error distribution: Gaussian, Laplacian or other distributions be-

longing to the exponential family. However, the distribution of residual error is far from Gaussian or Laplacian in real scenes as illustrated in Fig. 1. Robust scan matching requires accurate residual error modelling techniques.

Recently, several algorithms claim that better performance can be achieved by robust residual error modelling. The parameters of the residual error distribution used are seeked via optimization. Yang et al. [14] assumed a robust but implicit distribution for residual error, and proved that their method was much more robust against outliers. Cao et al. [15] analyzed the residual error of hyper spectral image rigorously, and proposed to model residual error as a mixture of exponential power (MoEP) distributions. The augmented Lagrange multipliers (ALM) is then employed to reconstruct the image.

In this paper, we propose a mixed-norm-based scan matching (MiNoM) algorithm. Our contributions in this paper are three folds: 1) A mixture of EP distributions is employed to approximate the residual error distribution, which is the first time used for scan matching problem according to our knowledge. 2) An efficient EM-like agorithm is proposed to solve both the parameters of residual error distribution and transformation. The algorithm includes an on-line parameter learning (OPL) and an iteratively reweighted least squares (IRLS)[16]. 3) The proposed MiNoM is applied to localization in GPS-denied environments, and achieves satisfying accuracy for autonomous navigation. Extensive experiments show that the proposed MiNoM gains better convergence characteristic and matching accuracy compared with several state-of-the-art scan matching algorithms.

## II. Residual Error Modelling

### A. Problem Formulation

The scan matching problem is often formulated as seeking a maximum of the likelihood function

$$\mathbf{x}^* = \arg\max_{\mathbf{x}} p\left(\mathbf{Q}|\mathbf{x},\mathbf{M}\right), \tag{1}$$

where $p\left(\mathbf{Q}|\mathbf{x},\mathbf{M}\right)$ is the likelihood function. The measurements $\mathbf{Q}$ and the map $\mathbf{M}$ are two point sets, i.e. $\mathbf{M} = \{\mathbf{m}_i\}_{i=1}^{N_m}, \mathbf{Q} = \{\mathbf{q}_i\}_{i=1}^{N_q}, \mathbf{m}_i, \mathbf{q}_i \in \mathbb{R}^D$, where $D = 2$ or 3 corresponding to 2D or 3D. The transformation $\mathbf{x}$ can be decomposed into a rotation and a translation $(\mathbf{R},\mathbf{t}) = \mathrm{Tf}\left(\mathbf{x}\right)$, where $\mathbf{R} \in \mathrm{SO}\left(D\right), \mathbf{t} \in \mathbb{R}^D$.

To compute the likelihood function, two primary steps are executed[17]:

*1) Computing Residual Error:* Given a transformation $(\mathbf{R},\mathbf{t})$, the correspondence can be determined via nearest neighbor searching, and the residual error $\mathbf{e}$ is calculated subsequently. Specifically, for each measurement point $\mathbf{q}_i$, its nearest Euclidean distance $e_i$ to map $\mathbf{M}$ is

$$\begin{aligned} c\left(i\right) &= \arg\min_j \left\|\mathbf{R}\mathbf{q}_i + \mathbf{t} - \mathbf{m}_j\right\|_2^2, \\ e_i &= \left\|\mathbf{R}\mathbf{q}_i + \mathbf{t} - \mathbf{m}_{c(i)}\right\|_2, \\ \mathbf{e} &= \{e_i\}_{i=1}^{N_q}, \end{aligned} \tag{2}$$

where $\mathbf{e}$ denotes residual error.

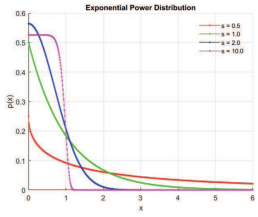



Fig. 2. Exponential power distributions with different shape parameters. Since the residual error is defined as Euclidean distance in Eq(2), the domain of EP is non-negative.

*2) Calculating Likelihood:* Assuming independence of each measurement, the likelihood $p\left(\mathbf{Q}|\mathbf{x},\mathbf{M}\right)$ is calculated

$$p\left(\mathbf{Q}|\mathbf{x},\mathbf{M}\right) = \prod_{i=1}^{N_q} p\left(\mathbf{q}_i|\mathbf{x},\mathbf{M}\right) = \prod_{i=1}^{N_q} p\left(e_i|\boldsymbol{\theta},\boldsymbol{\pi}\right), \tag{3}$$

where $p\left(e_i|\boldsymbol{\theta},\boldsymbol{\pi}\right)$ denotes the residual error distribution, and it is often assumed as a mixture of distributions in [17], $\boldsymbol{\theta}$ denotes the parameter vector of the distribution, and $\boldsymbol{\pi}$ denotes the weight vector of each component distribution. $p\left(e_i|\boldsymbol{\theta},\boldsymbol{\pi}\right)$ depicts noise characteristics of LiDAR and environment.

Though the parameters $(\boldsymbol{\theta},\boldsymbol{\pi})$ are assumed to be fixed in the original likelihood field model (LFM), they can also be optimized in the residual error modelling. The scan matching now can be formulated as

$$\left(\mathbf{R}^*,\mathbf{t}^*,\boldsymbol{\pi}^*,\boldsymbol{\theta}^*\right) = \arg\max_{\mathbf{R},\mathbf{t},\boldsymbol{\pi},\boldsymbol{\theta}} \prod_{i=1}^{N_q} p\left(e_i|\boldsymbol{\theta},\boldsymbol{\pi}\right), \tag{4}$$

where $e_i = \left\|\mathbf{R}\mathbf{q}_i + \mathbf{t} - \mathbf{m}_{c(i)}\right\|_2$.

### B. Revisit Residual Error Modelling of Point Set Registration

With a given correspondence $c(i)$ and $p\left(e|\boldsymbol{\theta},\boldsymbol{\pi}\right)$ is assummed as a single exponential power (EP) distribution (more specific details of EP are given in APPENDIX), $p\left(\mathbf{Q}|\mathbf{x},\mathbf{M}\right)$ is exactly the objective function of the standard ICP[4] or the sparse ICP[10].

The efficacy of scan matching algorithm like standard ICP or sparse ICP can be inferred from EP's probability density function as illustrated in Fig. 2. For standard ICP, the residual error distribution corresponds to a Gaussian distribution, which severely restricts the magnitude of residual error. Thus the standard ICP is sensitive to partially missing information or outliers. For sparse ICP, the shape parameter $s$ in residual error distribution is less than 1.0, and the corresponding probability density function has a thick tail which can tolerate large magnitude of residual error. Though SICP is much robust to outliers, it is sensitive to the initial transformation since the $s < 1.0$ introduces local minimals to its objective function.

Unfortunately, the real scenes as illustrated in Fig. 1(a) are abundant with partially missing information or outliers, and a good initial transformation is hard to find. Therefore, more reliable residual error modelling is required for scan matching algorithms.

### C. Residual Error Modelling using MoEP

To model the residual error distribution accurately, we employ a mixture of EP (MoEP) distributions. This is validated in Fig. 1(b), where the empirical residual error distribution can be well approximated by MoEP

$$\begin{aligned} &p\left(e|\boldsymbol{\theta},\boldsymbol{\pi}\right) \sim \sum_{k=1}^{K} \pi_k EP\left(e|\theta_k, s_k\right), \\ &s.t. \quad \pi_k \geq 0, \sum_{k=1}^{K} \pi_k = 1, \end{aligned} \tag{5}$$

where $\mathbf{s}$ and $\boldsymbol{\theta}$ denote shape/precision parameter's vector (SPV/PPV), which collects each EP's shape/precision parameter in the MoEP, and $K = |\mathbf{s}|$ denotes the length of SPV. Note that SPV should be given as a prior knowledge.

The MoEP-based residual error modelling is robust, but may lead to difficulties in solve Eq(4). Latent variable $\gamma_{ik}$ is introduced to make the problem tractable. $\gamma_{ik}$ indicates whether $e_i$ is from the $k^{th}$ component of the mixture or not,

$$\begin{aligned} &\gamma_{ik} = \{0, 1\}, \\ &s.t. \sum_{k=1}^{K} \gamma_{ik} = 1. \end{aligned} \tag{6}$$

With the latent variable, the residual error distribution can be easily written as

$$p\left(e_i|\boldsymbol{\theta},\boldsymbol{\pi}\right) = p\left(\boldsymbol{\gamma}_i\right) \prod_{k=1}^{K} p(e_i|\theta_k, s_k)^{\gamma_{ik}}, \tag{7}$$

where $p\left(\boldsymbol{\gamma}_i\right) = \prod_{k=1}^{K} \pi_k^{\gamma_{ik}}$ denotes the probability of latent variable $\boldsymbol{\gamma}_i = \{\gamma_{ik}\}_{k=1}^{K}$.

Using Eq(3) and Eq(7), the log likelihood function becomes

$$\begin{aligned} &L\left(\mathbf{R},\mathbf{t},\boldsymbol{\pi},\boldsymbol{\theta}\right) = \log\left(\prod_{i=1}^{N_q} p\left(e_i|\boldsymbol{\theta},\boldsymbol{\pi}\right)\right) \\ &= \sum_{k=1}^{K} \omega_k \left(\log \pi_k + \frac{1}{s_k}\log\theta_k\right) - \sum_{i=1}^{N_q}\sum_{k=1}^{K} \gamma_{ik}\theta_k e_i^{s_k} + const \end{aligned}, \tag{8}$$

where $\omega_k = \sum_{i=1}^{N_q} \gamma_{ik}$ is the weight.

With the introduction of MoEP-based residual error modelling, the Eq(4) can be written as

$$\left(\mathbf{R}^*,\mathbf{t}^*,\boldsymbol{\pi}^*,\boldsymbol{\theta}^*\right) = \underset{(\mathbf{R},\mathbf{t},\boldsymbol{\pi},\boldsymbol{\theta})}{\arg\max}\, L\left(\mathbf{R},\mathbf{t},\boldsymbol{\pi},\boldsymbol{\theta}\right). \tag{9}$$

Scan matching algorithm based on Eq(9) is referred as mix-norm-based scan matching (MiNoM). By configuring SPV $\mathbf{s}$ properly, the mix-norm-based objective function in Eq(8) can take both advantages of standard ICP and sparse ICP.

## III. EM-like Algorithm

### A. Algorithm Overview

To maximize the likelihood function in Eq(8), we propose an EM-like algorithm. E step and M step are executed iteratively: given initial parameters $\left(\mathbf{R}^{(t)},\mathbf{t}^{(t)},\boldsymbol{\pi}^{(t)},\boldsymbol{\theta}^{(t)}\right)$, the latent variable $\gamma_{ik}^{(t+1)}$ is calculated in E step, and $\left(\mathbf{R}^{(t+1)},\mathbf{t}^{(t+1)},\boldsymbol{\pi}^{(t+1)},\boldsymbol{\theta}^{(t+1)}\right)$ is optimized in M step.

**Algorithm 1** EM-like Algorithm

**Input:** Map $\mathbf{M}$, measurements $\mathbf{Q}$, SPV $\mathbf{s}$, initial transformation $\left(\mathbf{R}^{(0)},\mathbf{t}^{(0)}\right)$.
**Output:** $(\mathbf{R},\mathbf{t})$ and $(\boldsymbol{\pi},\boldsymbol{\theta})$.

```
for r = 1:1:r_max do
    Calculating c(i) and e^(r) via Eq(2).
    Calculating π^(r), θ^(r), γ_ik^(r) via Algorithm 2.
    Calculating (R^(r), t^(r)) via Algorithm 3.
end for
return (R, t) and (π, θ).
```

**Algorithm 2** On-line Parameter Learning (OPL)

**Input:** Residual error $\mathbf{e}$, SPV $\mathbf{s}$.
**Output:** $\boldsymbol{\pi}, \boldsymbol{\theta}, \gamma_{ik}$.

```
Initialize (π^(0), θ^(0)) randomly.
for r = 1:1:r_max do
    E-step via Eq(12) given (π^(r-1), θ^(r-1)).
    M-step via maximizing Eq(13) to obtain (π^(r), θ^(r)).
end for
return π, θ, γ_ik.
```

However, the M step involving Eq(8) is highly non-linear and difficult to optimize. Instead of solving $(\mathbf{R},\mathbf{t},\boldsymbol{\pi},\boldsymbol{\theta})$ directly, the EM-like algorithm solves the followed two sub-problems:

$$\left(\boldsymbol{\pi}^{(t+1)},\boldsymbol{\theta}^{(t+1)}\right) = \underset{\boldsymbol{\pi},\boldsymbol{\theta}}{\arg\max}\, L\left(\mathbf{R}^{(t)},\mathbf{t}^{(t)},\boldsymbol{\pi},\boldsymbol{\theta}\right) \tag{10}$$

corresponds to the on-line parameter learning, and

$$\left(\mathbf{R}^{(t+1)},\mathbf{t}^{(t+1)}\right) = \underset{\mathbf{R},\mathbf{t}}{\arg\max}\, L\left(\mathbf{R},\mathbf{t},\boldsymbol{\pi}^{(t+1)},\boldsymbol{\theta}^{(t+1)}\right) \tag{11}$$

corresponds to the the transformation calculation.

The above two optimization sub-problems can be solved efficiently. The whole algorithm is listed in Algorithm 1.

### B. On-line Parameter Learning

The parameters $(\boldsymbol{\pi},\boldsymbol{\theta})$ depict noise characteristics. Thus it is reasonable to learn accurate $(\boldsymbol{\pi},\boldsymbol{\theta})$ before updating transformation $(\mathbf{R},\mathbf{t})$. The on-line parameter learning can be interpreted as fitting a MoEP distributions given current residual error $\mathbf{e}$ and SPV $\mathbf{s}$. The complete on-line parameter learning algorithm is listed in Algorithm 2. OPL consists of mainly two steps:

*1) E Step:* Calculating latent variables via Bayesian rule given $(\boldsymbol{\pi},\boldsymbol{\theta})$

$$\gamma_{ik} = \frac{\pi_k p\left(e_i|\theta_k, s_k\right)}{\sum_{k=1}^{K} \pi_k p\left(e_i|\theta_k, s_k\right)}. \tag{12}$$

**Algorithm 3** Calculating Transformation via IRLS

**Input:** Map $\mathbf{M}$, measurements $\mathbf{Q}$, SPV $\mathbf{s}$, latent variable $\gamma_{ik}$, PPV $\boldsymbol{\theta}$, and correspondence $c(i)$.
**Output:** $(\mathbf{R}, \mathbf{t})$.

```
Set w^(0) = 1.
for r = 1 : 1 : r_max do
    Update (R^(r), t^(r)) via Eq(15).
    Update weights w^(r) via Eq(16).
end for
return (R, t).
```

*2) M Step:* Calculating parameters of residual error distribution given $\gamma_{ik}$ by seeking a maximum of Eq(8). To solve it, the Lagrangian function

$$\mathcal{L}(\boldsymbol{\pi}, \boldsymbol{\theta}, \lambda) = L(\mathbf{R}, \mathbf{t}, \boldsymbol{\pi}, \boldsymbol{\theta}) + \lambda \left( \sum_{k=1}^{K} \pi_k - 1 \right) \tag{13}$$

is constructed, and the parameters of residual error distribution are easily obtained by setting the derivatives of $\mathcal{L}(\boldsymbol{\pi}, \boldsymbol{\theta}, \lambda)$ with respect to $(\boldsymbol{\pi}, \boldsymbol{\theta})$ equal to zero.

### C. Calculating Transformation via IRLS

After obtaining parameters of residual error distribution, the transformation can be obtained by solving the following problem

$$\begin{aligned} (\mathbf{R}^*, \mathbf{t}^*) &= \arg\min_{\mathbf{R},\mathbf{t}} \sum_{i=1}^{N_q} \sum_{k=1}^{K} \gamma_{ik} \theta_k e_i^{s_k}, \\ e_i &= \left\| \mathbf{R}\mathbf{q}_i + \mathbf{t} - \mathbf{m}_{c(i)} \right\|_2^{s_k}. \end{aligned} \tag{14}$$

According to [4], Eq(14) has analytical solution when SPV $\mathbf{s}$ contains only 2.0 and the problem reduces to original point-to-point ICP algorithm. However, it becomes highly nonlinear when provided with other SPV. Observing the structure of Eq(14), IRLS is a feasible choice.

The complete IRLS algorithm is listed in Algorithm 3, and it iteratively runs two primary step:

*1) Updating Transformation:*

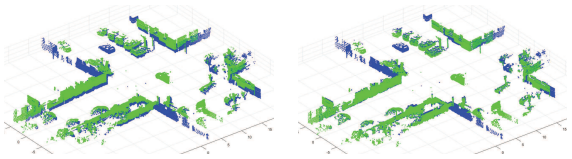
$$\left( \mathbf{R}^{(r)}, \mathbf{t}^{(r)} \right) = \arg\min_{\mathbf{R},\mathbf{t}} \sum_{i=1}^{N_q} \omega_i^{(r-1)} \left\| \mathbf{R}\mathbf{q}_i + \mathbf{t} - \mathbf{m}_{c(i)} \right\|_2^2. \tag{15}$$

The above weighted least square (WLS) problem can be solved efficiently via singular value decomposition (SVD)[4].

*2) Updating Weights:*

$$\begin{aligned} e_i^{(r)} &= \max\left( \left\| \mathbf{R}^{(r)}\mathbf{q}_i + \mathbf{t}^{(r)} - \mathbf{m}_{c(i)} \right\|_2, \varepsilon \right), \\ \omega_i^{(r)} &= \sum_{k=1}^{K} \gamma_{ik} \theta_k \left( e_i^{(r)} \right)^{s_k - 2}, \end{aligned} \tag{16}$$

where the small positive number $\varepsilon = 0.0001$ is introduced for numerical stability, when residual error $e_i^{(r)}$ is close to zero.



(a) (b)

Fig. 3. Groundtruth pose may be wrong for some frames in KITTI. (a) illustrates alignment using pose provided by KITTI, and the ratio score is 37.5%. (b) illustrates alignment using proposed MiNoM, and the ratio score is 78.5%. The two frames are from sequence indexed by 07, and their distance is around 3.0 meter.

## IV. Experimental Results and Discussion

In this section, the proposed MiNoM is verified via matching of 3D point clouds from real traffic scenes and compared with several state-of-the-art scan matching algorithms. The convergence and matching accuracy are extensively analyzed in GPS available and GPS-denied scenes. Furthermore, the sparsity analysis offers us an insight of why the proposed MiNoM gains better performance than other scan matching algorithms. In the attached video, the detailed process of residual error modelling in MiNoM and several localization applications using MiNoM are presented.

### A. Dataset

*1) KITTI Dataset:* The KITTI[18] dataset provides high definition point cloud captured from Velodyne-64E (10Hz), and accurate pose data from high-end GNSS. Representative urban scenes are chosen from KITTI dataset, which include sequences indexed by 00, 02, and 07 from "Visual Odometry / SLAM Evaluation 2012". To produce frames of spatially overlapped point clouds, each sequence is sampled at 3.0 meters according to the groundtruth pose data. The typical LiDAR data from KITTI are illustrated in Fig. 3.

By using ratio score introduced in the followed subsection, we found that groundtruth poses in KITTI may be in low quality for some frames as illustrated in Fig. 3. These poor groundtruth poses (quite few frames in each sequence) are corrected manually before comparing scan matching algorithms.

*2) Underground Garage Dataset:* Accurate GPS signal is not always available in real traffic scenes, if considering some typical urban scenes which include garage, tree-abundant campus and high building urban districts. To validate scan matching algorithms in GPS-denied environments, a typical underground garage around $60m \times 60m$ is chosen. An accurate 3D occupancy map is built via GraphSLAM[19] and is illustrated in Fig. 5(a).

### B. Pre-processing

For each point cloud in KITTI or underground garage, preprocessings are done before applying scan matching: 1) Each point cloud is segmented into ground and obstacle points, and only obstacle points are reserved for scan matching, 2) the obstacle points are further random sampled at a rate of 10%, which produces around 5,000 to 8,000 3D points for each point cloud. 10% is empirically selected since we found that

this percent is enough to attain a satisfying matching, and a higher sampling percent promotes negligible improvements. The intuitive behind ground removal is due to the scanning mode of 3D LiDAR sensor. Because Velodyne-64E weights more lower angle and captures a large portion of ground points, and these ground points provide few geometrical information than obstacle points. However, it should be noted that the rotation and translation are still computed in 3D even ground points are removed.

### C. Implementation Details

The implementation of point-to-point ICP[4], point-to-plane ICP[5], generalized ICP[6] and normal distribution transformation[11] are from Point Cloud Library (PCL)[1], which are denoted as p2pICP, p2plICP, GICP, and NDT, respectively. For NDT there exists another implementation maintained by its authors (http://wiki.ros.org/perception_oru), but we use PCL implementation in this paper for consistency. The sparse ICP (SICP)[10] is employed with source code from its authors[2]. The correntropy ICP[8] (CICP) and the proposed MiNoM are implemented in MATLAB R2017a. All the experiments are performed with a laptop with Intel Core i7 processor and 16G RAM.

For NDT, the grid size is set as 1.0 meter. For CICP, the selection of scale factor $\sigma$ is quite important. After discussing with its authors, $\sigma = 1.0$ meter is a proper choice which obtains satisfying results. For SICP, the norm is selected as $p = 0.4, 0.8, 1.0$. The maximal iteration of all the aforementioned scan matching algorithms is set to be 300 for a fair comparison.

Two configurations of SPV are both introduced, MiNoM0 is provided with $\mathbf{s} = \begin{bmatrix} 1.0 & 2.0 \end{bmatrix}$, and MiNoM1 is provided with $\mathbf{s} = \begin{bmatrix} 0.5 & 1.0 & 2.0 \end{bmatrix}$.

### D. Performance Evaluation

We choose two scores for fair evaluation comparison of the proposed MiNoM with several other scan matching algorithms.

*1) Percent Score:* For two point clouds, assuming their groundtruth transformation $(\mathbf{R}_g, \mathbf{t}_g)$ is obtained from GNSS, and the estimated transformation from scan matching is $(\mathbf{R}_e, \mathbf{t}_e)$, the error metrics[20] are defined as

$$\begin{aligned} \Delta\mathbf{R} &= \mathbf{R}_g^{-1}\mathbf{R}_e, \\ \varepsilon_{\mathbf{R}} &= \cos^{-1}\left(\frac{\mathbf{Tr}(\Delta\mathbf{R})-1}{2}\right), \\ \varepsilon_{\mathbf{t}} &= \|\mathbf{t}_e - \mathbf{t}_g\|_2, \end{aligned} \tag{17}$$

where $\mathbf{Tr}()$ denotes a trace operator on square matrix. Considering the accuracy of GNSS in KITTI, a match is classified as high accuracy if

$$\begin{aligned} \varepsilon_{\mathbf{R}} &\leq 0.5^\circ, \\ \varepsilon_{\mathbf{t}} &\leq 0.1m. \end{aligned} \tag{18}$$

For a sequence, the percent score is defined as

$$percent = N_{good}/N_{total}, \tag{19}$$

[1] http://pointclouds.org/documentation/tutorials/registration_api.php.

[2] https://github.com/opengp/sparseicp.

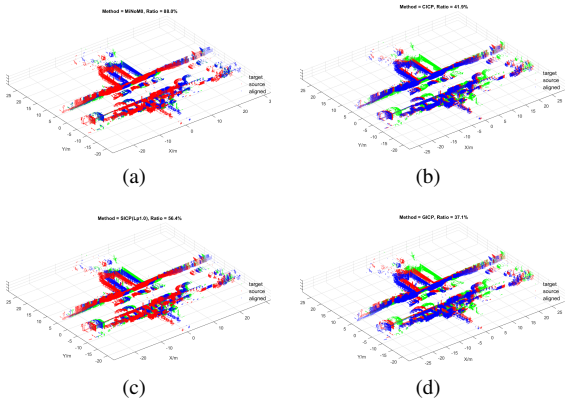


Fig. 4. The matching results for two frames from sequence indexed by 00. (a)(b)(c)(d) are match results via MiNoM, CICP, SICP and GICP, respectively. Only the proposed MiNoM produces high accuracy matching.

where $N_{good}$ is the frame number of high accuracy and $N_{total}$ is the total frame number of this sequence. The percent score indicates the convergence characteristic of scan matching algorithms.

*2) Ratio Score:* It is based on $L_0$ norm of residual error

$$\begin{aligned} U(e_i, \varepsilon) &= \begin{cases} 1 & e_i \leq \varepsilon \\ 0 & e_i > \varepsilon \end{cases}, \\ ratio &= \sum_{i=1}^{N_q} U(e_i, \varepsilon) \Big/ N_q, \end{aligned} \tag{20}$$

where $\epsilon = 0.2$ meter. The ratio score indicates the matching accuracy of scan matching algorithms.

### E. Convergence and Accuracy Analysis

*1) GPS Available:* To evaluate scan matching algorithms with different sequences of KITTI, each point cloud is matched with its previous point cloud, under the initial transformation being identity rotation matrix and zero translation vector. The comparison results for these algorithms in different sequences are listed in Tab I. For each score, the top three are colored in green, blue and red.

In terms of percent score, the proposed MiNoM, either MiNoM0 or MiNoM1, is the top or the second places of the scan matching algorithms, indicating that MiNoM has a better convergence characteristic than other scan matching algorithms. In total, MiNoM0 outperforms other scan matching algorithms as large as 7.3 %, which is equivalent to 195 frames' scan matching. A typical scan matching is illustrated in Fig. 4, where all of scan matching algorithms except MiNoM fail due to complicated environments.

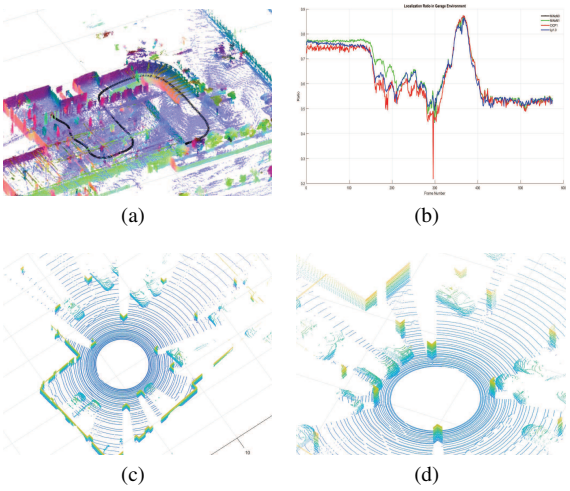
In terms of ratio score, MiNoM0 outperforms other scan matching algorithms around 0.9%, which is equivalent to around 450 to 720 points per scan matching, indicating that MiNoM attains better matching accuracy.

However, in terms of time cost, the proposed MiNoM is higher than p2pICP, p2plICP and GICP, but it is still comparable with CICP, SICP and NDT. Notice that SICP consumes most of time, since the alternating Direction Method of Multipliers (ADMM) used in SICP is much more

TABLE I

CONVERGENCE AND ACCURACY ANALYSIS BASED ON KITTI DATASET.

| SeqIdx/Length | Metric | MiNoM0 | MiNoM1 | CICP | $L_{0.4}$ | $L_{0.8}$ | $L_{1.0}$ | p2pICP | p2plICP | GICP | NDT |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 00/1079 | Percent/% | **76.18** | **74.70** | 61.82 | 14.27 | 56.63 | **70.16** | 36.24 | 57.00 | 64.41 | 29.75 |
| | Ratio/% | **84.35** | **83.19** | 81.02 | 53.88 | 76.70 | 82.54 | 60.96 | 70.62 | **84.25** | 62.32 |
| | Time/ms | 6065 | 9077 | 4933 | 31240 | 20122 | 11659 | **1013** | **615** | **264** | 3080 |
| 02/1397 | Percent/% | **69.01** | **67.86** | **67.79** | 18.83 | 51.32 | 60.63 | 33.43 | 57.48 | 63.28 | 41.95 |
| | Ratio/% | **85.78** | **85.22** | **85.84** | 64.22 | 80.05 | 83.57 | 68.36 | 76.03 | 84.38 | 74.23 |
| | Time/ms | 5541 | 7436 | 5406 | 13754 | 9534 | 4635 | **892** | **530** | **259** | 4063 |
| 07/208 | Percent/% | **85.58** | **82.21** | 71.15 | 10.10 | 57.21 | **78.85** | 35.10 | 61.54 | 62.98 | 39.90 |
| | Ratio/% | **83.44** | **82.55** | **81.97** | 50.51 | 74.76 | 81.75 | 57.91 | 69.26 | 81.60 | 66.05 |
| | Time/ms | 4335 | 7036 | 2961 | 14473 | 10183 | 4544 | **853** | **541** | **247** | 2309 |
| all/2684 | Percent/% | **73.17** | **71.72** | 65.65 | 16.32 | 53.91 | **65.87** | 34.69 | 57.60 | 63.71 | 36.89 |
| | Ratio/% | **85.02** | **84.20** | 83.61 | 59.00 | 78.29 | 83.01 | 64.57 | 73.33 | **84.11** | 68.81 |
| | Time/ms | 5658 | 8065 | 5027 | 20839 | 13841 | 7452 | **937** | **565** | **260** | 3532 |



(a) (b) (c) (d)

Fig. 5. Localization in underground garage. (a) illustrates the resulting trajectory produced by MiNoM. (b) illustrates ratio scores of various scan matching algorithms, and proposed MiNoM outperforms other scan matching algorithms obviously. The ratio scores of p2pICP and p2plICP are not displayed due to their worse matching accuracy. (c) and (d) depict point cloud at frame of 134 and 195, which contain large portion of outliers (cars) and occlusion.

computational expensive than IRLS. If source/target point cloud are sampled efficiently and MiNoM is implemented in C/C++, a moderate time cost reduction can be expected.

Also, we can see that the performance of MiNoM0 is slightly better than MiNoM1 since the SPV of MiNoM1 contains $s_k < 1.0$, which may introduce local minimals into the objective function. This can be validated by observing the performance of $L_{0.4}$-based, $L_{0.8}$-based, and $L_{1.0}$-based SICP. Thus, MiNoM0 is recommended for scan matching in this paper.

To summarize, experiments on KITTI dataset prove that scan matching algorithms including MiNoM, CICP, $L_{1.0}$-based SICP and GICP are of good quality in real traffic scenes. MiNoM gains the best convergence characteristic and matching accuracy among all scan matching algorithms.

*2) GPS-denied Situations:* During localization in GPS-denied environments, live point cloud from LiDAR sensor is served as source point cloud, and 3D occupancy map is served as target point cloud. The source and target clouds are aligned via scan matching algorithms, with initial transformation from the last matching result or re-localization[21] when the ratio score is lower than 0.5. The localization results are listed in Tab. II.

TABLE II

LOCALIZATION IN UNDERGROUND GARAGE.

| Metric | MiNoM0 | MiNoM1 | CICP | $L_{1.0}$ | p2pICP | p2plICP |
|---|---|---|---|---|---|---|
| Ratio | **65.15** | **65.14** | 62.51 | **63.87** | 58.88 | 59.77 |
| ReLoc | **12** | **12** | **35** | **5** | 106 | 89 |

From the Tab. II, the convergence of scan matching algorithm can be inferred by the number of re-localization. CICP, p2pICP and p2plICP all require more re-localization, and they still produce wrong matches (producing ratio score lower than 50%). For example, CICP failed at frame of 296, producing a ratio of 21.93%. The scan matchings generated by proposed MiNoM and $L_{1.0}$-based SICP attain good matching quality, indicating they both have a good convergence characteristic. Though the number of re-localization MiNoM required is higher than SICP, it is acceptable considering MiNoM's higher matching accuracy than SICP.

The ratio scores for scan matching algorithms are plotted in Fig. 5(b). It shows that the ratio scores of MiNoM0 and MiNoM1 are almost overlapped in all frames, and they outperform other scan matching algorithms by 1.28%. In Fig. 5(c) and 5(d), surrounding cars incur a large portion of outliers and occlusion, and only the proposed MiNoM gains an accurate matching, producing ratio scores of 77.9% and 67.7%, respectively. The final trajectory generated by MiNoM is presented in Fig. 5(a), which is smooth and accurate for autonomous navigation application.

### F. Sparsity Analysis

The advantage of proposed MiNoM is that it can find outliers automatically, with the sparsity introduced by mixed-norm-based objective function in Eq(8). A norm contributor

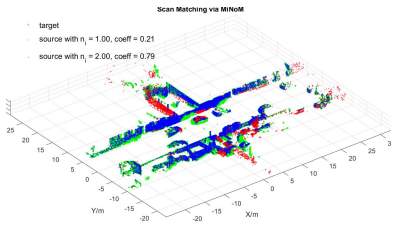



Fig. 6. Sparsity in MiNoM. Two point clouds are matched via MiNoM0, and source point cloud can be divided into two point sets: one belong to $n_i = 1.0$, the other belong to $n_i = 2.0$. The proposed MiNoM can select outliers automatically by observing point set belonging to $n_i = 1.0$.

$n_i$ is calculated based on the latent variable $\gamma_{ik}$

$$\begin{aligned} k_i^* &= \arg\max_k \gamma_{ik}, \qquad k = 1, 2, \dots K. \\ n_i &= \mathbf{s}_{k_i^*}, \end{aligned} \tag{21}$$

where $n_i$ indicates which component of $\mathbf{s}$ contributes most for a single measurement point.

As illustrated in Fig. 6, the point sets belonging to $n_i = 1.0$ and $n_i = 2.0$ are quite different: the point set belonging to $n_i = 1.0$ ensembles more likely as outliers, which includes un-observed walls and vehicles. The point set belonging to $n_i = 2.0$ is cling to the target point cloud, which indicates that this portion of source cloud is generated via target point cloud adding with Gaussian noise.

To summarize, the proposed MiNoM gains better performance than other scan matching algorithms because it can model residual error distribution better than others.

## V. CONCLUSIONS

In this paper, a highly accurate scan matching algorithm via residual error modelling is proposed. We show that the MoEP-based residual error distribution is flexible and robust in modelling residual error in real scenes. Meanwhile, the online parameter learning and IRLS are efficient for solving the mix-norm-based objective function. The experimental results show that the proposed MiNoM has a better convergence characteristic and matching accuracy than some state-of-art scan matching algorithms. Though promising improvements in robustness and accuracy are achieved as we expected, there still exist some work need to be done in the future: a) facilitating real-time implementation via multi-thread programming and fast nearest neighbor search, b) analyzing the selection of SPV rigorously, and c) integrating MiNoM into Monte Carlo localization (MCL).

## VI. APPENDIX

### A. Exponential Power Distribution

Exponential power (EP) distribution[15] is a generalized distribution belonging to exponential family

$$\begin{aligned} p\left(e|\theta, s\right) &= \lambda \exp\left(-\theta|e|^s\right), \\ \lambda &= s\theta^{1/s}/2\Gamma\left(1/s\right), \end{aligned} \tag{22}$$

where $\theta$ is the precision parameter, $s$ is the shape parameter, and $\Gamma()$ is the Gamma function. For simplify, Eq(22) can be denoted as $p\left(e|\theta, s\right) \sim EP\left(e|\theta, s\right)$. It should be noted that the residual error $e_i$ in Eq(2) is always non-negative, thus the exponential power distribution should be modified slightly as

$$p\left(e|\theta, s\right) \approx 2.0 \times EP\left(e|\theta, s\right), \qquad e \geq 0. \tag{23}$$

The probability density functions of EP with different shape parameter are illustrated in Fig. 2.

## REFERENCES


[1] L. Li, M. Yang, L. Guo, C. Wang, and B. Wang, "Hierarchical neighborhood based precise localization for intelligent vehicles in urban environments," *IEEE Transactions on Intelligent Vehicles*, vol. 1, no. 3, pp. 220–229, 2017.

[2] D. Cui, J. Xue, S. Du, and N. Zheng, "Real-time global localization of intelligent road vehicles in lane-level via lane marking detection and shape registration," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2014, pp. 4958–4964.

[3] F. Pomerleau, F. Colas, R. Siegwart, and S. Magnenat, "Comparing ICP variants on real-world data sets," *Autonomous Robots*, vol. 34, no. 3, pp. 133–148, 2013.

[4] P. J. Besl and N. D. Mckay, "A method for registration of 3-D shapes," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 14, no. 2, pp. 239–256, 2002.

[5] K. L. Low, "Linear least-squares optimization for point-to-plane ICP surface registration," *Chapel Hill*, 2004.

[6] A. Segal, D. Hähnel, and S. Thrun, "Generalized-ICP," in *Robotics: Science and Systems*, 2009.

[7] J. Serafin and G. Grisetti, "NICP: Dense normal based point cloud registration," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2015, pp. 742–749.

[8] G. Xu, S. Du, and J. Xue, "Precise 2D point set registration using iterative closest algorithm and correntropy," in *International Joint Conference on Neural Networks*, 2016, pp. 4627–4631.

[9] P. Bergström and O. Edlund, "Robust registration of point sets using iteratively reweighted least squares," *Computational Optimization & Applications*, vol. 58, no. 3, pp. 543–561, 2014.

[10] S. Bouaziz, A. Tagliasacchi, and M. Pauly, "Sparse iterative closest point," *Symposium on Geometry Processing*, vol. 32, no. 5, pp. 113–123, 2013.

[11] M. Magnusson, A. Nuchter, C. Lorken, and A. J. Lilienthal, "Evaluation of 3D registration reliability and speed - a comparison of ICP and NDT," in *IEEE International Conference on Robotics and Automation*, 2009, pp. 2263–2268.

[12] P. Biber, S. Fleck, and W. Strasser, "A probabilistic framework for robust and accurate matching of point clouds," in *Pattern Recognition, Tübingen, Germany, Proceedings*, 2004, pp. 480–487.

[13] R. W. Wolcott and R. M. Eustice, "Robust LIDAR localization using multiresolution Gaussian mixture maps for autonomous driving," *International Journal of Robotics Research*, vol. 36, no. 3, pp.292–319, 2017.

[14] M. Yang, L. Zhang, J. Yang, and D. Zhang, "Robust sparse coding for face recognition," in *Computer Vision and Pattern Recognition*, 2011, pp. 625–632.

[15] X. Cao, Y. Chen, Q. Zhao, D. Meng, Y. Wang, D. Wang, and Z. Xu, "Low-rank matrix factorization under general mixture noise distributions," in *International Conference on Computer Vision*, 2015, pp. 1493–1501.

[16] C. S. Burrus, "Iterative re-weighted least squares," *Comm.pure Appl.math*, vol. 44, no. 6, pp. 1–9, 2009.

[17] S. Thrun, W. Burgard, and D. Fox, *Probabilistic robotics*. MIT press, 2005.

[18] R. Urtasun, P. Lenz, and A. Geiger, "Are we ready for autonomous driving? the KITTI vision benchmark suite," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 3354–3361.

[19] G. Grisetti, R. Kummerle, C. Stachniss, and W. Burgard, "A tutorial on graph-based SLAM," *IEEE Intelligent Transportation Systems Magazine*, vol. 2, no. 4, pp. 31–43, 2011.

[20] Q. H. Du, "Metrics for 3d rotations: Comparison and analysis," *Journal of Mathematical Imaging & Vision*, vol. 35, no. 2, pp. 155–164, 2009.

[21] R. Dubé, D. Dugas, E. Stumm, J. Nieto, R. Siegwart, and C. Cadena, "Segmatch: Segment based loop-closure for 3D point clouds," in *2017 IEEE International Conference on Robotics and Automation*, 2017, pp. 5266–5272.