Estimating Rigid Transformation Between Two Range Maps Using Expectation Maximization Algorithm

Shuqing Zeng†

Abstract

We address the problem of estimating a rigid transformation between two point sets, which is a key module for target tracking system using Light Detection And Ranging (LiDAR). A fast implementation of Expectation-maximization (EM) algorithm is presented whose complexity is $O(N)$ with $N$ the number of scan points.

I. INTRODUCTION

Rigid registration of two sets of points sampled from a surface has been widely investigated (e.g., [1], [4]–[6], [9]) in computer vision literature. Generally, these methods are designed to tackle range maps with dense points for non-realtime applications.

In [2], [8] scans are matched using iterative closest line (ICL), a variant of “normal-distance” form of ICP algorithm [1] originally proposed in computer vision community by [3]. However, the convergence of this approach is sensitive to errors in normal direction estimations [10].

![Fig. 1. Illustration of the proposed algorithm.](image)

Fig. 1 illustrates the concept. The light green circles denote the contour of a target $M$. The red circles are the projection of $M$ under a rigid transformation $T$, denoted as $\tilde{M}$. Let $S$ be the current range image shown as upper triangles. We propose an Expectation-maximization (EM) algorithm [4], [7] to find the rigid transformation such that the projected range image best matches the current image. Each point $m_j$ in $\tilde{M}$ is treated as the center of a Parzen window. There is an edge between $s_k \in S$ and $m_j$ if $s_k$ lies in the window. The weight of the edge $(s_k, m_j)$ is based on the proximity between the two vertices. The larger weight of the edge, the thicker the line is shown, and the more force that pulls the corresponding point $m_j$ to $s_k$ through $T$.

This document describes a fast implementation of expectation maximization (EM) algorithm [7] to locally match between $M$ and $S$. By exploiting the sparsity of the locally matching matrix, this implementation scales linearly with the number of points.

†Department of Computer Science and Engineering, Michigan State University, East Lansing, MI 48824, e-mail: zengshuq@msu.edu.
II. ALGORITHM DERIVATION

This section is devoted to the problem of how to estimate the rigid transformation $T$ using (EM) algorithm, giving scan map $\mathcal{S}$ and a contour model $\mathcal{M}$.

We constructed a bipartite graph $B = (\mathcal{M}, \mathcal{S}, E_B)$ between the vertex set $\mathcal{M}$ to $\mathcal{S}$ with $E_B$ the set of edges. Let $m \in \mathcal{M}$ and $s \in \mathcal{S}$. An edge exists between the points $s$ and $m$ if and only if $\|s - m\| < W$ with $W$ a distance threshold. By $\mathcal{N}(s) \equiv \{m \mid (s, m) \in E_B\}$ we denote the neighborhood of $s$.

Scan points are indexed using a lookup hash-table with $W/2$ resolution. Find the points $m$ near a point $s$ within the radius $W$ involving searching through all the three-by-three neighbor grid of the cell containing $s$. Since hash table is used, and $|\mathcal{N}(s)|$ is bounded, construction graph $B$ is an $O(N)$ operation with $N$ the number of points in a scan.

Let $s_j \in \mathcal{S}$ be one of the $n_S$ scan points, and $m_k \in \mathcal{M}$ be one of the $n_M$ points from the model. We denote $\bar{T}$ a rigid transformation from the model to the new scan frame, with the parameter vector $y$. If $s_j$ is the measure of $m_k$ (i.e., $(s_j, m_k) \in B$) with a known noise model, we write the density function as $p(s_j \mid m_k, y) = p(s_j \mid T(m_k, y))$. In case of an additive and centered Gaussian noise of precision matrix $\Gamma$, $p(s_j \mid m_k, T) = e^{-\frac{1}{2}||s_j - \bar{T}(m_k, y)||^2}$ where the Mahalanobis norm is defined as $||x||^2_\Gamma \equiv x^\Gamma x$.

We use the binary matrix $A$ to represent the correspondence between $s_j$ and $m_k$. The entry $A_{jk} = 1$ if $s_j$ matches $m_k$ and 0 otherwise. Assume each scan point $s_j$ corresponds to at most one model point. We have

$$\Sigma_k A_{jk} = \begin{cases} 1 & \text{if } \mathcal{N}(s_j) \neq \emptyset \\ 0 & \text{Otherwise.} \end{cases}$$

for all scan point index $j$.

For the above equation, we note that for the case $\mathcal{N}(s_j) = \emptyset$, $s_j$ is an outlier, and the correspondence $s_j$ to $m_k$ can be treated as a categorical distribution. In order to apply EM procedure we use a random matching matrix $\mathcal{A}$ with each element a binary random variable. Each eligible matching matrix $\mathcal{A}$ has a probability $p(\mathcal{A}) \equiv p(\mathcal{A} = A)$. One can verify that $\bar{A}_{jk} = E\{A_{jk}\} = P(\mathcal{A}_{jk} = 1)$, and the following constraint holds

$$\Sigma_k \bar{A}_{jk} = \begin{cases} 1 & \text{if } \mathcal{N}(s_j) \neq \emptyset \\ 0 & \text{Otherwise.} \end{cases}$$

Considering the distribution of $\mathcal{A}_j$, the $j$-th row of the $\mathcal{A}$, which is the distribution of assigning the scan point $s_j$ to the model point $m_k$, i.e.,

$$p(\mathcal{A}_j) = \prod_{m_k \in \mathcal{N}(s_j)} (\bar{A}_{jk})^{A_{jk}}$$

Assuming the scan points are independent, we can write

$$p(\mathcal{A}) = \prod_{s_j \in \mathcal{S}} \prod_{m_k \in \mathcal{N}(s_j)} (\bar{A}_{jk})^{A_{jk}} = \prod_{(s_j, m_k) \in E_B} (\bar{A}_{jk})^{A_{jk}}$$

(1)

An example of $p(\mathcal{A})$ is the noninformative prior probability of the matches: a probability distribution that a given scan point is a measure of a given model point without knowing measurement information:

$$\bar{A}_{jk} = \pi_{jk} = \begin{cases} \frac{1}{|\mathcal{N}(s_j)|} & \text{if } \mathcal{N}(s_j) \neq \emptyset \\ 0 & \text{Otherwise.} \end{cases}$$

The joint probability of the scan point $s_j$ and the corresponding assignment $\mathcal{A}_j$ can be expressed as

$$p(s_j, \mathcal{A}_j \mid \mathcal{M}, y) = \prod_{m_k \in \mathcal{N}(s_j)} (\pi_{jk} p(s_j \mid m_k, y))^{A_{jk}}$$
Providing that the scan points are conditionally independent, the overall joint probability is the product of the each row of $\mathcal{A}$:

$$p(\mathcal{S}, \mathcal{A} \mid \mathcal{M}, y) = \prod_{(s_j, m_k) \in \mathcal{E}} (\pi_{jk} p(s_j \mid m_k, y))^{A_{jk}}$$

and the logarithm of marginal distribution can be written as

$$\text{ML}(T) = \log p(\mathcal{S} \mid \mathcal{M}, y) = \log \left( \sum_{\mathcal{A}} p(\mathcal{S}, \mathcal{A} \mid \mathcal{M}, y) \right)$$

Unfortunately, Eq. (3) has no closed-form solution and no robust and efficient algorithm to directly minimize it with respect to the parameter $y$. Noticing that Eq. (3) only involves the logarithm of a sum, we can treat the matching matrix $\mathcal{A}$ as latent variables and apply the EM algorithm to iteratively estimate $y$. Assuming after $n$-th iteration, the current estimate for $y$ is given by $y_n$, we can compute an updated estimate such that $\text{ML}(T)$ is monotonically increasing, i.e.,

$$\Delta(y \mid y_n) = \text{ML}(y) - \text{ML}(y_n) > 0$$

Namely, we want to maximize the difference $\Delta(y \mid y_n)$.

Now we are ready to state two propositions whose proofs are relegated to Appendix.

**Proposition 2.1:**

$$\Delta(y \mid y_n) = E_{\mathcal{A} \mid \mathcal{S}, \mathcal{M}, y_n} \{ \log (p(\mathcal{S}, \mathcal{A} \mid \mathcal{M}, y)) \}$$

**Proposition 2.2:** Given the transformation estimate $y_n$, scan points $\mathcal{S}$ and model points $\mathcal{M}$, the posterior of the matching matrix $\mathcal{A}$ can be written as

$$p(\mathcal{A} \mid \mathcal{S}, \mathcal{M}, y_n) = \prod_{j, k: (s_j, m_k) \in \mathcal{E}} (\hat{A}_{jk})^{A_{jk}}$$

where

$$E\{A\}_{jk} = \hat{A}_{jk} = \left\{ \begin{array}{ll}
\frac{\pi_{jk} p(s_j \mid m_k, y_n)}{\sum_{k} \pi_{j} p(s_j \mid m_k, y_n)} & \text{If } \mathcal{N}(s_j) \neq \emptyset \\
0 & \text{Otherwise.}
\end{array} \right.$$ 

Therefore, we have the following EM algorithm to compute $y$ that maximizes the likelihood defined in Eq. (3). We assume there exists an edge in the graph $\mathcal{B}$ between $s_j$ and $m_k$ in the following derivation.

- **E-step:** Given the previous estimate $T_n$, we update $\hat{A}_{jk}$ using Eq. (5). The conditional expectation is computed as

$$\Delta(y \mid y_n) = E_{\mathcal{A} \mid \mathcal{S}, \mathcal{M}, y_n} \{ \log p(\mathcal{S}, \mathcal{A} \mid \mathcal{M}, y) \}$$

$$= E \left\{ \log \left( \prod_{j, k} \pi_{jk} p(s_j \mid m_k, y)^{A_{jk}} \right) \right\}$$

$$= \sum_{j, k} E\{A_{jk}\} \{ \log p(s_j \mid m_k, y) + \log \pi_{jk} \}$$

$$= \sum_{j, k} \hat{A}_{jk} \| s_j - T(m_k, y) \|_2^2 + \text{const.}$$

where $E$ is $E_{\mathcal{A} \mid \mathcal{S}, \mathcal{M}, y_n}$ in short, and const. is the terms irrelevant to $y$.

- **M-step:** Compute $y$ to maximize the least-squares expression in Eq. (6).

The above EM procedure is repeated until the model is converged, i.e., the difference of log-likelihood between two iterations $\Delta(y \mid y_n)$ is less than a small number. The complexity of the above computation for a target in each iteration is $O(|E_{\mathcal{B}}|)$. Since the number of neighbors for $s_j$ is bounded, the complexity is reduced to $O(|\mathcal{S}|)$. Since experimental result shows that only 4-5 epochs are needed for EM iteration
to converge. Consequently, the overall complexity for all of the tracked objects is \( O(N) \) with \( N \) the number of scan points.

The following proposition shows how to compute the covariance matrix for the transformation parameters \( \mathbf{y} \).

**Proposition 2.3:** Given \( \mathbf{y} \), the covariance matrix \( \mathbf{R} \) is

\[
\mathbf{R} = \frac{1}{n_P} \sum_{(s, m_k) \in \mathcal{E}_B} \hat{A}_{jk}(s_j - T(m_k, \mathbf{y}))(s_j - T(m_k, \mathbf{y}))^T
\]

(7)

where \( n_P \) is the number of the nonzero rows of the matrix \( \hat{A} \).

### III. Proof of Propositions

#### A. Proof of Proposition 2.1

\[
\Delta(\mathbf{y} \mid \mathbf{y}_n) = \text{ML}(\mathbf{y}) - \text{ML}(\mathbf{y}_n)
= \log \left( \sum_{\mathcal{A}} p(\mathbf{S}, \mathcal{A} | \mathcal{M}, \mathbf{y}) \right) - \log \left( p(\mathbf{S} \mid \mathcal{M}, \mathbf{y}_n) \right)
= \log \left( \sum_{\mathcal{A}} p(\mathcal{A} | \mathcal{M}, \mathbf{y}) p(\mathbf{S}, \mathcal{A} | \mathcal{M}, \mathbf{y}) \right)
- \log p(\mathbf{S} | \mathcal{M}, \mathbf{y}_n)
\geq \sum_{\mathcal{A}} p(\mathcal{A} | \mathcal{M}, \mathbf{y}_n) \log \left( \frac{p(\mathbf{S}, \mathcal{A} | \mathcal{M}, \mathbf{y}) p(\mathcal{A} | \mathcal{M}, \mathbf{y})}{p(\mathbf{A}, \mathbf{y}|\mathcal{M})} \right)
- \log p(\mathbf{S} | \mathcal{M}, \mathbf{y}_n)
= \sum_{\mathcal{A}} p(\mathcal{A} | \mathcal{M}, \mathbf{y}_n) \log \left( \frac{p(\mathbf{S}, \mathcal{A} | \mathcal{M}, \mathbf{y}) p(\mathcal{A} | \mathcal{M}, \mathbf{y})}{p(\mathbf{A}, \mathbf{y}|\mathcal{M})} \right)
\]

(8)

where Jansen’s inequality and convexity of logarithm function are applied in deriving Eq. (8). Since we are maximizing \( \Delta(\mathbf{y} \mid \mathbf{y}_n) \) with respect to \( \mathbf{y} \), we can drop terms that are irrelevant to \( \mathbf{y} \), thus

\[
\Delta(\mathbf{y} \mid \mathbf{y}_n) = \sum_{\mathcal{A}} p(\mathcal{A} | \mathcal{M}, \mathbf{y}_n) \log \left( \frac{p(\mathbf{S}, \mathcal{A} | \mathcal{M}, \mathbf{y}) p(\mathcal{A} | \mathcal{M}, \mathbf{y})}{p(\mathbf{A}, \mathbf{y}|\mathcal{M})} \right)
\]

(9)
B. Proof of Proposition 2.2

If $\mathcal{N}(s_j) \neq \emptyset$, the marginal PDF of the $j$-th row of $A$ is $p(s_j | M, y) = \sum_k \pi_{jk} p(s_j | m_k, y)$. We assume there exists an edge in the graph $B$ between the scan and model points $s_j$ and $m_k$, and scan points are independent each other. One can verify that

$$p(S|M, y_n) = \prod_j \left( \sum_k \pi_{jk} p(s_j | m_k, y_n) \right)$$

Using Bayesian theorem, we have

$$p(A|S, M, y_n) = \frac{p(S, A|M, y_n)}{p(S|M, y_n)}$$

$$= \frac{\prod_{j,k} (\pi_{jk} p(s_j | m_k, y_n))^{A_{jk}}}{\prod_j (\sum_k \pi_{jk} p(s_j | m_k, y_n))}$$

$$= \frac{\prod_{j,k} (\pi_{jk} p(s_j | m_k, y_n))^{A_{jk}}}{\prod_{j,k} (\sum_k \pi_{jk} p(s_j | m_k, y_n))^{A_{jk}}}$$

Comparing with Eq. (4), the equation Eq. (5) holds.

C. Proof of Proposition 2.3

We treat the precision matrix $\Gamma$ as the uncertainty of unknown transformation parameter $y$. We use a maximum likelihood approach, which amounts to minimizing Eq. (6) with respect to $\Gamma$ given a transformation and a set of matches with probabilities:

$$\frac{\partial}{\partial \Gamma} \Delta(y|y_n)$$

$$= \frac{\partial}{\partial \Gamma} \sum_{(s_j, m_k) \in E_B} \hat{A}_{jk} \left( \frac{\|s_j - T(m_k, y)\|^2}{2} + \log |\Gamma|^{-\frac{1}{2}} \right)$$

$$= \frac{1}{2} \sum_{(s_j, m_k) \in E_B} \hat{A}_{jk} (s_j - T(m_k, y)) (s_j - T(m_k, y))^T$$

$$- \frac{n_p}{2} \Gamma^{-1} = 0$$

where $n_p$ is the number of nonzero rows of the matrix $\hat{A}$. Thereby the covariance matrix $R$ is computed as

$$R = \Gamma^{-1}$$

$$= \frac{1}{n_p} \sum_{(s_j, m_k) \in E_B} \hat{A}_{jk} (s_j - T(m_k)) (s_j - T(m_k))^T$$

REFERENCES

[1] P. Besl and N. McKay. A method for registration of 3-D shapes. IEEE Trans. Pattern Analaysis and Machine Intelligence, 14(2):239–256, 1992.

[2] A. Censi. An icp variant using a point-to-line metric. In IEEE International Conference on Robotics and Automation, pages 19–25, New York, NY, 2008.
[3] Y. Chen and G. Medioni. Object modeling by registration of multiple range images. *Image and Vision Computing*, 10(3):145–155, 1992.

[4] S. Granger and X. Pennec. Multi-scale EM-ICP: A fast and robust approach for surface registration. In *ECCV*, pages 418–432, 2002.

[5] A. Jagannathan and E. Miller. Unstructure point cloud matching within graph-theoretic and thermodynamic frameworks. In *CVPR*, pages 1008–1015, 2005.

[6] A. Makadia, A. Patterson, and K. Daniilidis. Fully automatic registration of 3D point clouds. In *CVPR*, pages 1297–1304, 2006.

[7] G. McLachain and T. Krishnan. *The EM algorithm and extensions*. John Wiley & Sons Inc., New York, second edition, 2008.

[8] E. Olson. Real-time correlative scan matching. In *IEEE International Conference on Robotics and Automation*, pages 4387–4393, Kobe, Japan, 2009.

[9] G. Sharp, S. Lee, and D. Wehe. ICP registration using invariant features. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(1):90–102, 2002.

[10] C. Stewart. Uncertainty-driven, point-based image registration. In N. Paragios, Y. Chen, and O. Faugeras, editors, *Handbook of Mathematical Models in Computer Vision*, chapter 14, pages 221–235. Springer, 2006.