A Novel Method For Map Alignment Assessment Using Synthetic Displacement Fields

Rema Daher, Theodor Chakhachiro, and Daniel Asmar

Abstract—In the era of autonomous robot mapping, assessing the goodness of the generated maps is important, and is usually performed by first aligning them to blueprints and then calculating the displacement error. Accuracy of map alignment is also critical in other applications such as collaborative mapping in multi-robot applications and the use of prior maps in real time robot localization and navigation. However, map alignment is difficult for two reasons: first, one map can be significantly distorted from the other, and second, establishing what constitutes a ground truth for alignments of different types is challenging. Most map alignment techniques to this date have addressed the first problem, while paying too little importance to the second. In this paper, we propose a ground truth, which consists of synthetically transformed maps with their corresponding displacement fields. Furthermore, we propose a new system for comparison, where the displacement field of any occupancy grid map alignment technique, can be computed and compared to the ground truth using statistical measures. The local information in displacement fields renders the evaluation system applicable to any alignment technique, whether it is linear or not. In our experiments, the proposed method was applied to different alignment techniques from the literature.

I. INTRODUCTION

Many robot applications today rely on map alignment for environment modeling. For instance, in real-time robot mapping, the blueprint semantic information can be merged with the sensor map to create a more meaningful map for more accurate localization [1] [2]. Another important application is collaborative mapping in multi-robot systems that require data sharing [3]. The main challenge in map alignment is feature matching which is a critical part of evaluating robot mapping [4].

In order to encompass all the above mentioned applications, map alignment methods should be applicable on map pairs with different modalities, scales, percentage of space coverage, and local and global deformations. For example, when aligning a sensor map of occupancy grid format to a blueprint, the algorithm should be able to align them despite the deformations in the sensor map, the scale differences, the partial coverage of the sensor map, and the blueprint’s RGB format. The map alignment problem also presents challenges that arise from the limited information or features present in occupancy grid maps.

To solve these challenges, most of the methods to date try to generate reliable features and descriptive information that do not rely on modalities, scales, percentage of space coverage, or local and global deformations.

Solutions to the alignment of occupancy grid maps include approaches such as image registration, optimization, Hough/Radon transform, and graph matching. In this paper, machine learning methods such as [5] will not be discussed since these methods have focused on aligning images or raster maps of simple displacement types and not on aligning occupancy grid maps, which differ greatly in their features and complexity. To elaborate, in [5], a satellite image is aligned with a cadastral map, which is a different application than ours, whereby we are only aligning maps that can be represented as occupancy grid maps, such as architectural drawings and robot maps. First, image registration techniques were adapted to perform map alignment, where optimization and feature based methods were among the most widespread. However, indoor maps have much fewer features than regular photos due to their repetitive structures and the lack of texture in the occupancy grid format. Thus, methods based on features such as SIFT [6] and Enhanced Correlation Coefficient (ECC) Maximization [7] were affected by local minima. As for alignment through optimization, the dependency of the optimization on the distance between points for correspondence renders it also sensitive to local minima and only able to give satisfying results with small deformations and close input maps. Such optimization methods include Iterative Closest Point (ICP) [8], Coherent Point Drift (CPD) [9], and the algorithms proposed in [10] [11].

Consequently, other approaches started to appear that transformed these maps into a parametric space using the Hough and Radon transform, which introduce more relevant features to be matched [12] [13]. Nonetheless, these methods align through rigid transformations and thus fail in considering the scale of these maps.

In contrast, graph matching methods interpret abstract representations of the maps and perform nonrigid transformations. Graph matching is able to capture geometric and topological features [14] [15] [1]. Saeedi [16] proposed an approach that uses a graph based representation, Voronoi diagram, and solves the alignment problem through radon transforms and edge matching methods.

In all the above mentioned methods, map modalities are disregarded. In contrast, the state of the art technique that deals with multi-modal maps uses a decomposition-based algorithm
abstracting maps into 2D arrangements [17]. Shahbandi et al. expand their work to include local deformations resulting in a nonlinear transformation method, which uses an optimization formulation that fine-tunes their previous work [18].

The problem addressed in this paper is the lack of an accurate evaluation method for these alignment methods. In the literature, manual evaluations are most commonly used along with automatic evaluations based on unreliable ground truth transformations. The main contributions of this work include the following:

- Proposing an automatic system that can evaluate all types of occupancy grid map alignment techniques
- Creating a ground truth of synthetic maps with their corresponding deformations fields
- Validating our system by evaluating and comparing various map alignment methods

The paper is organized as follows: In Section II, a literature review of the related work is laid out and discussed. In Section III, the proposed system is described. The ground truth generation is first detailed followed by an explanation of our map alignment evaluation system. Furthermore, the experiments are described in Section IV accompanied by a visualization and both a quantitative and a qualitative analysis of the results found in Section V. Finally, Section VI concludes the paper.

II. RELATED WORK

In order to evaluate map alignment methods, ground truth transformations, represented as displacement fields, are needed. Due to the fact that it is not possible to generate ground truth displacement from a real dataset, there is an absence of a reliable ground truth in the literature. Thus, available error metrics depend heavily on the map alignment algorithm itself. For example, previous work includes quantitative and qualitative assessments of the map merging methods. In [19], qualitative and quantitative approaches are used to judge the performance of multi-robot SLAM algorithms (MRSLAM). The qualitative assessment is a discussion of the advantages and disadvantages of each MRSLAM map merging algorithm, which range from sensor noise to the quality of the optimization technique. The quantitative assessment measured the algorithm time, the number of features extracted from the map, and the distance traveled by the robot [19]. This assessment lacks a measure of the alignment transformation error.

In contrast to [19], several metrics exist that assess the quality and accuracy of the alignment transformation. One of these metrics is the success rate, which is a visual categorization of successful alignments. The categorization is binary, where the success rate is the ratio of the number of successful alignment experiments over the total number of experiments [17] [18]. A more objective metric is the keypoint method, where matching keypoints in both maps are manually selected through visual inspection. Then, the root mean square error between the manually associated keypoints is calculated and used as an accuracy measure. In both these metrics, manual intervention is required as well as the absence of a comprehensive evaluation assessing local map transformations [18].

In comparison to manual metrics, [20] adopted an automatic measure of merging techniques. The measure computes the rotation and translation errors between the rigid transformation of a merging technique and the ground truth transformation. The ground truth used is from the Stanford 3D Scanning Repository, which provides rigid transformations calculated using a modified ICP algorithm to align multi-view scans. Since these ground truth rigid transformations are generated using another map alignment technique, the ground truth is error prone and therefore unreliable.

It is also worth mentioning metrics that rely on robot pose as a ground truth [21]. However, the robot pose is not only error prone, but it is also not available in every map alignment application.

All these techniques fall short in providing an accurate evaluation metric with a reliable ground truth that can automatically assess nonlinear and nonrigid map alignments.

Automatic evaluation needs a reliable ground truth transformation, however these transformations cannot be obtained for real data. Thus, for global transformations that can be represented as affine transformations, a synthetic ground truth can be created from layout maps that are transformed with random affine transformations. However, this ground truth would not be representative of the real scenarios, where the map is not only deformed globally through affine transformations, but also through local deformations and noise. In this paper, we present a ground truth (found here1) of layout maps transformed with local deformations and noise while keeping coherency of the maps at the same time. Local deformations and noise represent errors that occur during robot mapping, such as odometry drift and sensor errors. This ground truth is then used with the assessment method to compare various map alignment methods.

III. PROPOSED SYSTEM

The proposed system is a map evaluation technique relying on a ground truth, which includes synthetically transformed maps and their respective displacement fields, followed by the generation of a displacement field for the map alignment to be compared with the ground truth using various statistical functions.

A. Ground truth

In order to generate the ground truth for evaluating map alignment methods, the HouseExpo dataset [22] is used as a basis. The HouseExpo dataset is a set of 2D binary layout maps that resemble accurate blueprints with no noise or deformations. Maps in the HouseExpo dataset are first preprocessed then transformed into synthetic robot sensor maps while annotating the displacement field. Synthetic maps are

1https://www.dropbox.com/sh/wn5kzh3q1s6jpcn/AABe8Mw-UhyPKE9axXU7bHI4a?dl=0
needed since the displacement fields can’t be tracked in an otherwise real world dataset.

In the pre-processing step, the maps are reformatted into robot sensor map occupancy grid format by first creating borders through edge detection and then thickening them. These pre-processed layout maps are then distorted to resemble robot sensor maps or any other error prone mapping technique.

While distorting the map locally, coherency, and thus the general shape of the map, must be maintained. Coherency was achieved by adapting the Elastic Distortion for data expansion algorithm presented in [23] to the pre-processed maps at hand. The Elastic Distortion algorithm was demonstrated on a handwriting recognition task MNIST [24], which has similar properties to the task at hand, but differs in scale, distortion formulation, and complexity.

In the elastic distortion algorithm, a random displacement field of values between $-1$ and $1$ is first computed in the $x$ and $y$ direction of each pixel in the original map. Then, this field is normalized, smoothed using a 2D Gaussian filter, and scaled. This distortion field models the random noise generated mainly by the sensors.

In our adaptation of the elastic distortion algorithm additional steps are added including:

- Erasing a random rectangular part of the map to recreate partial mapping
- Adding a nonlinear displacement field that represents large distortions and brokenness
- Adding a displacement field that scales the map
- Adding a weighted sum of the various computed displacement fields

To model global distortions, rotation and translation can also be added. However, in this paper we will only focus on scale, since rotation and translation can be easily handled by all the alignment methods that will be compared in Section IV.

With these modifications, the adapted overall system becomes as illustrated in Fig. 1. First the pre-processed layout maps are transformed into partial or incomplete maps. Partial mapping is achieved by substituting random rectangles from the image with the value of unknown cells, shown in a blue outline in Fig. 2. One corner of the random rectangle is constrained between $1/5$ and $1/2$ of the map image size.

Subsequently, a random displacement field $D_{x1}$, $D_{y1}$, between $-1$ and $1$, is generated and acts as small scale local distortions in the map edges that usually results from robot mapping. This distortion is seen in green in Fig. 2.

The second displacement field is responsible for large scale distortions affecting clusters in the map, which is also a
common form of distortion in robot mapping. For instance, Fig. 2 highlights a large deformation in red. This field is experimentally formulated as a nonlinear function of \( x \) and \( y \):

\[
D_{x2} = \sqrt{\text{rand}(a \in (-1,1))x + \text{rand}(a \in (-1,1))y} \\
D_{y2} = \sqrt{\text{rand}(a \in (-1,1))x + \text{rand}(a \in (-1,1))y}
\]

These equations were experimentally formulated due to the lack of a parametric model that correctly reproduces the additive effect of the slippage and friction of the robot. To visually validate this formulation, a set of real robotic sensor maps was generated; using the Robotic Operating System (ROS) Kinetic distribution [25], where the Explore Lite as well as the Turtlebot3 Gazebo, Navigation and Mapping packages were implemented on the Waffle Pi robot. These simulated maps are compared to the synthetically transformed maps generated by the proposed system. Both sets of generated maps are shown in Fig. 3, where the deformations seem to be of a similar nature. This visually validates the choice of the deformation formulation. A more quantitative validation is not possible, since simulated and synthetic deformations cannot be exactly the same due to the randomness of the noise. Another reason is that even if the deformations were the same, the displacement fields in the simulated robotic sensor maps cannot be tracked.

Both displacement fields are then normalized to be able to scale them later depending on the application:

\[
D_{x2,norm} = \|D_{x2}\| \\
D_{y2,norm} = \|D_{y2}\| \\
D_{x1,norm} = \|D_{x1}\| \\
D_{y1,norm} = \|D_{y1}\|
\]

The scaling of the displacement fields depends on the resolution of the input map image and the amount of distortion needed to portray a robot sensor map. Through experimentation, the scale factors, \((s_{x1}, s_{y1})\) and \((s_{x2}, s_{y2})\), for \((D_{x1,norm}, D_{y1,norm})\) and \((D_{x2,norm}, D_{y2,norm})\), respectively, that produced the most realistic output were each between \((10, 15)\) and \((-3000, 3000)\).

Consequently, the weighted sum of the displacement fields in the \( x \) and \( y \) directions becomes:

\[
D_x = s_{x1}D_{x1,norm} + s_{x2}D_{x2,norm} \\
D_y = s_{y1}D_{y1,norm} + s_{y2}D_{y2,norm}
\]

For smoothing the overall displacement field, a Gaussian filter \((GF_x, GF_y)\) is applied, where the standard deviation of the convolution \(\sigma_{GF}\) is responsible for controlling how much the pixels are scattered while maintaining coherency of the distortion. A scaling factor \(\alpha\) is also used to control the intensity of the deformation. In addition, a convolution filter size \((GF_{sx}, GF_{sy})\) is chosen to have sufficient overlap over the displacement field in order to avoid losing information, but not too large to avoid causing redundant computations. These parameters are selected according to the Matlab implementation [26] of the elastic distortion algorithm [23] with \(\sigma_{GF} = 4, \alpha = 34, GF_{sx} = 2 \times 3 \times \text{ceil}(\sigma_{D_x}) + 1\), and \(GF_{sy} = 2 \times 3 \times \text{ceil}(\sigma_{D_y}) + 1\), where \(\sigma_{D_x}\) and \(\sigma_{D_y}\) are the standard deviation of the displacement field. The smoothed displacement field now becomes:

\[
SD_x = \alpha GF_x(D_x) \\
SD_y = \alpha GF_y(D_y)
\]

Following the filtering, a random scale between \(1/2\) and \(1\) is applied to the map image coordinates transforming \((u, v)\) into \((x, y)\). The displacement field caused by the scaling \((D_{x3}, D_{y3})\) is computed by subtracting \(x\) from \(u\) and \(y\) from \(v\). This displacement is added to \((SD_x, SD_y)\):

\[
D_{x,final} = SD_x + D_{x3} \\
D_{y,final} = SD_y + D_{y3}
\]

The entire deformation procedure is presented in blue in Fig. 1. Finally, the map image is transformed by applying the final displacement field to its pixels.

The original map image, \((D_{x,final}, D_{y,final})\), and the transformed images are used as the ground truth upon which the error metric compares the alignment methods.

### B. Map Alignment Evaluation System

To evaluate the map alignment methods, a displacement field \((D_{x,method}, D_{y,method})\) is generated for each map pair using various map alignment methods. This is achieved by warping the original pixel coordinates of the map with the transformation matrix generated from the map alignment method. The warped mesh grid is then subtracted from the original mesh grid to get the displacement field in the \( x \) and \( y \) direction of these pixel coordinates. Since each method is implemented using different coding languages and uses different transformation tools, additional processing is required to extract
the displacement information. After the displacement field is extracted, the method is assessed by first computing the error between the generated displacement field and its corresponding ground truth:

\[
\text{Error}_x = D_{x, \text{method}} - D_{x, \text{final}} \quad (13)
\]
\[
\text{Error}_y = D_{y, \text{method}} - D_{y, \text{final}} \quad (14)
\]

The matrices \( \text{Error}_x \) and \( \text{Error}_y \) are then concatenated in order to have all the errors in one matrix \( \text{Error} \). The flow from the ground truth generation to the \( \text{Error} \) matrix is highlighted in green in Fig. 1. After that, various statistical functions are performed on the absolute value of \( \text{Error} \), since the direction of the error is not needed to calculate the closeness of the map alignment method’s displacement field to its ground truth. These functions include:

\[
\text{Error} = \text{concat}(\text{Error}_x, \text{Error}_y) \quad (15)
\]
\[
\text{Mean} = \frac{1}{n+m} \sum_{i=1}^{n} \sum_{j=1}^{m} \text{Error}(n,m) \quad (16)
\]
\[
\text{rmse} = \sqrt{\text{Mean}(\text{Error}^2)} \quad (17)
\]
\[
\sigma_{\text{err}} = \sqrt{\frac{1}{n+m} \sum_{i=1}^{n} \sum_{j=1}^{m} (\text{Error}(n,m) - \text{Mean}(\text{Error}))^2} \quad (18)
\]
\[
\text{Max} = \text{maximum}(\text{Error}) \quad (19)
\]
\[
\text{Median} = \text{median}(\text{error}) \quad (20)
\]

Without loss of generality, these functions were selected due to their widespread use in the computer vision field and their ability to give a better understanding of the data. However, any other statistical metric can be used in addition or as a replacement of these functions.

**IV. EXPERIMENTS**

From our proposed ground truth, map pairs with their associated displacement fields are used to test various map alignment methods from the literature. It is worth mentioning that this comparison is not the purpose of this paper, but an application to validate our system. The alignment methods used in the comparison include: CPD [9] Voronoi diagram-based [16] SIFT [6] Hough-based [12] ECC maximization [7] Decomposition-Optimization based (DO) [18]

<table>
<thead>
<tr>
<th>Map</th>
<th>Original</th>
<th>Synthetic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**TABLE I**

THE MAIN EMPHASIS OF MAP PAIRS IN TERMS OF DEFORMATION PARAMETERS

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Scale</td>
</tr>
<tr>
<td>2</td>
<td>local small deformations</td>
</tr>
<tr>
<td>3</td>
<td>Large deformations and brokenness</td>
</tr>
<tr>
<td>4</td>
<td>Partial maps</td>
</tr>
<tr>
<td>5</td>
<td>Very similar room structures</td>
</tr>
<tr>
<td>6</td>
<td>All elements of distortion including Scale, local small deformations, large deformations and brokenness and partial maps</td>
</tr>
</tbody>
</table>

A similar comparison is made in [18], which is based on manually selecting corresponding keypoints. Thus, the validity of our results will be based on their similarity with the output of [18].

To capture the strength of each method, the map pairs are uniquely chosen to highlight the main challenges of alignment, which include but are not limited to, scaling, local deformations, large deformations and brokenness, partial mapping, and similar room structures. To elaborate, 6 maps were deformed using all the previously mentioned elements of distortion, but each with a specific emphasis as shown in Table I and Fig. 4.

The transformed maps in Fig. 4 were generated on Matlab and the evaluation of the map alignment methods was per-
formed on Python. This work was implemented on a Toshiba Satellite Intel Core i7-5500U running at 2.3/3.0 Turbo GHz with an 8GB RAM and Intel HD Graphics 5500.

V. RESULTS

The displacement fields generated by each method were compared to the ground truth to validate our system. The Root Mean Square (RMS), standard deviation ($\sigma$), maximum, mean, and median of the displacement field errors were tabulated in Table II with the exception of ECC method with Map pair 1, which could not converge. The tabulated results were also verified by a visual representation of the alignment from Map 1 through 6 as shown in Fig. 5 6 7 8 9 10.

Comparing the methods in every experiment in Table II and Fig. 5 6 7 8 9 10, the Decomposition-Optimization based method (DO) proved the closest to the ground truth in terms of all the statistical error metrics performed on the displacement field. This is in line with the comparison of [18], which is based on manually selecting corresponding keypoints.

Fig. 5. Map 1 alignment results, where the main emphasis is on scale as described in Table I

(Hough) (SIFT) (CPD)

(Voronoi) (DO)

Fig. 6. Map 2 alignment results, where the main emphasis is on local small deformations as described in Table I

ECC Hough SIFT

CPD Voronoi DO

Fig. 7. Map 3 alignment results, where the main emphasis is on large deformations and brokenness as described in Table I

ECC Hough SIFT

CPD Voronoi DO

In addition to the previous lateral analysis of Table II, a comparison between maps is made to find the strength of each method with respect to the various alignment challenges represented in 6 Maps, Fig. I. However, the data in Table II between different maps is not comparable longitudinally in its current form, since the errors are in pixels and a meter/pixel conversion resolution is not available for synthetically generated maps. This is remedied through the computation of the normalized relative errors with respect to the Decomposition-Optimization based method (DO), the winning method in the lateral comparison. Through these calculations, as shown in Table III and IV, the performance of each method is sorted in decreasing order along the various maps and their associated challenges. It should be noted that the purpose of this analysis is showing the possible applications of our system in the field. However, for producing an accurate survey of each method’s strengths, which is not the purpose of this paper, more experiments should be carried out.

ECC relies on photometric similarities between the reference image and the warped image. Moreover, it cannot perform well when introduced to a Gaussian noise with high variance. Mainly, The ECC algorithm uses a Mean Square Distance between the reference points and the warped points. This requires high computational power when the distance is very large, which is the case with a scaling parameter, and will halt the convergence of the algorithm. This explains its failure at aligning Map 1, which emphasized on scaling.

CPD lacks the ability to handle distant maps leading to a bad performance on Map 1, which has an emphasis on scale. However, CPD is a shape-based point set registration, which relies on distinct and consistent patterns present in maps. Thus,
CPD can detect large deformations and patterns, but still relies on distance between points.

In contrast, the Hough based method was able to handle large deformations, brokenness, and partial mapping. However, it can only perform rigid transformations and thus as expected performed poorly on Map 1 with its high scale. Furthermore, its largest weakness lies in local small deformations. This can be explained by the reliance of the Hough representation on line detection, which can be affected by noisy maps that blur the available lines.

The Voronoi method used performs Radon transform for matching, which improves on the Hough transform but still has very similar strength and weaknesses including good performance in the face of large deformations and brokenness, and poor alignment of scaled and locally deformed maps.

SIFT method is an image registration technique relying on SIFT features, which are invariant to scale, local deformations, and partial occlusion. However, it has a disadvantage when it comes to aligning very similar room structures with low unique feature count. This is verified in our experiments where the SIFT method aligns partial Map 4 and scaled Map 1 more than Map 5, which has very similar room structures.

This analysis proves that the results are logical and can be attributed to unique features of every alignment technique. The lateral and longitudinal comparisons show the success of the evaluation system proposed in this paper in assessing alignment methods.

VI. CONCLUSION

This paper provided a new evaluation system to assess the performance of map alignment techniques. The proposed evaluation technique automatically generates error measurements between the alignment displacement field and its corresponding ground truth. In addition, we proposed a ground truth upon which any occupancy grid map alignment method can be assessed. Experiments demonstrated the utility of the proposed assessment system in the comparison between the state of the art in map alignment. Further analysis was implemented to
assess the ability of each method to solve common map alignment challenges, which include various local displacements.

In the future, a more detailed parametric model will be established to better represent the deformations occurring in robot mapping; thus, creating a more accurate ground truth of deformed maps with their deformation fields. Further analysis will also be carried out to test the effectiveness of our system in evaluating the alignment of partial maps pairs.

ACKNOWLEDGEMENT

This work was supported by the University Research Board at the American University of Beirut

REFERENCES


