Aligning the Dissimilar: A Probabilistic Feature-Based Point Set Registration Approach

Martin Danelljan, Giulia Meneghetti, Fahad Shahbaz Khan and Michael Felsberg

The self-archived postprint version of this journal article is available at Linköping University Institutional Repository (DiVA):

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-137895>

N.B.: When citing this work, cite the original publication.

Danelljan, M., Meneghetti, G., Khan, F. S., Felsberg, M., (2016), Aligning the Dissimilar: A Probabilistic Feature-Based Point Set Registration Approach, *Proceedings of the 23rd International Conference on Pattern Recognition (ICPR) 2016*, , 247-252.
<https://doi.org/10.1109/ICPR.2016.7899641>

Original publication available at:

<https://doi.org/10.1109/ICPR.2016.7899641>

Copyright: IEEE

<http://www.ieee.org/>

©2016 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

Aligning the Dissimilar: A Probabilistic Method for Feature-Based Point Set Registration

Martin Danelljan*, Giulia Meneghetti*, Fahad Shahbaz Khan, Michael Felsberg
Computer Vision Laboratory, Department of Electrical Engineering, Linköping University, Sweden
{martin.danelljan, giulia.meneghetti, fahad.khan, michael.felsberg}@liu.se

Abstract—3D-point set registration is an active area of research in computer vision. In recent years, probabilistic registration approaches have demonstrated superior performance for many challenging applications. Generally, these probabilistic approaches rely on the spatial distribution of the 3D-points, and only recently color information has been integrated into such a framework, significantly improving registration accuracy. Other than local color information, high-dimensional 3D shape features have been successfully employed in many applications such as action recognition and 3D object recognition. In this paper, we propose a probabilistic framework to integrate high-dimensional 3D shape features with color information for point set registration. The 3D shape features are distinctive and provide complementary information beneficial for robust registration. We validate our proposed framework by performing comprehensive experiments on the challenging Stanford Lounge dataset, acquired by a RGB-D sensor, and an outdoor dataset captured by a Lidar sensor. The results clearly demonstrate that our approach provides superior results both in terms of robustness and accuracy compared to state-of-the-art probabilistic methods.

I. INTRODUCTION

Registration of 3D-point sets is a challenging problem in computer vision, with many potential applications, such as scene reconstruction, robotics, and 3D object localization. The registration problem involves estimating the relative rigid transformations between two or more point sets. In recent years, the probabilistic registration methods have demonstrated promising results on challenging datasets. These approaches model the distribution of 3D-points as a density function. The registration is then performed by either maximizing a similarity measure between the density models [1], [2], or applying the Expectation Maximization (EM) algorithm to iteratively find the registration parameters [3], [4], [5], [6].

Initially, most existing probabilistic registration approaches only utilized the spatial distribution of the 3D-points. Recently, feature information, such as color, has been integrated in such a probabilistic framework [6]. The integration of local color information is performed by constructing a density model of the joint spatial-color space. Other than local color information, high-dimensional 3D shape features, e.g. 3D-SIFT [7] and PFH [8], have been successfully employed in many applications such as action recognition [9] and 3D object recognition [10]. The 3D features extract information from a local neighborhood in the point set, and are therefore

Fig. 1. Registration of four different Lidar scans (a) of an outdoor scene. The two state-of-the-art GMM-based methods JRMPs [3] (b) and CPPSR [6] (c) fail to register the point sets due to severe occlusions and non-uniform point density. Our method (d) accurately registers the sets by exploiting high dimensional shape features in the registration.

distinctive and possess high discriminative power. Additionally, when color is insufficient, 3D shape information provides complementary information beneficial for such registration tasks. However, the problem of *integrating* high-dimensional shape features and *fusing* them with color information are yet to be investigated for probabilistic point set registration.

The recently introduced color-based probabilistic framework [6] efficiently models the distribution of feature observations with a mixture model in the color space. However, this strategy cannot be directly employed for 3D shape features due to their high dimensionality. In the context of object recognition, high-dimensional features are quantized by learning a task-specific codebook, to obtain a bag-of-words (BOW) based probabilistic representation. In this work, we propose a probabilistic representation of the feature space, that is reminiscent to the

*Both authors contributed equally to this work.

BOW methodology, for integrating high-dimensional shape features. In our approach, the feature space is first quantized by clustering based on the feature descriptors extracted from the point sets. This leads to a compact feature representation that is data adaptive. The clustered shape representation then serves as a basis for estimating the local feature distributions in the point set.

As discussed above, the 3D shape features are expected to provide improved registration performance in terms of robustness due to their distinctiveness and high discriminative power. On the other hand, local information, such as color observations, have a high spatial resolution, leading to improved accuracy [6]. A careful strategy when integrating 3D shape information is necessary for improved robustness, while preserving the accuracy of the registration. In this work, we tackle this problem by introducing an adaptive fusion strategy for robust point set registration.

Contributions: In this paper, we propose to integrate high-dimensional 3D-shape features in a probabilistic framework for point set registration. To construct a compact probabilistic representation of the feature space, we introduce an efficient model that is reminiscent to the popular bag-of-words paradigm. To aid the registration process, our approach jointly estimates the local feature distribution parameters in an EM-framework. We further introduce an adaptive strategy for integrating the shape features with color information to obtain improved robustness, while maintaining the accuracy.

We evaluate our approach on two challenging datasets: one indoor scene captured by an RGB-D sensor and one outdoors scene captured by a Lidar. The results demonstrate that integration of shape feature significantly improves the robustness of the registration, with a significant reduction in failure rate. Further, our adaptive fusion strategy ensures the improved robustness is obtained without any significant degradation in accuracy. Figure 1 shows a visualization of the registration results on the Lidar Outdoor dataset. Our method accurately registers the four, highly dissimilar views, while state-of-the-art probabilistic approaches [3], [6] struggle in these challenging situations.

II. RELATED WORK

In recent years, the problem of point sets registration has received much attention. Most of the earlier approaches were based on the classical Iterative Closest Points (ICP) algorithm [11]. The ICP approach computes the rigid transformation that minimizes the distance between the assumed point correspondences. The assumed correspondences are then iteratively updated by finding the nearest neighbors. The standard ICP algorithm requires proper initialization for correct convergence. Several approaches exist in literature that extend the standard ICP algorithm to improve robustness to large initialization errors [12], [13], [14].

Other than the ICP based approaches, probabilistic registration methods have shown promising results in recent years. The probabilistic methods employ a density model of the distribution of the points. The GMMReg [1] approach optimizes

a distance measure between Gaussian mixture models (GMM) of the two point sets, to find the transformation. A correlation based statistical approach for point set registration has been proposed by Tsin and Kanade [2]. In this approach, the KL divergence is maximized with respect to the transformation parameters. Evangelidis et al. [3] propose a generative approach for jointly registering multiple point sets (JRMPS). Different to previous methods [4], [5], this approach assumes the point sets to be generated from the same GMM. The rigid transformation and mixture parameters are jointly estimated using a EM-based Maximum Likelihood (ML) optimization.

Despite the success of the above mentioned probabilistic approaches, feature information such as color and shape have been largely ignored. In a recent work, Danelljan et al. [6] propose a probabilistic framework to integrate color information for point set registration. The approach extends the probabilistic model of [3] by modeling the density of the joint point-color space. In [6] only color information was investigated. However, the model is generic and can be used with any invariant feature.

In the context of 3D object recognition, 3D shape features have demonstrated promising results [10], [15]. These features typically integrate information from a spatial neighborhood in the point set into a high-dimensional descriptor. Feature-based registration approaches are typically based on matching to obtain correspondences. Poreba et al. [16] developed a method based on features consisting of two steps: an initial estimation based on robust feature matching using RANSAC and a second step that refines the initial transformation. Basdogan et al. [17] propose a registration framework based on a geometric descriptor and an efficient k-NN search for finding correspondences. Different from these methods, we do not employ explicit matching of feature descriptors. Instead, we investigate the integration of 3D shape features in a probabilistic registration framework.

III. PROBABILISTIC POINT SET REGISTRATION FRAMEWORK

Our registration framework is based on the recent Color-based Probabilistic Point Set Registration (CPPSR) method [6]. This framework extends the probabilistic registration model [3] with feature observations for increased robustness and accuracy. In [6], only the incorporation of color measurements obtained from the sensor, e.g. an RGB-D camera or a Lidar, was investigated. The model can however be extended to any feature information that is invariant to rigid transformations. Unlike most registration methods, the approaches [6], [3] allows joint registration of multiple point sets.

In the CPPSR, a point cloud $\mathcal{X}_i$ is modeled as a set of observed 3D-points $\mathbf{x}_{ij} \in \mathbb{R}^3$ and corresponding feature values $y_{ij} \in \Omega$. Here, Ω denotes the feature space and $(\mathbf{x}_{ij}, y_{ij}) \in \mathcal{X}_i$ is the j th observation in the i th point set. We denote the random variables associated with the corresponding observations with capital letters $\mathbf{X}_{ij}, Y_{ij}$. All observations are assumed to originate from a common joint distribution $p_{\mathbf{V}, \mathbf{Y}}$ that represents the scene in a reference coordinate

system (reference frame). The points $\mathbf{X}_{ij}$ in set i are related to the reference frame by an unknown rigid transformation $\phi_i(\mathbf{x}) = R_i\mathbf{x} + \mathbf{t}_i$, that maps the points in $\mathcal{X}_i$ to the reference frame. The transformed observations are thus distributed as $(\phi_i(\mathbf{X}_{ij}), Y_{ij}) \sim p_{\mathbf{V}, Y}$. The registration problem is then formulated as finding the transformation parameters $R_i, \mathbf{t}_i$ along with the model parameters for the density $p_{\mathbf{V}, Y}$, given the observations $(\mathbf{x}_{ij}, y_{ij})$.

The density of the joint point-feature space is described as a mixture model. Gaussian components are used in the spatial dimensions to represent the spatial distribution of 3D-points. The feature distribution at each spatial component is modeled by a mixture of non-parametric components. For an observation $(\mathbf{X}, Y)$, a pair of discrete latent random variables (Z, C) are introduced. These assign the observation to the spatial mixture component $Z \in \{0, \dots, K\}$ and the feature component $C \in \{1, \dots, L\}$. Here, K and L denote the number of spatial and feature components respectively. The mixture model of the joint point-feature space is based on the conditional independence assumption $\mathbf{X} \perp\!\!\!\perp Y, C | Z$, which enables the following factorization of the complete-data likelihood, $p_{\mathbf{X}, Y, C, Z} = p_{\mathbf{X}|Z} p_{Y|C, Z} p_{C|Z} p_Z$. The factor p_Z is defined by the mixture weight $\pi_k = p_Z(k)$ of component k . The first factor is given by a Gaussian function $p_{\mathbf{X}|Z}(\mathbf{x}|k) = \mathcal{N}(\phi_i(\mathbf{x}); \boldsymbol{\mu}_k, \Sigma_k)$, where i is the set from which the observation originates. In addition, a uniform component is used for $k = 0$ to model outliers.

The feature distribution is modeled by the factors $p_{Y|C, Z}$ and $p_{C|Z}$. In the CPPSR, a general non-parametric component density function $p_{Y|C, Z}(y|l, k) = B_l(y)$ is used for $k \geq 0$. As for the spatial case, a uniform component is used for $k = 0$. Since the components B_l are non-parametric, the feature distribution for each spatial component k is completely determined by the parameters $\rho_{kl} = p_{C|Z}(l|k)$. They represent the feature component weights for each k . By marginalizing over the latent variables Z, C , the mixture model of the observation (X, Y) is computed as,

$$p_{\mathbf{X}, Y}(\mathbf{x}, y) = \sum_{k=1}^K \sum_{l=1}^L \pi_k \rho_{kl} B_l(y) \mathcal{N}(\phi_i(\mathbf{x}); \boldsymbol{\mu}_k, \Sigma_k) + \pi_0 \mathcal{U}_U(\phi_i(\mathbf{x})) \mathcal{U}_\Omega(y). \quad (1)$$

Here, $\mathcal{U}_U$ and $\mathcal{U}_\Omega$ denote uniform distributions over the scene and the feature space respectively.

The registration is performed by finding a Maximum Likelihood (ML) estimate of the mixture model parameters $\Theta = (\{\pi_k, \boldsymbol{\mu}_k, \Sigma_k, \rho_{k1}, \dots, \rho_{kL}\}_{k=1}^K, \{R_i, \mathbf{t}_i\}_{i=1}^M)$. This is obtained using Expectation Conditional Maximization (ECM) [18] as described in [6], [3]. In the Expectation step, the posterior distributions of the latent variables are updated. The two consecutive Conditional Maximization steps updates the transformation and mixture parameters respectively. The ECM process thus jointly estimates the rigid transformations and the density model of the scene.

IV. OUR APPROACH

Here, we present our feature-based probabilistic registration approach. Our framework integrates high-dimensional 3D shape features in a probabilistic manner, for increased robustness of the registration.

A. Feature Description

In this work, we investigate the use of descriptive high-dimensional features in a probabilistic registration framework. Different from point-wise color observations, such high-dimensional features capture the geometrical properties of the local neighborhood and are typically based on histograms. In our experiments, we employ the Point Feature Histograms (PFH) [8] due to its discriminative power and invariance to rigid transformations. However, other types of invariant features can also be employed in our framework. The PFH uses both the locations and normals of points in a fix sized neighborhood of N points. For each pair of points in the neighborhood, three angular features are extracted using an invariant reference frame. The descriptor is then constructed as a 3-dimensional histogram of the extracted angles for all pairs, resulting in a 5^3 -dimensional feature vector. We refer to [8] for more details.

The incorporation of high-dimensional features into the probabilistic framework presented in section III, requires a set of mixture components B_l to be defined in the feature space. Danelljan et al. [6] used products of B-spline functions to construct a probabilistic model of the point-wise color feature observations. The components were placed in a regular grid in the 3-dimensional HSV space. However, this strategy implies an exponential increase of the number of feature components L with the dimensionality of the feature space. It is therefore not suitable for high-dimensional 3D shape features. Instead, we cluster the feature space, using a methodology reminiscent to the Bag of Words (BoW) for image classification. This enables a compact and data adaptive representation of the feature space.

As a first step in our registration pipeline, 3D shape features are extracted from all point sets. The feature space is then clustered using K-means, based on all extracted feature vectors. The observed feature value of a point $\mathbf{x}_{ij}$ is represented by the index $y_{ij} \in \{1, \dots, L\}$ of the closest cluster centroid. Here, L is the number of K-means clusters. This effectively transforms the features to the discrete space $\Omega = \{1, \dots, L\}$. The feature components are set to the indicator functions $B_l = \delta_l$. Thus, $B_l(y_{ij}) = 1$ whenever y_{ij} belongs to the l th cluster and $B_l(y_{ij}) = 0$ otherwise. In our model, the feature component weights ρ_{kl} specify the categorical distribution of a feature vector from cluster l appearing at the spatial component k . This resembles a normalized BoW histogram computed in a spatial neighborhood.

B. Feature Distribution Initialization

Since our Maximum Likelihood estimation problem is non-convex, the initialization of the parameters Θ is an important step in EM-based frameworks. Here, we propose a robust

Fig. 2. A visualization of the Lidar Outdoor Dataset, consisting of 4 Lidar scans of the same scene. The dataset is extremely challenging due to severe occlusions and varying point density.

initialization procedure for the feature component parameters ρ_{kl} . The proposed approach is more suitable to the feature representation introduced in section IV-A. Compared to the low-level color observations employed in [6], high-dimensional 3D shape features are more descriptive since they integrate information from a spatial neighborhood. Such features are therefore more discriminative as a spatial neighborhood typically contains features from a few different components. To fully exploit the descriptiveness of 3D shape features, we enhance the initialization procedure of the mixture weights ρ_{kl} . In [6], these weights were initialized by uniform sampling on the $L - 1$ simplex for each component k . Instead, we draw independent samples from a Dirichlet distribution $(\rho_{k1}^{(0)}, \dots, \rho_{kL}^{(0)}) \sim \text{Dir}(\alpha \mathbf{m})$ to obtain the initial feature weights for each spatial component k . Here, α is a concentration parameter and $\mathbf{m} = (m_1, \dots, m_L)$ is the frequency of feature values $m_l = \frac{1}{N} \sum_{ij} \delta_l(y_{ij})$ normalized with the number of observations $N = \sum_{ij} 1$.

The normalized frequency $\mathbf{m}$ specifies the expectation of the Dirichlet distribution, the concentration parameter α specifies its shape. Increasing α concentrates the probability mass around $\mathbf{m}$. Setting $\alpha = L$ leads to approximately uniform sampling, provided that $\mathbf{m}$ is almost uniform. On the other hand, a decreased value of α moves the probability mass towards the boundaries of the simplex. This provides samples of more distinctive categorical distributions $(\rho_{kl}^{(0)})_{l=1}^L$, where most values are close to zero. A small concentration parameter value proved important for the convergence speed and robustness of our registration method. Throughout our experiments, we use $\alpha = 1$.

C. Adaptive Feature Fusion

3D shape features, such as PFH, integrate information from a spatial neighborhood. This leads to a larger discriminative power but also to a reduced spatial resolution in the feature representation. An incorporation of such features can therefore lead to increased robustness at the cost of a reduced accuracy. To avoid this issue, we employ an approach which corresponds to multi-resolution search strategy. The 3D shape features are used in the first half of the EM iterations in the registration process. This alleviates the problem of converging to the correct local Maximum Likelihood mode. The estimate is then

Fig. 3. A baseline comparison of three different versions of our approach. The recall plot is computed from 200 multi-view registrations on the Lidar outdoor dataset. We compare our final approach (FPPSR-CA), employing both the clustering-based representation and the adaptive fusion strategy, with a straightforward feature representation (FPPSR-A) and no adaptive fusion strategy (FPPSR-C). Our final version (FPPSR-CA) achieves significantly improved robustness and accuracy compared to the two baseline versions.

refined by performing the second half of the EM iterations without the 3D features, for preserved accuracy.

V. EXPERIMENTS

We evaluate our approach on two challenging datasets: an outdoor dataset [19] acquired by a FARO Focus 3D Lidar and the Stanford Lounge [20] acquired by a Kinect RGB-D sensor.

A. Details and parameters

In our experiments, we fix the number of spatial components $K = 500$, the outlier ratio parameter $\pi_0 = 0.005$ and the number of ECM iterations (100) for the JRMPS, CPPSR and our approach. The number of feature clusters L in our method, controls the trade-off between distinctiveness and invariance of the feature representation. We found $L = 10$ to provide a good balance on all datasets.

We use the Frobenius norm between the rotation matrices [3], [6] to quantitative evaluation. The rotation error is defined as $\|\hat{R} - R\|_F$, where $\hat{R}$ represents the estimated rotation and R denotes the ground-truth rotation. Our approach is implemented in MATLAB.

B. Baseline Comparison

We first perform a baseline comparison on the Lidar Outdoor Dataset [19]. The dataset consists of four scans, contain-

TABLE I

THE RESULTS FROM THE MULTI-VIEW REGISTRATION EXPERIMENTS ON THE LIDAR OUTDOOR DATASET, IN TERMS OF AVERAGE ERROR AND STANDARD DEVIATION FOR INLIERS AND FAILURE RATE. OUR APPROACH ACHIEVES A SIGNIFICANT REDUCTION IN FAILURE RATE COMPARED TO THE STATE-OF-THE-ART APPROACHES.

	Avg. err	Std. dev.	Failure rate (%)
JRMPS [3]	$1.26 \cdot 10^{-2}$	$3.18 \cdot 10^{-3}$	52.9
CPPSR [6]	$1.20 \cdot 10^{-2}$	$4.40 \cdot 10^{-3}$	52.6
Our	$1.53 \cdot 10^{-2}$	$9.22 \cdot 10^{-3}$	31.2

ing more than one million points each. Figure 2 shows all the four views. The dataset is challenging due to severe occlusions and variations in the point density. This dissimilarity between the individual point sets significantly complicates the task of registering the different scans. We perform a multi-view registration experiment by jointly aligning all four views. We perform 200 registrations by initializing each point set with a uniformly sampled random rotation. In each view, about 5000 points are sampled using a keypoint detector¹ in order to partially alleviate the uneven distribution of the points. The PFH descriptor for each keypoint is computed using the full point set to ensure a sufficient amount of neighboring points.

Three different feature-based versions of our method are evaluated. To verify the feature model proposed in section IV-A, we compare with a version, called FPPSR-A, that instead employs a straightforward model of the feature space. For this purpose, we normalize the PFH descriptors and let the feature components be the coordinate projections $B_l(y) = y^{(l)}$, where $y^{(l)}$ denotes the l th dimension in the normalized histogram y . To validate the adaptive feature fusion, presented in section IV-C, we compare with FPPSR-C, that does not employ this component. Note that FPPSR-A includes the adaptive fusion (section IV-C) and that FPPSR-C employs the clustering-based feature representation (section IV-A). Lastly, we evaluate the version FPPSR-CA that employs both the proposed feature representation and the adaptive fusion. For a fair comparison, we use the initialization strategy described in section IV-B for all three versions.

Figure 3 shows the recall plot of the comparison between the three versions. The recall is obtained by computing the fraction (vertical axis) of pairwise registration errors that are smaller than a rotation error threshold (horizontal axis). Compared to the straightforward feature representation (FPPSR-A), our approach (FPPSR-CA) achieves superior robustness, shown by the increased recall for larger thresholds. Further, our adaptive fusion strategy significantly improves the accuracy of the registration compared to FPPSR-C, while obtaining similar robustness. For the remaining experiments, we use FPPSR-CA as our final approach.

C. Lidar Dataset

Here, we perform a comprehensive comparison of our feature-based approach with the two state-of-the-art probabilistic joint registration approaches: the JRMPS [3], employing

¹We use the PCL implementation of the Scale Invariant Features Transform (SIFT) 3D detector <http://www.pointclouds.org/>.

Fig. 4. A comparison of our method with the JRMPS [3] and CPPSR [6] for the multi-view registration on the Lidar outdoor dataset. Our approach achieves significantly improved robustness while maintaining the accuracy.

no feature information, and CPPSR [6] that utilizes the color observations. We use an experimental setup similar to the baseline experiment (section V-B), but perform over 500 registrations for a more extensive evaluation. Table I reports the results in terms of average inlier error, standard deviation and failure rate. The failure rate measures the robustness and is defined as the percentage of pair-wise rotation errors that are greater than 0.1 (approximately 4 degrees). A registration is regarded an inlier if the error is smaller than 0.1. To evaluate the accuracy of the registrations, we report the average and standard deviation of the pair-wise inlier rotation errors.

Both the JRMPS and the CPPSR struggle in registering the four views, with a failure rate of 52.9% and 52.6% respectively. Our method significantly improves the state-of-the-art with a failure rate of 31.3%, while maintaining a competitive accuracy of $1.50 \cdot 10^{-2}$ with respect to both JRMPS ($1.26 \cdot 10^{-2}$) and CPPSR ($1.20 \cdot 10^{-2}$). Note that the error measures are computed on the inlier registrations. Therefore, the vastly increased number of successful registrations of our method leads to a slightly lower average accuracy. On the other hand, in the recall plot (figure 4) our approach provides consistently better results for all error thresholds. This indicates that our method in fact maintains the accuracy of JRMPS and CPPSR, while significant improving the robustness.

In summary, the results clearly demonstrate that a proper integration of high-dimensional 3D shape features leads to superior registration performance in challenging scenarios. Further, our approach efficiently registers multiple views, despite partial overlap, occlusions and varying point density. A visualization of an example registration is shown figure 1.

D. Stanford Lounge Dataset

Finally, we present results on the Stanford Lounge Dataset [20], consisting of 3000 RGB-D frames acquired by Kinect sensor. As ground-truth, we employ the poses provided by the authors [20]. For computational efficiency, we randomly downsample the frames to 10k points. In our experiments, we do not observe any improvements when using keypoint sampling techniques. This is partially attributed to the fact that the variations of the point density is less significant in this dataset. For every subsampled frame, the PFH descriptors

Fig. 5. A qualitative comparison on the Stanford Lounge dataset. Registration results are shown for an example RGB-D frame. Our approach (d) more accurately registers the point sets (a), compared to both JRMPS (b) and CPPSR (c).

TABLE II

A COMPARISON OF OUR APPROACH WITH STATE-OF-THE-ART REGISTRATION METHODS ON THE STANFORD LOUNGE DATASET. THE RESULTS ARE REPORTED IN TERMS OF FAILURE RATE, AVERAGE, AND STANDARD DEVIATION OF THE INLIER ROTATION ERRORS. OUR APPROACH SIGNIFICANTLY REDUCES THE RELATIVE FAILURE RATE BY 40% ON THIS DATASET.

	Avg. err	Std. dev.	Failure rate (%)
ICP [11]	$4.32 \cdot 10^{-2}$	$2.53 \cdot 10^{-2}$	15.70
Color GICP [21]	$1.72 \cdot 10^{-2}$	$1.75 \cdot 10^{-2}$	1.27
JRMPS [3]	$1.78 \cdot 10^{-2}$	$1.35 \cdot 10^{-2}$	3.67
CPPSR [6]	$1.54 \cdot 10^{-2}$	$1.08 \cdot 10^{-2}$	1.00
Our	$1.51 \cdot 10^{-2}$	$1.08 \cdot 10^{-2}$	0.60

are computed for each sampled point, by utilizing the complete point set. As in [6], we perform pairwise registration between frame number n and $n + 5$ for all the frames in the dataset.

Table II reports the average error, the standard deviation, and the failure rate for the compared methods. In addition to the probabilistic methods, we also compare with the standard ICP and the Color-GICP. The JRMPS approach provides a failure rate of 3.67%. The recently introduced CPPSR, employing the color features, obtains competitive results with a failure rate of 1.00%. Our approach significantly improves the state-of-the-art on this dataset, with a failure rate of 0.60%. It is worth to mention that our approach provides this significant reduction of failure rate without any degradation in accuracy. Figure 5 shows a qualitative comparison of our approach with both the JRMPS and CPPSR on the Stanford dataset.

VI. CONCLUSION

We propose a probabilistic framework to integrate high dimensional 3D features for point set registration. Our approach constructs a compact probabilistic representation by clustering the high-dimensional feature space. The local feature distribution parameters are jointly estimated in an EM-framework. Moreover, we introduce an adaptive fusion strategy to integrate high-dimensional 3D shape features with local color information. Experiments on two challenging datasets clearly demonstrate that our approach leads to significant improvement in robustness without any degradation in accuracy. Future work involves further investigations on the impact of the clustered shape representation employed in our framework. Another research direction is to perform a comprehensive evaluation of the 3D shape descriptors for probabilistic point set registration.

Acknowledgments: This work has been supported by SSF (VPS), VR (EMC²), Vinnova (iQMatic), EU’s Horizon 2020 R&I program grant No 644839, the Wallenberg Autonomous Systems Program, the NSC and Nvidia.

REFERENCES

- [1] B. Jian and B. C. Vemuri, “Robust point set registration using gaussian mixture models,” *PAMI*, vol. 33, no. 8, pp. 1633–1645, 2011.
- [2] Y. Tsin and T. Kanade, “A correlation-based approach to robust point set registration,” in *ECCV*, 2004, pp. 558–569.
- [3] G. D. Evangelidis, D. Kounades-Bastian, R. Horaud, and E. Z. Psarakis, “A generative model for the joint registration of multiple point sets,” in *ECCV*, 2014, pp. 109–122.
- [4] R. Horaud, F. Forbes, M. Yguel, G. Dewaele, and J. Zhang, “Rigid and articulated point registration with expectation conditional maximization,” *PAMI*, vol. 33, no. 3, pp. 587–602, 2011.
- [5] A. Myronenko and X. B. Song, “Point set registration: Coherent point drift,” *PAMI*, vol. 32, no. 12, pp. 2262–2275, 2010.
- [6] M. Danelljan, G. Meneghetti, F. Shahbaz Khan, and M. Felsberg, “A probabilistic framework for color-based point set registration,” in *CVPR*, 2016.
- [7] P. Scovanner, S. Ali, and M. Shah, “A 3-dimensional sift descriptor and its application to action recognition,” in *ACM MM*, 2007.
- [8] R. B. Rusu, Z. C. Marton, N. Blodow, and M. Beetz, “Learning informative point classes for the acquisition of object model maps,” in *ICARCV*, 2008, pp. 643–650.
- [9] A. Kläser, M. Marszalek, and C. Schmid, “A spatio-temporal descriptor based on 3d-gradients,” in *BMVC*, 2008.
- [10] Y. Guo, F. A. Sohel, M. Bennamoun, M. Lu, and J. Wan, “Rotational projection statistics for 3d local surface description and object recognition,” *IJCV*, vol. 105, no. 1, pp. 63–86, 2013.
- [11] P. J. Besl and N. D. McKay, “A method for registration of 3-d shapes,” *PAMI*, vol. 14, no. 2, pp. 239–256, 1992.
- [12] A. Rangarajan, H. Chui, and F. L. Bookstein, “The softassign procrustes matching algorithm,” in *IPMI*, 1997.
- [13] D. Chetverikov, D. Stepanov, and P. Krsek, “Robust euclidean alignment of 3d point sets: the trimmed iterative closest point algorithm,” *IMAVIS*, vol. 23, no. 3, pp. 299–309, 2005.
- [14] A. Segal, D. Hähnel, and S. Thrun, “Generalized-icp,” in *RSS*, 2009.
- [15] Y. Guo, M. Bennamoun, F. A. Sohel, M. Lu, J. Wan, and N. M. Kwok, “A comprehensive performance evaluation of 3d local feature descriptors,” *IJCV*, vol. 116, no. 1, pp. 66–89, 2016.
- [16] M. Poreba and F. Goulette, “A robust linear feature-based procedure for automated registration of point clouds,” *Sensors*, vol. 15, no. 1, pp. 1435–1457, 2015.
- [17] C. Basdogan and A. C. zireli, “A new feature-based method for robust and efficient rigid-body registration of overlapping point clouds,” *The Visual Computer*, vol. 24, no. 7-9, pp. 679–688, 2008.
- [18] X. L. Meng and D. B. Rubin, “Maximum Likelihood Estimation via the ECM Algorithm: A General Framework,” *Biometrika*, vol. 80, no. 2, pp. 267–278, 1993.
- [19] J. Unger, A. Gardner, P. Larsson, and F. Banterle, “Capturing reality for computer graphics applications,” in *Siggraph Asia Course*, 2015.
- [20] Q.-Y. Zhou and V. Koltun, “Dense scene reconstruction with points of interest,” *ACM Trans. Graph.*, vol. 32, no. 4, pp. 112:1–112:8, 2013.
- [21] M. Korn, M. Holzkothen, and J. Pauli, “Color supported generalized-icp,” in *VISAPP*, 2014.