A Probabilistic Framework for Color-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

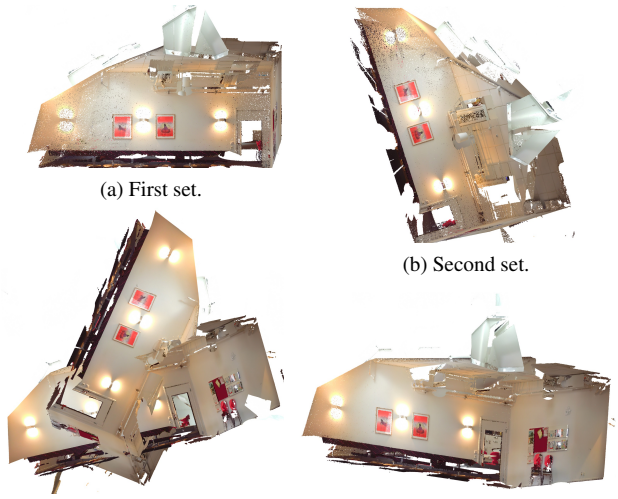
In recent years, sensors capable of measuring both color and depth information have become increasingly popular. Despite the abundance of colored point set data, state-of-the-art probabilistic registration techniques ignore the available color information. In this paper, we propose a probabilistic point set registration framework that exploits available color information associated with the points. Our method is based on a model of the joint distribution of 3D-point observations and their color information. The proposed model captures discriminative color information, while being computationally efficient. We derive an EM algorithm for jointly estimating the model parameters and the relative transformations.

Comprehensive experiments are performed on the Stanford Lounge dataset, captured by an RGB-D camera, and two point sets captured by a Lidar sensor. Our results demonstrate a significant gain in robustness and accuracy when incorporating color information. On the Stanford Lounge dataset, our approach achieves a relative reduction of the failure rate by 78% compared to the baseline. Furthermore, our proposed model outperforms standard strategies for combining color and 3D-point information, leading to state-of-the-art results.

1. Introduction

3D-point set registration is a classical computer vision problem with important applications. Generally, the points originate from measurements of sensors, such as time-of-flight cameras and laser range scanners. The problem is to register observed point sets from the same scene by finding their relative geometric transformations. One class of approaches [2, 16], based on the Iterative Closest Point (ICP) [1], iteratively assumes pairwise correspondences and then finds the transformation by distance minimization. Alternatively, probabilistic methods [5, 7, 9, 14] model the distribution of points using *e.g.* Gaussian Mixture Models (GMMs).

Recently, probabilistic approaches demonstrated promising results for point set registration [5, 7]. The improved performance in probabilistic methods is achieved by modeling the distribution of points as a density function. The probabilistic approaches can be further categorized into correlation-based and Expectation Maximization (EM) based methods. The correlation-based approaches [9, 17] estimate the transformation parameters by maximizing a similarity measure between the density models of the two point sets. Instead, the EM-based methods simultaneously estimate the density model and the transformation parameters [5, 7, 14]. In this paper, we explore probabilistic models for EM-based *colored* point set registration.



(a) First set. (b) Second set. (c) Baseline registration [5]. (d) Our color-based registration.

Figure 1. Registration of the two colored point sets (a) and (b), of an indoor scene captured by a Lidar. The baseline GMM-based method (c) fails to register the two point sets due to the large initial rotation error of 90 degrees. Our method accurately registers the two sets (d), by exploiting the available color information.

State-of-the-art probabilistic techniques [5, 7, 14] rely on the distribution of points in 3D-space, while ignoring additional information, such as color, for point set registration. On the other hand, the increased availability of cheap RGB-D cameras has triggered the use of colored 3D-point sets in many computer vision applications, including 3D object recognition [4], scene reconstruction [3] and robotics [6]. Besides RGB-D cameras, many laser range scanners also capture RGB or intensity information. Additionally, col-

ored point sets are produced by stereo cameras and ordinary cameras by using structure from motion. In this paper, we investigate the problem of incorporating color information for probabilistic point set registration, regardless of the sensor used for capturing the data.

When incorporating color information in probabilistic point set registration, the main objective is to find a suitable probability density model of the joint observation space. The joint space consists of the 3D-point observations and their associated color information. Color information can be incorporated into a probabilistic point set model in two standard ways. (i) A first approach is to directly introduce joint mixture components in the complete observation space. This model requires large amounts of data due to the high dimensionality of the joint space, leading to a high computational cost. (ii) A second approach is to assume stochastic independence between points and color, which enables separable modeling of both spaces. However, this assumption ignores the crucial information about the spatial dependence of color. The aforementioned shortcomings of both fusion approaches motivate us to investigate alternative probabilistic models for incorporating color information.

Contributions: In this paper, we propose a color-based probabilistic framework for point set registration. Our model combines the advantages of (i) and (ii), by assuming *conditional* independence between the location of a point and its color value, given the spatial mixture component. In our model, each spatial component also contains a non-parametric density estimator of the local color distribution. We derive an efficient EM algorithm for joint estimation of the mixture and the transformation parameters. Our approach is generic and can be used to integrate other invariant features, such as curvature and local shape.

Comprehensive experiments are performed on the Stanford Lounge dataset [19] containing 3000 RGB-D frames with ground-truth poses. We also perform experiments on two colored point sets captured by a Lidar: one indoor scene and one outdoor scene [18]. The results clearly demonstrate that our color-based registration significantly improves the baseline method. We further show that the proposed color-based registration method outperforms standard color extensions, leading to state-of-the-art performance. Figure 1 shows registration results on the indoor Lidar dataset, using the baseline [5] and our color-based registration model.

2. Related Work

Initially, most point set registration methods [2, 16] were based on the classical ICP [1] algorithm. The ICP-based approaches alternate between assuming point-to-point correspondences between the two sets and finding the optimal transformation parameters. The standard ICP [1] is known to require a good initialization, since it is prone to get stuck in local minima. Several methods [2, 15, 16] have been proposed to tackle this robustness issue.

Probabilistic registration techniques employ, *e.g.*, Gaussian mixtures to model the distribution of points. In correlation based probabilistic approaches [9, 17], the two point sets are modeled separately in a first step. A similarity measure between the density models, *e.g.* the KL divergence, is then maximized with respect to the transformation parameters. However, these methods lead to nonlinear optimization problems with non-convex constraints. To avoid complex optimization problems, several recent methods [5, 7, 14] simultaneously estimate the density model and the registration parameters in an EM-based framework. Among these methods, the recent Joint Registration of Multiple Point Sets (JRMPS) [5] models all involved point sets as transformed realizations of a single common GMM. Compared to previous EM-based methods [7, 14], JRMPS does not constrain the GMM centroids to the points in a particular set. This further enables a joint registration of multiple point sets.

The use of color information for point set registration has been investigated in previous works [8, 11, 10, 12]. Huhle *et al*. [8] propose a kernel-based extension to the normal distributions transform, for aligning colored point sets. Most approaches [10, 11, 12] aim at augmenting ICP-based methods [1, 16] with color. In these approaches, a metric is introduced in a joint point-color space, to find correspondences in each iteration. A drawback of these ICP variants is that the metric relies on a data dependent parameter that controls the trade-off between spatial distance and color difference. Different to these methods, we incorporate color information in a probabilistic registration framework. The registration is performed using an EM-based maximum likelihood estimation. Next, we describe the baseline probabilistic registration framework.

3. Joint Registration of Point Sets

We base our registration framework on the JRMPS [5] method, since it has shown to provide improved performance compared to previous GMM based approaches [7, 14]. Contrary to these methods, JRMPS assumes both sets to be transformed realizations of one reference GMM. This avoids the underlying asymmetric assumption of using one of the sets as a reference model in the registration [7, 14]. Further, the JRMPS has the advantage of naturally generalizing to joint registration of multiple sets.

3.1. Point Set Observation Model

In the problem of joint registration of multiple point sets, the observations consist of 3D-points in M different views of the same scene. The aim is then to find the transformation of each set to a common reference coordinate system, called the reference frame. All observations of 3D-points are assumed to originate from the same spatial distribution $\mathbf{V} \sim p_{\mathbf{V}}$, representing the entire scene. Here, $\mathbf{V} \in \mathbb{R}^3$ is a

random variable (r.v.) of a point in the reference frame, and $p_{\mathbf{V}}$ is the probability density function (p.d.f.) of $\mathbf{V}$.

Let $\mathbf{X}_{ij} \in \mathbb{R}^3$ be the r.v. of the j:th observed point in view $i \in \{1, \dots, M\}$ and let $\mathbf{x}_{ij}$ be its observed value. Observations in view i are related to the reference frame by the unknown rigid transformation $\phi_i(\mathbf{x}) = R_i\mathbf{x} + \mathbf{t}_i$, such that $\phi_i(\mathbf{X}_{ij}) \sim p_{\mathbf{V}}$. The transformed observations $\phi_i(\mathbf{X}_{ij})$ thus have the distribution $p_{\mathbf{V}}$ in the reference frame. Consequently, the p.d.f. of the observation $\mathbf{X}_{ij}$ is given by $p_{\mathbf{X}_{ij}}(\mathbf{x}_{ij}) = p_{\mathbf{V}}(\phi_i(\mathbf{x}_{ij}))$. To simplify notation, we often write $p_{\mathbf{X}_{ij}}(\mathbf{x}_{ij}) = p(\mathbf{x}_{ij})$.

As described above, the observed points are assumed to be transformed samples of the distribution $p_{\mathbf{V}}$. The point distribution $p_{\mathbf{V}}$ is modeled as a mixture of Gaussian distributions. Let K be the number of Gaussian components. We then introduce the discrete latent r.v. $Z \in \{0, \dots, K\}$ that assigns the point $\mathbf{V}$ to the mixture component $Z = k$. The extra 0th component is a uniform distribution that models the occurrence of outlier points. The joint p.d.f. of $\mathbf{V}$ and Z factorizes as $p(\mathbf{v}, z) = p(\mathbf{v}|z)p(z)$. For discrete variables, we use the notation $p(Z = k) = p_Z(k)$. The mixture component weights π_k are defined as the prior probabilities $\pi_k = p(Z = k)$ of the latent variable Z. The conditional distribution of $\mathbf{V}$ given $Z = k$ is then defined as,

$$p(\mathbf{v}|Z = k) = \begin{cases} \mathcal{U}_U(\mathbf{v}), & k = 0 \\ \mathcal{N}(\mathbf{v}; \boldsymbol{\mu}_k, \Sigma_k), & k \neq 0. \end{cases} \quad (1)$$

Here, $\mathcal{U}_U$ denotes a uniform distribution in the convex hull $U \subset \mathbb{R}^3$ of the observations [7]. The multivariate normal distribution with expectation $\boldsymbol{\mu}$ and covariance Σ is denoted by $\mathcal{N}(\cdot\,; \boldsymbol{\mu}, \Sigma)$. The point density function $p_{\mathbf{V}}$ is obtained by marginalizing over the latent variable Z,

$$p_{\mathbf{V}}(\mathbf{v}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{v}; \boldsymbol{\mu}_k, \Sigma_k) + \pi_0 \mathcal{U}_U(\mathbf{v}). \quad (2)$$

Next, we describe how the above described observation model is used for point set registration.

3.2. Point Set Registration

The registration is performed by jointly estimating the transformation and the GMM parameters, in (2), using the EM algorithm. We denote the set of all observations by $\mathcal{X} = \{\mathbf{x}_{ij}\}_{j=1,i=1}^{N_i,M}$ and the collection of corresponding latent variables by $\mathcal{Z} = \{Z_{ij}\}_{j=1,i=1}^{N_i,M}$. Here, N_i denotes the number of observations in point set i. All observations are assumed to be independent. As in [5], a fix outlier weight π_0 is assumed. The model parameters are summarized as,

$$\Theta = \left(\{\pi_k, \boldsymbol{\mu}_k, \Sigma_k\}_{k=1}^{K}, \{R_i, \mathbf{t}_i\}_{i=1}^{M}\right). \quad (3)$$

The point registration is performed by jointly estimating the parameters Θ from the observed data $\mathcal{X}$. In [5], a Maximum Likelihood (ML) estimate of Θ is obtained using the Expectation Maximization (EM) framework. The E-step evaluates the conditional expectation of the complete data log-likelihood $\log p(\mathcal{X}, \mathcal{Z}|\Theta)$. The expectation is taken with respect to the latent variables $\mathcal{Z}$ given the observed data $\mathcal{X}$ and the current estimate of the parameters $\Theta^{(n)}$,

$$\begin{aligned} Q(\Theta; \Theta^{(n)}) &= \mathrm{E}_{\mathcal{Z}|\mathcal{X},\Theta^{(n)}}\left[\log p(\mathcal{X}, \mathcal{Z}|\Theta)\right] \\ &= \sum_{\mathcal{Z}} p(\mathcal{Z}|\mathcal{X}, \Theta^{(n)}) \log p(\mathcal{X}, \mathcal{Z}|\Theta) \end{aligned} \quad (4)$$

In the M-step, the aim is to find the optimizer of (4) as $\Theta^{(n+1)} = \arg\max_{\Theta} Q(\Theta; \Theta^{(n)})$. To obtain a closed form solution, the M-step is divided into two conditional maximization (CM) steps [13], where the transformation and GMM parameters are updated separately [7].

Using the definitions in section 3.1 and the independent observations assumption, the complete data likelihood is expressed as $p(\mathcal{X}, \mathcal{Z}|\Theta) = \prod_{ij} p(\mathbf{x}_{ij}, z_{ij}|\Theta)$, where

$$p(\mathbf{x}_{ij}, Z_{ij} = k|\Theta) = \pi_k \mathcal{N}(\phi_i(\mathbf{x}_{ij}); \boldsymbol{\mu}_k, \Sigma_k)\,,\ k \neq 0. \quad (5)$$

The posterior density of the latent variables factorizes as $p(\mathcal{Z}|\mathcal{X}, \Theta^{(n)}) = \prod_{ij} p(z_{ij}|\mathbf{x}_{ij}, \Theta^{(n)})$. The E-step then reduces to computing the posterior probabilities of the latent variables $\alpha_{ijk}^{(n)} := p(Z_{ij} = k|\mathbf{x}_{ij}, \Theta^{(n)})$ [5]. Eq. 4 now simplifies to,

$$Q(\Theta; \Theta^{(n)}) = \sum_{ijk} \alpha_{ijk}^{(n)} \log p(\mathbf{x}_{ij}, Z_{ij} = k|\Theta). \quad (6)$$

By applying (5) and ignoring constant terms, (6) can be rewritten to the equivalent minimization problem,

$$\begin{aligned} f(\Theta; \Theta^{(n)}) = \sum_{ij} \sum_{k=1}^{K} \alpha_{ijk}^{(n)} \Bigg(&\frac{1}{2} \log |\Sigma_k| \\ &+ \frac{1}{2} \|R_i\mathbf{x}_{ij} + \mathbf{t}_i - \boldsymbol{\mu}_k\|_{\Sigma_k^{-1}}^2 - \log \pi_k \Bigg). \end{aligned} \quad (7)$$

Here, $|\Sigma_k|$ denotes the determinant of Σ_k and we have defined $\|\mathbf{x}\|_{\Sigma_k^{-1}}^2 = \mathbf{x}^{\mathrm{T}}\Sigma_k^{-1}\mathbf{x}$. For simplicity, isotropic covariances are assumed $\Sigma_k = \sigma_k^2 I$, as in [5].

The parameters Θ are updated in the two CM-steps of the algorithm. The first CM-step minimizes (7) with respect to the transformation parameters $\{R_i, \mathbf{t}_i\}_{i=1}^{M}$, given the current GMM parameters $\{\pi_k^{(n-1)}, \boldsymbol{\mu}_k^{(n-1)}, \Sigma_k^{(n-1)}\}_{k=1}^{K}$. The second CM-step minimizes (7) with respect to the GMM parameters given the new $\{R_i^{(n)}, \mathbf{t}_i^{(n)}\}_{i=1}^{M}$. We refer to [5] for the closed form solutions of the two CM-steps. Next we introduce our color based registration technique.

4. Feature Based Point Set Registration

We reformulate the registration problem from section 3 to incorporate feature information associated with each 3D-point. In this work, we investigate the incorporation of *color*

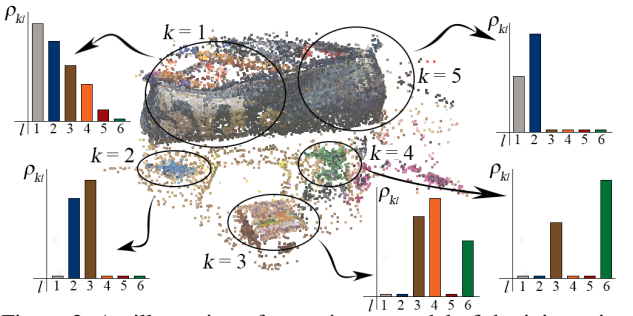


Figure 2. An illustration of our mixture model of the joint point-color space. The ellipses represent spatial mixture components $p(\mathbf{v}|Z=k)$ in our model. Each spatial component k is associated with a mixture model in the color space, given by the weights ρ_{kl} (visualized as histograms). This mixture model encodes the color distribution of points associated with the spatial component k.

information for point set registration. However, our framework is not restricted to color features. It also enables the use of, *e.g.*, structural features that describe the local shape or curvature of the point set.

4.1. Feature Based Observation Model

Our framework assumes the observations to consist of a 3D-point and its associated feature value, *e.g.* color. Let $Y \in \Omega$ denote the r.v. of the feature value associated with the 3D-point $\mathbf{V}$. Here, Ω is the set of all possible feature values, called the feature space. For example, if Y is the color of the 3D-point in normalized HSV coordinates, then the feature space is the unit cube $\Omega = [0,1]^3$. We assume observations of points and features to originate from a common joint distribution $(\mathbf{V}, Y) \sim p_{\mathbf{V},Y}$. The aim of this paper is to propose an efficient yet distinctive mixture model of the joint point-feature density $p_{\mathbf{V},Y}$. Next, we investigate three different strategies to construct a mixture model of the joint point-feature space.

4.1.1 The Direct Approach

A direct generalization of the GMM based registration technique (section 3), is to introduce joint mixture components in the point-feature space $\mathbb{R}^3 \times \Omega$. In general, let $F(\mathbf{v}, y; \theta_k)$ denote the density function of a mixture component in the joint space $(\mathbf{v}, y) \in \mathbb{R}^3 \times \Omega$. Here, θ_k denote the parameters of the k:th component. A mixture model in the joint point-feature space is expressed as

$$p_{\mathbf{V},Y}(\mathbf{v}, y) = \sum_{k=1}^{K} \pi_k F(\mathbf{v}, y; \theta_k). \quad (8)$$

However, this strategy of directly introducing joint components $F(\mathbf{v}, y; \theta_k)$ requires a large amount of data, due to the exponential growth of volume with the number of dimensions (*i.e.* the curse of dimensionality). This leads to a higher computational cost.

4.1.2 The Independent Approach

To alleviate the problems induced by the direct strategy (8), a simple approach is to assume stochastic independence between 3D-points and feature values. The joint distribution $p_{\mathbf{V},Y}$ then factorizes as the product of the marginal distributions for the 3D-points $p_{\mathbf{V}}$ and feature values p_Y, such that $p_{\mathbf{V},Y} = p_{\mathbf{V}} p_Y$. This assumption enables the spatial distribution of points $p_{\mathbf{V}}$ and the distribution of features p_Y to be modeled separately. Let $\tilde{F}$, $\tilde{\theta}_l$ and $\tilde{\pi}_l$ denote the components, parameters and weights respectively for the mixture model of the feature density p_Y. We denote the number of feature components by L. The joint distribution can then be expressed as

$$p_{\mathbf{V},Y}(\mathbf{v}, y) = \sum_{k=1}^{K} \sum_{l=1}^{L} \pi_k \tilde{\pi}_l \mathcal{N}(\mathbf{v}; \boldsymbol{\mu}_k, \Sigma_k) \tilde{F}(y; \tilde{\theta}_l). \quad (9)$$

Here, we have used the GMM presented in section 3.1 for the spatial distribution $p_{\mathbf{V}}$ and ignore the uniform component for simplicity. While the independence assumption allows for a separation of the mixture models, it completely removes information regarding the spatial dependence of feature values. Such information is crucial for aiding the registration process.

The aforementioned approaches have major limitations when incorporating feature information for point set registration. Next, we describe an approach that combines the discriminative power of the direct approach with the efficiency of the independent approach.

4.1.3 Our Approach

We propose a mixture model of the joint point-feature space $\mathbb{R}^3 \times \Omega$ that tackles the drawbacks of the aforementioned approaches. Contrary to the direct strategy (section 4.1.1), our method does not require an increased amount of points to infer the model parameters. We thereby avoid the problems induced by the higher dimensionality of the observation space. Additionally, our model accurately captures the local characteristics in the distribution of features, *e.g.*, how colors are distributed in the scene. This enables our framework to exploit the underlying discriminative feature information associated with each 3D-point.

The proposed mixture model contains a separate feature distribution for each spatial mixture component (illustrated in figure 2). In addition to the spatial latent variable Z, we introduce a second latent r.v. $C \in \{1, \dots, L\}$. This variable assigns a point-feature pair $(\mathbf{V}, Y)$ to one of the L mixture components in the feature space Ω. Our model is based on the conditional independence assumption between the point $\mathbf{V}$ and the feature variables Y, C given the spatial mixture component Z. This is symbolically expressed as

$\mathbf{V} \perp Y, C \mid Z$. Our model assumption enables the following factorization of the joint p.d.f. of $(\mathbf{V}, Y, C, Z)$,

$$\begin{aligned} p(\mathbf{v}, y, c, z) &= p(\mathbf{v}, y, c|z)p(z) = p(\mathbf{v}|z)p(y, c|z)p(z) \\ &= p(\mathbf{v}|z)p(y|c, z)p(c|z)p(z). \end{aligned} \tag{10}$$

The first and fourth factor of (10) do not depend on the feature information, and are defined in section 3.1 (see (1)).

Each spatial component is given a separate feature distribution that characterizes the occurrences of feature values in the vicinity of the component. These distributions are defined by the feature component weights, determined by the conditional probability of a feature component $C = l$ given a spatial component $Z = k$,

$$p(C = l|Z = k) = \rho_{kl}\,, \quad k \neq 0. \tag{11}$$

This expression defines the third factor in (10). The feature mixture weights must satisfy $\rho_{kl} \geq 0$ and $\sum_l \rho_{kl} = 1$ for each spatial component k. For the outlier component $k = 0$, we assume uniform weights $p(C = l|Z = 0) = {}^1\!/_L$.

The second factor $p(y|c, z)$ in (10) is determined by the mixture components in the feature space. Since the feature space Ω can be compact or discrete, we do not restrict our choice to Gaussian distributions. Instead, we consider arbitrary non-negative functions $B_l : \Omega \to \mathbb{R}$ satisfying $\int_\Omega B_l = 1$. We define,

$$p(y|C = l, Z = k) = \begin{cases} \mathcal{U}_\Omega(y), & k = 0 \\ B_l(y), & k \neq 0. \end{cases} \tag{12}$$

As for the spatial mixture components (1), we also use a uniform component in the feature space for $Z = 0$ to model outliers. The integration feature information into the registration process comes at an increased computational cost. In order to minimize this cost, we use non-parametric feature components B_l in our model. This allows the probabilities $B_l(y_{ij})$ to be precomputed and avoids additional costly maximizations of in the M-step.

The proposed mixture model of the joint space is computed by marginalizing over the latent variables Z, C in (10) and using the definitions (1), (11) and (12),

$$\begin{aligned} p_{\mathbf{V},Y}(\mathbf{v}, y) = &\sum_{k=1}^{K}\sum_{l=1}^{L} \pi_k \rho_{kl} B_l(y) \mathcal{N}(\mathbf{v}; \boldsymbol{\mu}_k, \Sigma_k) \\ &+ \pi_0 \mathcal{U}_U(\mathbf{v}) \mathcal{U}_\Omega(y). \end{aligned} \tag{13}$$

Our model (13) differs from the direct approach (8) in that it enables a separation between the point and feature components. It also differs from the independent approach (9) in that the feature component weights ρ_{kl} depend on the spatial component k. Our model thus shares distinctiveness with the direct approach (8) and efficiency with the independent approach (9).

4.2. Registration

Different from the standard GMM based registration (section 3), our model includes the feature observations y_{ij} and the corresponding latent feature variables C_{ij}. In our framework, the set of all observations is $\mathcal{X} = \{(\mathbf{x}_{ij}, y_{ij})\}_{j=1,i=1}^{N_i,M}$ and the collection of corresponding latent variables is $\mathcal{Z} = \{(Z_{ij}, C_{ij})\}_{j=1,i=1}^{N_i,M}$. The model parameters have been extended with the feature distribution weights ρ_{kl} in (11), and are given as

$$\Theta = \left(\{\pi_k, \boldsymbol{\mu}_k, \Sigma_k, \rho_{k1}, \ldots, \rho_{kL}\}_{k=1}^{K}, \{R_i, \mathbf{t}_i\}_{i=1}^{M}\right). \tag{14}$$

We apply an EM procedure, as described in section 3.2, to estimate the parameters (14) of our model. The model assumptions in section 4.1.3 imply the complete data likelihood $p(\mathcal{X}, \mathcal{Z}|\Theta) = \prod_{ij} p(\mathbf{x}_{ij}, y_{ij}, c_{ij}, z_{ij}|\Theta)$, where the joint probability of an observation and its latent variables is

$$\begin{aligned} &p(\mathbf{x}_{ij}, y_{ij}, C_{ij} = l, Z_{ij} = k|\Theta) = \\ &\quad = \pi_k \rho_{kl} B_l(y_{ij}) \mathcal{N}(\phi_i(\mathbf{x}_{ij}); \boldsymbol{\mu}_k, \Sigma_k)\,,\ k \neq 0. \end{aligned} \tag{15}$$

The independence of observations imply the factorization $p(\mathcal{Z}|\mathcal{X}, \Theta^{(n)}) = \prod_{ij} p(z_{ij}, c_{ij}|\mathbf{x}_{ij}, y_{ij}, \Theta^{(n)})$. By applying (15), the latent posteriors are expressed as,[1]

$$\alpha_{ijkl}^{(n)} := p(Z_{ij} = k, C_{ij} = l|\mathbf{x}_{ij}, y_{ij}, \Theta^{(n)}) = \tag{16}$$
$$\frac{\pi_k^{(n)} \rho_{kl}^{(n)} B_l(y_{ij}) \mathcal{N}\left(\phi_i^{(n)}(\mathbf{x}_{ij}); \boldsymbol{\mu}_k^{(n)}, \Sigma_k^{(n)}\right)}{\sum_{q=1}^{K}\sum_{r=1}^{L} \pi_q^{(n)} \rho_{qr}^{(n)} B_r(y_{ij}) \mathcal{N}\left(\phi_i^{(n)}(\mathbf{x}_{ij}); \boldsymbol{\mu}_q^{(n)}, \Sigma_q^{(n)}\right) + \lambda}.$$

Here, the constant in the denominator, originating from the outlier component is given by $\lambda = \frac{\pi_0}{m(U)m(\Omega)}$, where m denotes the reference measure of the space.

For our mixture model, the expected complete data log-likelihood (4) reduces to,

$$Q(\Theta; \Theta^{(n)}) = \sum_{ijkl} \alpha_{ijkl}^{(n)} \log p(\mathbf{x}_{ij}, y_{ij}, C_{ij} = l, Z_{ij} = k|\Theta). \tag{17}$$

As in section 3.2, maximization of the expected complete data log-likelihood (17) can be reformulated as an equivalent minimization problem by applying (15),[1]

$$\begin{aligned} g(\Theta; \Theta^{(n)}) = \sum_{ij}\sum_{k=1}^{K}\sum_{l=1}^{L} \alpha_{ijkl}^{(n)} \Bigg(&\frac{1}{2}\log|\Sigma_k| \\ &+ \frac{1}{2}\|R_i\mathbf{x}_{ij} + \mathbf{t}_i - \boldsymbol{\mu}_k\|_{\Sigma_k^{-1}}^2 - \log\pi_k - \log\rho_{kl}\Bigg). \end{aligned} \tag{18}$$

To simplify the expression (17), we first define the marginal latent posteriors by summing over the latent feature variable $\alpha_{ijk}^{(n)} = \sum_l \alpha_{ijkl}^{(n)}$. This enables our loss (18) to be rewritten as,

[1] See the supplementary material for a detailed derivation.

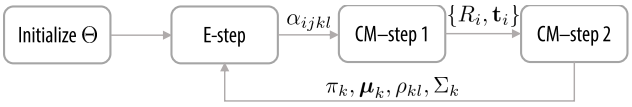


Figure 3. Overview of our EM-based registration. The parameters updated after each step are indicated on the arrow.

$$g(\Theta;\Theta^{(n)}) = f(\Theta;\Theta^{(n)}) - \sum_{ij}\sum_{k=1}^{K}\sum_{l=1}^{L}\alpha_{ijkl}^{(n)}\log\rho_{kl}. \quad (19)$$

Here, $f(\Theta;\Theta^{(n)})$ is the corresponding loss (7) in the standard GMM-based registration. This implies that the transformation parameters $(R_i, \mathbf{t}_i)$ and the spatial mixture parameters $(\pi_k, \boldsymbol{\mu}_k, \Sigma_k)$ can be obtained as in section 3.2. However, in our method, the latent posteriors given by (16) are used in the M-step. Different from section 3, our marginal latent posteriors $\alpha_{ijk}^{(n)}$ thus also integrate feature information into the EM-procedure. Finally, the feature distribution weights are obtained by minimizing the second term in (19) using Lagrangian multipliers,[1]

$$\rho_{kl}^{(n)} = \frac{\sum_{ij}\alpha_{ijkl}^{(n)}}{\sum_{ij}\alpha_{ijk}^{(n)}}, \quad k = 1,\dots,K. \quad (20)$$

We incorporate the estimation of the feature distribution parameters (20) in the second CM-step (see section 3.2), along with the estimation of the other mixture parameters. Figure 3 shows an overview of our approach.

4.3. Feature Description

Here, we provide a detailed description of how the distribution of features is modeled, by the selection of feature mixture components B_l. We restrict our discussion to color features. In our model, the feature observations are represented by an HSV triplet $y = (y^H, y^S, y^V) \in \Omega = [0,1]^3$. In this work, we use second order B-splines to construct the feature components B_l. However, other functions with similar characteristics can also be used. Each component B_l is a separable function $B_l(y) = a_l B_l^1(y^H) B_l^2(y^S) B_l^3(y^V)$. In each dimension, the component is given by a scaled and shifted second order B-spline function B_l^i. The constant a_l is a normalization factor given by the condition $\int_\Omega B_l = 1$. The components B_l are placed in a regular grid inside the unit cube $\Omega = [0,1]^3$. The spacing between the components is set to $1/L_d$ along feature dimension d, where L_d denotes the number of components in dimension d. The total number of components is hence $L = \prod_d L_d$.

Similar to GMMs, our method is able to model multimodal color distributions. However, our choice of nonparametric mixture components B_l is computationally beneficial. In contrast, employing a standard GMM in the color space requires computation of the color means and covariances in the EM-procedure. Our approach further allows the probabilities $B_l(y_{ij})$ to be precomputed for all points.

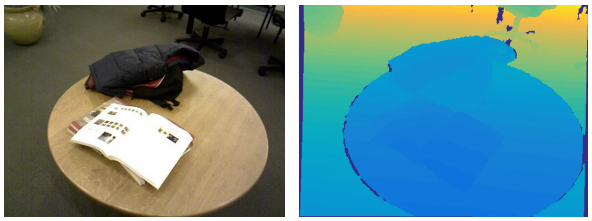


Figure 4. An RGB-D frame from the Stanford Lounge dataset, containing the RGB image (left) and the depth (right).

5. Experiments

We perform a comprehensive quantitative and qualitative evaluations on one RGB-D and two Lidar datasets.

5.1. Details and Parameters

We use the same number of spatial components $K = 500$, the same outlier ratio $\pi_0 = 0.005$ and 100 EM-iterations for both the standard JRMPS and our color-based versions. We also initialize all methods with the same parameters for the spatial GMM. The initial means $\boldsymbol{\mu}_k^{(0)}$ are uniformly sampled on a sphere with the radius equal to the standard deviation of the point distribution. As in [5], we fix the spatial component weights π_k to uniform, since we did not observe any improvement in updating them. The feature component weights ρ_{kl} are initialized by uniformly sampling the $L-1$ simplex for each k. Our approach is implemented in Matlab. Compared to the baseline JRMPS, our approach marginally increases the computation time (25 to 27 sec. on a single core), for 2000 points per set.

For the direct approach, presented in section 4.1.1, the joint components are constructed as products of a spatial Gaussian and a feature component $F(\mathbf{v}, y; \theta_k) = \mathcal{N}(\mathbf{v}; \boldsymbol{\mu}_k, \Sigma_k) B_{l_k}(y)$. Here, B_{l_k} is constructed as in section 4.3, and the index $l_k \in \{1,\dots,L\}$ is selected randomly for each component k. For the independent approach (section 4.1.2), we also set the feature components based on the B-splines presented in section 4.3. That is, we set $\tilde{F}(y;\tilde{\theta}_l) = B_l(y)$ in (9). For all methods, we use $L_d = 4$ feature components in each dimension of the HSV space, which gives $L = 64$ feature components in total. For both the direct and independent approaches, we also employ the additional uniform outlier component (see section 3.1).

Evaluation Criteria: We compute the rotation errors compared to the ground truth by measuring the Frobenius distance between rotation matrices [5]. The rotation error is defined as $\|\hat{R} - R\|_F$, where $\hat{R}$ and R are the estimated and ground-truth relative rotations between two point sets.

5.2. Stanford Lounge Dataset

We perform experiments on the Stanford Lounge Dataset [19], consisting of 3000 RGB-D frames taken by a Kinect. Figure 4 contains an example frame. We use the estimated poses, provided by the authors, as ground truth.

	Avg. error	Std. dev.	Failure rate (%)
ICP [1]	$4.32 \cdot 10^{-2}$	$2.53 \cdot 10^{-2}$	15.70
GMMReg [9]	$6.09 \cdot 10^{-2}$	$2.31 \cdot 10^{-2}$	59.04
Color GICP [11]	$1.72 \cdot 10^{-2}$	$1.75 \cdot 10^{-2}$	1.27
JRMPS [5]	$1.68 \cdot 10^{-2}$	$1.24 \cdot 10^{-2}$	3.41
Direct Approach	$1.91 \cdot 10^{-2}$	$1.30 \cdot 10^{-2}$	2.14
Independent Approach	$1.68 \cdot 10^{-2}$	$1.24 \cdot 10^{-2}$	3.41
Our Approach	$\mathbf{1.47} \cdot 10^{-2}$	$\mathbf{1.01} \cdot 10^{-2}$	**0.74**

Table 1. A comparison with other registration methods on the Stanford Lounge dataset. We report the failure rate along with the average and standard deviation of the inlier rotation errors. Compared to the baseline JRMPS [5], our approach achieves significantly better robustness with a relative reduction in the failure rate by 78%. Further, our approach outperforms other color based methods, including Color GICP [11].

5.2.1 Pairwise Registration

We compare our approach with several state-of-the-art methods with publicly available code, namely ICP[2] [1], GMMReg [9], Color GICP[3] [11], and the baseline JRMPS [5]. To ensure a significant initial transformation, we perform registration between frame number n and $n+5$, for all frames n in the dataset. We randomly downsample the frames to 10000 points. As a measure of robustness, we report the failure rate defined as the percentage of rotation errors larger than 0.1 (approximately 4 degrees). We further define a registration to be an inlier if the error is smaller than 0.1. We compute the average and standard deviation of the inlier rotation errors, as measures of accuracy.

The results are reported in Table 1. The standard ICP obtains inferior performance with a failure rate of 15.7%. The baseline JRMPS achieves a failure rate of 3.41%. The Color GICP provides competitive results with a failure rate of 1.27%. The two standard color extensions, using the independent and direct approaches, provides the failure rates 3.41% and 2.14% respectively. Our approach achieves the best results on this dataset, with a failure rate of 0.74%. Additionally, our method obtains a significant reduction of the average rotation error by 12.5% compared to JRMPS.

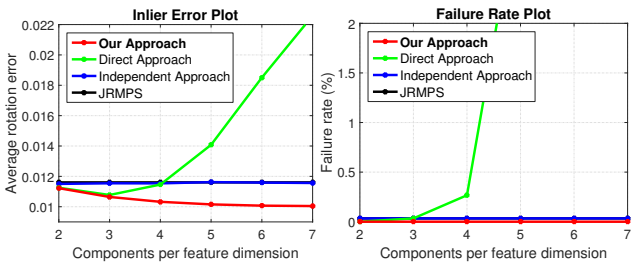


Figure 5. An analysis of the number of feature mixture components L, on the Stanford Lounge dataset. We compare our approach with the baseline JRMPS and the two standard color-extensions. We show the average inlier rotation error (left) and failure rate (right) for different numbers of components per feature dimension L_d in the HSV space. Our approach provides consistent improvements compared to the other probabilistic approaches.

	Avg. error	Std. dev.	Failure rate (%)
JRMPS [5]	$0.913 \cdot 10^{-2}$	$0.636 \cdot 10^{-2}$	0.467
Ours	$\mathbf{0.768} \cdot 10^{-2}$	$\mathbf{0.539} \cdot 10^{-2}$	**0.067**

Table 2. A comparison of joint multi-view registration on the Stanford Lounge dataset, in terms of average inlier error, standard deviation and failure rate. Our approach significantly reduces the relative failure rate with 86% compared to JRMPS.

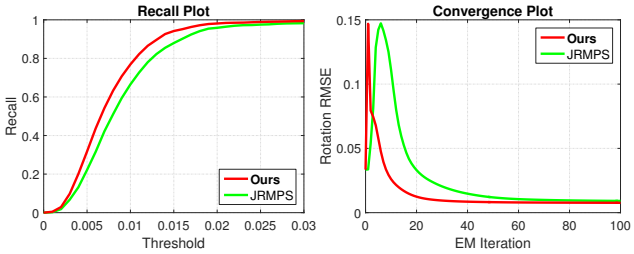


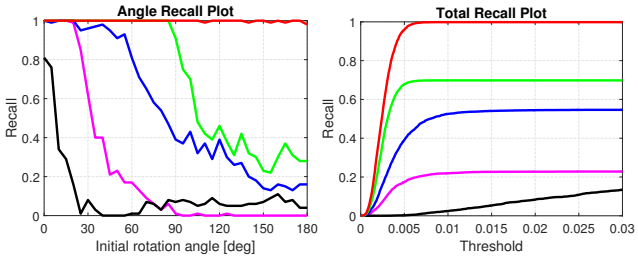
Figure 6. A joint multi-view registration comparison of our method with JRMPS [5] on the Stanford Lounge dataset. The recall plot (left) shows the fraction of correct registrations over a range of rotation-error thresholds. The convergence plot (right) shows the average frame-to-frame inlier rotation error after each EM iteration. Our method demonstrates superior accuracy and robustness, while achieving faster convergence.

In figure 5 we investigate the impact of varying the number of feature components L on the Stanford Lounge dataset, when using 2000 points per set.[4] The left plot shows the average frame-to-frame rotation error for inliers, when increasing the number of components per HSV-dimension from 2 to 7. As a reference, we also include the baseline JRMPS. The independent approach (section 4.1.2) provides similar results to JRMPS. The direct approach (section 4.1.1), requires a larger amount of data points when increasing the number of feature components. The performance therefore rapidly degrades as the number of feature components is increased. Contrary to this, our model benefits from increasing the number of feature components, leading to improved results.

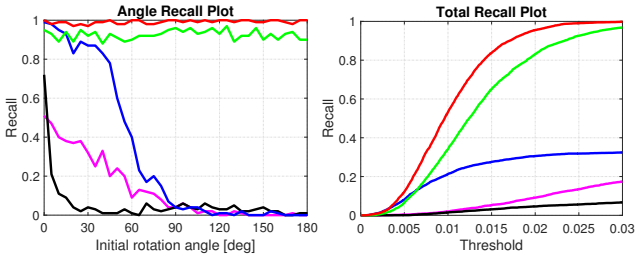
5.2.2 Joint Multi-view Registration

Here, we investigate the performance of our approach for joint registration of multiple point sets. Alignment of multiple point sets is important in many applications. Most registration methods [1, 7, 11] are however limited to pairwise registration. In these cases, multi-view registration must be performed either by sequential pair-wise alignment or by performing a one-versus-all strategy, leading to drift or biased solutions. Similar to JRMPS [5], our method is able to jointly register an arbitrary number of point sets. We perform joint registration of every 10 consecutive frames, with

[2] We use the built-in MATLAB implementation of ICP.
[3] We use the Color GICP implemented in Point Cloud Library.
[4] Analysis of K and π_0 is provided in the supplementary material.



(a) Lidar Indoor dataset.



(b) Lidar Outdoor dataset.

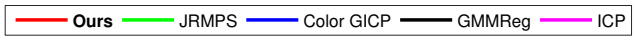


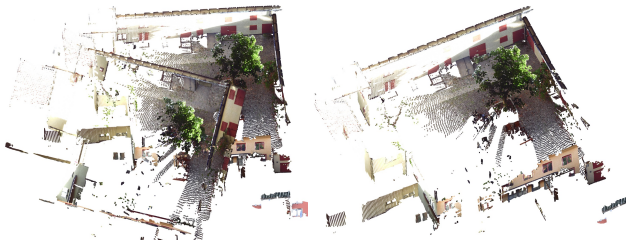
Figure 7. Initialization robustness comparison on the Lidar Indoor (a) and Outdoor (b) datasets. The left plots show the recall at a threshold of 0.025. The recall is computed over 100 randomly sampled rotation axes for each angle. The right plots contain the total recall over all registrations, plotted with respect to the error threshold. Compared to previous methods, our approach provides superior robustness, while maintaining the accuracy.

an interval of 9 frames, on the Stanford Lounge dataset. This implies that joint multi-view registration is performed on frame 1-10, 10-19, etc. Table 2 contains the results, by measuring the frame-to-frame rotation errors. Our color based model reduces the relative failure rate by 86% compared to the baseline JRMPS. In case of average rotation error, our approach provides a significant reduction of 15.9%.

Figure 6 shows the recall and convergence rate plots. Recall is computed as the fraction of frame-to-frame rotation errors smaller than a threshold. In figure 6, the recall is plotted over a range of error thresholds. To compare the convergence rate of our method with the baseline JRMPS, we plot the average frame-to-frame inlier rotation error after each EM iteration. Our method converges in significantly fewer iterations, enabling a more efficient registration.

5.3. Lidar Datasets

We experimented with two Lidar datasets, acquired by a FARO Focus3D. Both consist of more than a million colored 3D points in a single 360 degree view. The Indoor dataset is visualized in figure 1 and the Outdoor dataset is visualized in figure 8. We compare with state-of-the-art methods by evaluating the robustness to initial rotation errors. Registration is performed using initial rotation errors between 0 and 180 degrees with an interval of 5 degrees. For every angle, we uniformly sample 100 random rotation



(a) Color GICP. (b) Ours.

Figure 8. Registration of an outdoor scene captured by a Lidar. Color GICP (a) fails to register the point sets due to a large initial transformation. Our approach (b) accurately register the point sets.

axes. The point sets are constructed by randomly sampling points from the single Lidar scan. For each transformation, we sample two sets with 2000 points each. One of the sets is then transformed with the rotation defined by its corresponding axis and angle. We plot the recall at a rotation error threshold of 0.025 (approximately 1 degree) with respect to the initial angle. We also compare the total recall over all registrations.

Lidar Indoor Dataset: Figure 7a shows the angle robustness comparison in terms of angle recall and total recall. ICP, GMMreg and Color GICP struggle for initial angles larger than 60 degrees. The robustness of JRMSP starts to degrade at an initial angle of 90 degrees. Our approach provides consistent registrations for angles up to 180 degrees.

Lidar Outdoor Dataset: Figure 7b shows the initial angle robustness comparison on the Lidar Outdoor dataset. As in the Indoor dataset, the ICP and Color GICP provides inferior results due to large initial transformations. Our approach provides consistent improvements compared to the JRMPS. Figure 8 shows a qualitative comparison between Color GICP and our approach on this dataset.

6. Conclusions

In this work, we propose a novel probabilistic approach to incorporate color information for point set registration. Our method is based on constructing an efficient mixture model for the joint point-color observation space. An EM algorithm is then derived to estimate the parameters of the mixture model and the relative transformations.

Experiments are performed on three challenging datasets. Our results clearly demonstrate that color information improves accuracy and robustness for point set registration. We show that a careful integration of spatial and color information is crucial to obtain optimal performance. Our approach exploits the discriminative color information associated with each point, while preserving efficiency.

Acknowledgments: This work has been supported by SSF (VPS), VR (EMC^2), Vinnova (iQMatic), EU's Horizon 2020 R&I program grant No 644839, the Wallenberg Autonomous Systems Program, the NSC and Nvidia.

References

[1] P. J. Besl and N. D. McKay. A method for registration of 3-d shapes. *PAMI*, 14(2):239–256, 1992.

[2] D. Chetverikov, D. Stepanov, and P. Krsek. Robust euclidean alignment of 3d point sets: the trimmed iterative closest point algorithm. *IMAVIS*, 23(3):299–309, 2005.

[3] S. Choi, Q. Zhou, and V. Koltun. Robust reconstruction of indoor scenes. In *CVPR*, 2015.

[4] B. Drost, M. Ulrich, N. Navab, and S. Ilic. Model globally, match locally: Efficient and robust 3d object recognition. In *CVPR*, 2010.

[5] 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.

[6] Z. Fang and S. Scherer. Real-time onboard 6dof localization of an indoor mav in degraded visual environments using a rgb-d camera. In *ICRA*, 2015.

[7] R. Horaud, F. Forbes, M. Yguel, G. Dewaele, and J. Zhang. Rigid and articulated point registration with expectation conditional maximization. *PAMI*, 33(3):587–602, 2011.

[8] B. Huhle, M. Magnusson, W. Straßer, and A. J. Lilienthal. Registration of colored 3d point clouds with a kernel-based extension to the normal distributions transform. In *ICRA*, 2008.

[9] B. Jian and B. C. Vemuri. Robust point set registration using gaussian mixture models. *PAMI*, 33(8):1633–1645, 2011.

[10] A. E. Johnson and S. B. Kang. Registration and integration of textured 3d data. *IMAVIS*, 17(2):135–147, 1999.

[11] M. Korn, M. Holzkothen, and J. Pauli. Color supported generalized-icp. In *VISAPP*, 2014.

[12] H. Men, B. Gebre, and K. Pochiraju. Color point cloud registration with 4d ICP algorithm. In *ICRA*, 2011.

[13] X. L. Meng and D. B. Rubin. Maximum Likelihood Estimation via the ECM Algorithm: A General Framework. *Biometrika*, 80(2):267–278, 1993.

[14] A. Myronenko and X. B. Song. Point set registration: Coherent point drift. *PAMI*, 32(12):2262–2275, 2010.

[15] A. Rangarajan, H. Chui, and F. L. Bookstein. The softassign procrustes matching algorithm. In *IPMI*, 1997.

[16] A. Segal, D. Hähnel, and S. Thrun. Generalized-icp. In *RSS*, 2009.

[17] Y. Tsin and T. Kanade. A correlation-based approach to robust point set registration. In *ECCV*, 2004.

[18] J. Unger, A. Gardner, P. Larsson, and F. Banterle. Capturing reality for computer graphics applications. In *Siggraph Asia Course*, 2015.

[19] Q.-Y. Zhou and V. Koltun. Dense scene reconstruction with points of interest. *ACM Trans. Graph.*, 32(4):112:1–112:8, 2013.